In the Land of AKM:

Explaining the Dynamics of Wage Inequality in France

Documents de travail

N° 2023-20– Septembre 2023

Insee

Mesurer pour comprendre

Damien BABET

Olivier GODECHOT

Marco G. PALLADINO

Institut national de la statistique et des études économiques

2023/20

In the Land of AKM: Explaining the Dynamics of Wage Inequality in France*

DAMIEN BABET† OLIVIER GODECHOT‡ MARCO G. PALLADINO§

21 septembre 2023

Département des Études Économiques – Timbre G201
88, avenue Verdier – CS 70058 – 92541 MONTROUGE CEDEX – France
Tél. : 33 (1) 87 69 59 54 – E-mail : d3e-dg@insee.fr – Site Web Insee : http://www.insee.fr

* We dedicate this work to our professor, Christian Baudelot, who was one of the designers of the first version of the BTS database.We thank Kerstin Holzheu, HorngWong, Matthieu Lequien, Sébastien Roux, Pierre Cahuc, Pauline Givord, and the seminar participants at the département des études économiques seminar at Insee, and at the Coin meeting. Access to some confidential data, on which this work is based, was made possible within a secure environment provided by CASD – Centre d'accès sécurisé aux données (Ref. 10.34724/CASD). The research and writing of this paper benefited from the monetary support of the following institutions: Agence Nationale de la Recherche (grant ANR-17-CE41-0009-01)

† Insee
‡Sciences Po, CRIS-CNRS and AxPo
§ Banque de France; Sciences Po

Résumé

Nous utilisons une base de données appariée employeur-salarié nouvellement construite et quasi exhaustive, pour étudier la contribution des entreprises aux inégalités salariales en France. Nous utilisons le modèle d'Abowd, Kramarz et Margolis (1999) (ci-après AKM) pour décomposer la variance des log-salaires entre deux composantes inter- et intra-entreprises.
Notre analyse qui couvre la période de 2002 à 2019 révèle une augmentation significative des inégalités entre entreprises, principalement expliquée par une tendance croissante des travailleurs à haut salaire à se concentrer dans les mêmes entreprises, et dans des entreprises à haut niveau de rémunération. Ces phénomènes sont directement associés à l'évolution de la démographie des entreprises et de la composition de la main-d'œuvre. Au cours de la même période, les centiles inférieurs de salaire ont augmenté davantage que le reste de la distribution, en lien avec l'augmentation du salaire minimum. En conséquence, les inégalités au sein des entreprises ont diminué, ce qui a quasiment compensé la hausse des inégalités entre les entreprises.

Mots clés : Inégalités salariales, prime salariale offerte par les entreprises, décomposition AKM
Codes JEL : E24, J31, C33

Abstract

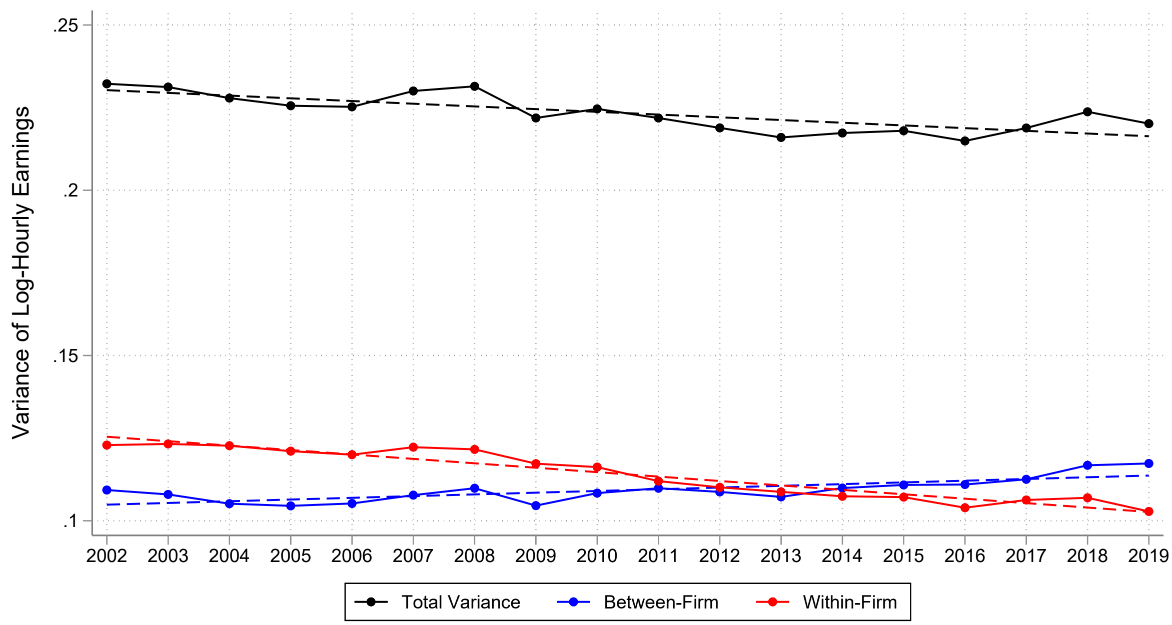
We use a newly built and quasi-exhaustive matched employer-employee database to study firms' contribution towage inequalities in France. We employ the Abowd, Kramarz, and Margolis (1999) model (hereafter AKM) to decompose log-wage variance into between- and within-firm components. Our analysis covering the period from 2002 to 2019 reveals a significant increase in between-firm inequalities, driven by a growing tendency of high-wage workers to cluster together in high premium firms. These phenomena are directly associated with changes in firms' demographics and workforce composition. Over the same period, bottom earnings percentiles increased more than the rest of the distribution, in line with the rise in the legal minimum wage. As a result, within-firm inequalities decreased, almost offsetting the rising between-firm inequalities.

Keyworks : Wage inequality, firm wage premium, AKM decomposition
Codes JEL : E24, J31, C33

Introduction

Wage inequality is a driving force of economic inequalities. Its rise for several decades in most rich countries is well documented[1]. Firms play a central role in driving these dynamics, as evidenced by studies conducted in Germany (Card, Heining, and Kline 2013) and the USA (Song et al. 2019), where rising inequality is largely attributed to between-firm wage disparities. France, however, represents an interesting case, with wage inequality remaining stable or even decreasing in recent decades, despite exhibiting similar polarizing dynamics observed in the other countries, characterized by an increase in between-firm inequalities (Figure 1).

FIGURE 1. Evolution of wage inequality - France



Notes: This figure shows the evolution over time of the variance of log-wage, the between-firm variance of log-wage, and the within-firm variance of log-wage. We compute the overall variance of log-earnings as $\frac{1}{N_t}\sum_i(w_{it}-\overline{w}_t)^2$, the between-firm variance as $\frac{1}{N_t}\sum_f N_{ft}(\overline{w}_{ft}-\overline{w}_t)^2$ and the within-firm variance as $\frac{1}{N_t}\sum_f\sum_{i\in f}(w_{it}-\overline{w}_{ft})^2$, where workers are indexed by i and time by t and firms by f. N_t and N_{ft} denote the number of workers in total and in each firm, respectively; w_{it}, $\overline{w}_t$ and $\overline{w}_{ft}$ are the log worker wage, the overall average log wage and the average log wage within each firm, respectively. All individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included.

To explore the factors contributing to the French exception, we employ the Abowd, Kramarz, and Margolis (1999) model (hereafter AKM) to decompose log-wage variance into between- and within-firm components.

[1] Tomaskovic-Devey et al. (2020) and OECD (2021) for a recent international comparison.

Despite the existence of wage premiums for similar workers across different firms, their contribution to the observed increase in between-firm wage inequalities is negligible, and if anything negative. Similarly to Germany and the USA, in France, we find that changes in the distribution of workers between firms are the primary drivers of the rise in between-firm inequalities. These changes manifest as a notable increase in *segregation*, wherein high-wage workers tend to work more often with similar workers, and a more limited in magnitude but robust rise in *sorting*, wherein high-wage workers are increasingly employed in high-premium firms.

During the same period, wages at the bottom have grown faster than middle and top percentiles. We provide suggestive evidence that these dynamics are associated with French labor market institutions' nature, reinforcing the findings of recent literature (Bozio, Breda, and Guillot 2020; Kramarz et al. 2022; Bilal and Lhuillier 2021). The mix of a significant increase in the minimum wage and the elimination of all employer-paid payroll taxes around it has proven successful at increasing redistribution and balancing the contemporaneous rise in between-firm inequalities.

This paper contributes to the existing literature in three main ways. Firstly, we create a new, quasi-exhaustive matched employer-employee dataset for France which enables us to provide the first decomposition of log-wage variance and its evolution from 2002 to 2019[2]. This dataset allows us to document the increase in segregation and sorting, alongside the decrease in within-firm inequalities.

Secondly, we propose a novel method to correct for the well-known "limited mobility" bias in the measurement of sorting through AKM models described in Abowd et al. (2004), Andrews et al. (2008), and Bonhomme et al. (2023). This bias stems from the limited number of observations available for individual firm and worker parameters. Individual parameter estimations remain consistent, but the variance of the error term is underestimated. We use a split-sampling strategy (building on Chanut (2018), Kline, Saggio, and Sølvsten (2020) and others) and provide proof that, under reasonable hypothesis analogous to Kline, Saggio, and Sølvsten, split-sampling corrects the limited mobility bias on quadratic terms. We also implement Bonhomme, Lamadon, and Manresa (2019) firm clustering method (without random effects) and find the results

[2]The AKM model of log-wages with additive workers and firm fixed effects was originally estimated on French data: a panel sample of 1/24th of French wage earners (without civil servants) from 1976 to 1987. The paper inspired much subsequent work on matched employer-employee datasets, most of them from countries where exhaustive, panelized administrative data was available to researchers: this exhaustivity proved essential to the quality of the estimation for these models. We build such a dataset for France, to bring AKM back to the state of the art in its original land. A first panelization of the BTS was developed in Godechot et al. (2020). We improved the algorithm and made it public for further use cf. Appendix C.

coherent with split sampling. Both Card, Heining, and Kline (2013) and Song et al. (2019) acknowledge the mobility bias in sorting measurement but assume its stability over time. Our corrected results challenge this assumption, revealing the potential impact of the mobility bias on observed dynamics.

Lastly, to gain a deeper understanding of the rise in segregation and sorting, we employ detailed descriptive statistics. We show that most of the rise in between-firm inequalities is driven by shifts in the composition of firms and changes in labor demand for specific occupations. We observe a stronger growth in two dimensions: the size of the workforce in firms employing high-premium, high-wage workers, and the number of firms offering low premiums. Specifically, the group of high-premium, high-wage firms experienced a consistent expansion over time, characterized by an increased presence of managers and engineers, accompanied by a decline in the proportion of blue-collar workers and associate professionals. On the other hand, the emergence of new firms with low premiums was accompanied by specialization in personal services and increased employment of health and social workers.

These trends reflect a structural evolution in the division of work between firms, arising also through outsourcing dynamics - already outlined in other contexts (Goldschmidt and Schmieder 2017; Dorn, Schmieder, and Spletzer 2018; Drenik et al. 2023) - and recently studied in France (Bergeaud et al. 2021; Bilal and Lhuillier 2021; Godechot et al. 2023). We rule out alternative stories potentially related to the rise in between-firm inequalities. We do not find evidence of an increasing return to skills and a change in firms' rent-sharing behavior.

Our findings provide insights into the dynamics of wage inequality and offer an explanation for the apparent divergence between France and other comparable countries. While France has managed to contain wage inequalities through institutional measures, the underlying dynamics driving segregation and sorting remain consistent with those observed elsewhere. The persistence of segregation and sorting in France highlights the strength of these processes, and their successful mitigation within the French context suggests that similar institutional tools could be effective in addressing inequalities elsewhere.

Section 1 provides an overview of our data construction process. In Section 2, we outline the methods used, including the variance decomposition of log-wage based on the AKM model, and the correction of bias through split-sampling or firm clustering. The results are presented in Section 3, followed by a discussion of our methods and findings in Section 4.

1. Data

1.1. Building an Exhaustive Pseudo-panel

We utilize BTS data, which stands for “Base tous salariés", or “all wage-earners file"[3]. It is an exhaustive yearly dataset built from tax returns files provided by firms on their payrolled employees. This dataset serves as the source for French official statistics on wage evolution.

To conduct panel analysis, we need to address the issue of pseudonymity in the data. Each individual is assigned a unique identifying code that changes every year, allowing for cross-sectional analysis but not for long panel analysis. In France, panel analysis on matched employer-employee wage data is typically performed using the “BTS panel" or "DADS panel," which is a narrower panel constructed from a sample of 1/24th of the data before 2002 and 1/12th after. This “narrow panel" samples the same individuals as a permanent demographic panel. The sampling also facilitates additional data quality control and correction work that would be more challenging with the exhaustive data. The oldest years of this narrow panel were the basis for the original AKM. The narrow panel remained the basis for later AKM estimations on French data, notably in Abowd, Kramarz, and Roux (2006) and Coudin, Maillard, and Tô (2018).

However, since 1999, AKM models have been more accurately estimated in countries where researchers have had access to exhaustive panel data, such as the USA, Germany, Sweden, Austria, Italy, Norway, and Denmark. The reduction in sample size and precision due to sampling raises uncertainty in the estimation process. For instance, in the narrow panel, firms need to be approximately 12 times larger to have their firm fixed effects estimated with the same precision as in the exhaustive data. This decrease in precision introduces a larger “limited mobility bias" for variance and sorting estimates. Additionally, the identification of AKM models relies on mobile workers moving between firms, and this is only possible within the group of firms interconnected through such workers. Sampling drastically reduces the proportion of firms belonging to the main connected component and biases the estimation towards larger and more connected firms.

Each yearly BTS file for a given year contains data for both the current year and the

[3] formerly known as “DADS", which stands for “déclaration annuelle de données sociales" or “annual social data declaration". DADS were the main source for BTS, completed with other administrative sources (public sector data for instance), and was progressively replaced by a new administrative source, the DSN (“déclarations sociales nominatives") starting in 2016.

previous year. This overlap allows for matching between yearly files based on common information, such as establishment ID, gender, number of hours, job duration, start and end dates of the job, municipality of work and residence, earnings, and age. The matching procedure provides a single match for approximately 98% of individuals between 2002 and 2019, except for the period 2016-2018 where it drops to 91%-93%. However, there are rare cases where matching is not possible, such as when all matching variables are identical for several individuals or when individual data is modified between yearly files. Workers who are not matched due to career interruptions or matching misses still appear in the panel, but they are represented by multiple identification numbers. This results in an almost exhaustive pseudo panel that we refer to as the “wide panel."

Before 2002, the matching procedure is not applicable due to the lack of linkage between various jobs of a single individual in the files. It is not possible to track a worker through different employers, even within the same year. Matching is only feasible for workers who held the same job for two consecutive years. As a result, the AKM estimation cannot be performed before 2002[4].

We have computed long-term series on sorting from 1976 to 2019 using the narrow panel, which provides wage and career information since 1976 with some missing years and varying data quality[5].

To supplement the analysis, we incorporate exhaustive firm financial data from administrative sources (FICUS/FARE files), which are matched to the wage files and provide information on value-added per worker and total workforce at the legal unit level. We use legal unit identification numbers (SIREN) as our empirical units for firms[6]. Estimates based on establishment identification numbers (SIRET) are very similar[7].

1.2. Sample Restrictions

Following similar works, we exclude public workers, both for comparability and data quality reasons[8]. We focus on ordinary jobs, excluding subsidized contracts, interns, and apprenticeships. Both men and women are included in the analysis. To ensure data preparation simplicity and avoid significant disconnection and high sorting between

[4]The matching procedure is detailed in Annex C.

[5]Historical series in the Annex, Figure A1.

[6]Our approach aligns with other studies that have applied the AKM method to analyze French data. More recently, Insee has started to provide datasets of groups, built from financial links between legal units. There is not enough historical depth yet to measure evolution at this observational unit level.

[7]See Table A3.

[8]Most notably, public servants are not included before 2009.

distant geographic regions, we limit our analysis to metropolitan France.

We divide the data into three adjacent six-year periods: 2002-2007, 2008-2013, and 2014-2019[9]. Each observation consists of a worker / firm / year triplet, where each individual worker is associated with the firm from which she earned the most during the year (or, when equal, for which she worked the most). We refer to these observations as a “wage", or a “job" for convenience. Each worker can appear up to six times during each sample period.

We have information about the number of hours paid, which is rare in this kind of data. Without it, it is common practice in the literature to set a minimum wage for inclusion and exclude women to reduce the risk of misidentifying part-time workers. We can keep both part-time and full-time employees and avoid these exclusions. Our target variable of earnings is the log hourly gross wage (including employees’ social contributions, but not employers’). We restrict the sample to individuals employed for the full year to minimize the impact of annualized payments[10]. We limit the impact of possibly erroneous extreme values with some additional selection. We exclude jobs with an hourly wage inferior to 80% of the legal minimum hourly wage for the corresponding year, or above 1000 times the minimum hourly wage, with less than 100 work hours a year, and observations with missing values for sex, age and employer. We keep only workers between the ages of 16 and 70. These restrictions are applied after matching when the wide panel is already constructed, selecting specific observations while retaining individuals who have been working during the year or receiving unemployment benefits. All sample restrictions might incur selection effects, that could change in time. The restriction to workers employed for the full year, for instance, excludes workers that are more frequently at the bottom of the wage distribution.

The estimation of AKM models relies on connected sets of firms and workers. Two sets of firms are considered disconnected if no worker in the data has worked for two different firms, one from each group. A connected set includes all workers who have ever worked in the firms belonging to that set. Estimations are performed on the main connected set for each period. Descriptive information characterizing both the full population and the largest connected set is provided in Table 1.

In our historical series computed on the narrow panel, paid hours are not reliable

[9]Card, Heining, and Kline (2013) divide their 1985-2009 data into four overlapping seven-year panels, and Song et al. (2019) divide their 1980-2013 data into five adjacent six-year panels.

[10]We found similar results when widening the selection to all individuals whose main job during the year lasts more than 90 days (Table A8). In this larger sample, the connectivity is better and the limited mobility bias is reduced.

TABLE 1. Summary statistics

				Log-Hourly Wage	
	Person*yr	Individuals	Firms	Mean	Std.Dev.
	Overall Sample				
2002-2007	65,457,069	21,356,960	1,203,830	2.67	0.46
2008-2013	68,998,598	18,595,890	1,232,452	2.81	0.45
2014-2019	67,928,369	20,894,672	1,397,646	2.91	0.46
	Largest Connected Set				
2002-2007	58,666,317	18,842,389	536,814	2.69	0.46
2008-2013	61,413,372	16,266,899	560,778	2.83	0.45
2014-2019	59,550,287	18,065,244	564,113	2.94	0.46

Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. The largest connected set entails the group of firms connected by worker mobility.

before 1996, but job duration in days and an indicator variable for part-time jobs exist since 1976. There were other changes in variables in the 44-year period. For instance, the distinction between the public and private sectors is not consistent[11]. Because of these changes, it is not possible to precisely replicate the selection criteria used in the 2002-2019 wide panel. We use a slightly different selection based on Insee's long-term, private sector series selection. In our long-term series, we also compute sorting on the daily wage of full-time workers. While the level of sorting is lower than the hourly wage narrow panel sorting, the trends are generally parallel in the post-1996 period where both measures are available.

[11]Partly because of nationalization and privatization of major firms, such as banks, airlines, posts, and telecommunications, etc. over the period that is difficult to track in the data.

2. Methodology

2.1. The AKM Model

We follow AKM with an additive model of log-wages :

(1) $$y_{it} = \beta x_{it} + \theta_i + \psi_{j(i,t)} + u_{it}$$

Here y_{it} is the logarithm of the hourly wage of worker $i = 1, 2, \ldots, N$ during year $t = 1, \ldots, T$, demeaned by the average log-hourly wage for all workers during year t so that $\bar{y}_t = 0$[12]. Time-varying covariates x_{it} are limited to age and age squared. θ_i is the fixed effect of individual worker i, and ψ_j is the fixed effect of individual firm $j = 1, 2 \ldots, J$, both supposed constant in time for the duration of the panel, firm $j(i, t)$ being the employer of worker i during year t. u_{it} is the idiosyncratic error term. We further note $\mathbf{F} = (\mathbf{1}_{j=j(i,t)})$ the $N^* \times J$ matrix of the bipartite graph of workers/firms connections through time, with $N^* = NT$.

This model is built on two notable hypotheses. Firstly, it assumes no interaction effect between firm type and worker type, meaning that the fixed effects are additive in the log-wage. This assumption posits that the firm-specific wage premium will be the same for all workers regardless of their characteristics, such as gender, age, or skill level. Secondly, the model assumes exogenous mobility, where the residual term u_{it} has a null expectation conditional on the variables x_{it}, i, t, and j, as is traditionally assumed. Additionally, it is conditioned on the matrix $\mathbf{F}$. This hypothesis implies that wages before or after a job change, on average, remain the same as if there had been no job change.

Although both hypotheses can appear unrealistic and have been subjected to scrutiny, they seem to provide reasonable approximations. Bonhomme, Lamadon, and Manresa (2019) build a model that allows for interaction and finds only slight departures to the additive linear model. Di Addario et al. (2023) develop an extension of the two-way fixed effects model à la AKM with two firm fixed effects. They add to the fixed effect for the "destination" firm hiring the worker (the classic firm fixed effect) a fixed effect

[12] We also checked with a model inspired by Card et al. (2018), where log-wages are not centered, years are right-hand explanatory variables and age is included as a cubic polynomial constrained to be flat at 40 to avoid collinearity. Results in Table A7 are generally robust to this change in specification but show a lower rise in sorting. Figures A7A and A8A suggest this specification is more sensitive to firm fixed effect changes during the great recession. This alternate specification addresses the same classic underlying problem that age, generation and year effects are not separately identifiable without an additional source of variation or constraining hypothesis.

for the “origin” firm, reflecting the wage level necessary to “poach" a worker from a given firm. They found that destination effects are more than thirteen times as variable as origin effects across firms, implying that a more dynamic specification does not increase much the model explanatory power. Another potential specification error results from the possible evolution of “fixed" effects. Firm premium might change with time, and workers’ age profiles are heterogeneous, as shown in the French case in Magnac and Roux (2021) on BTS data. Formally, this simply contradicts the hypothesis of a null conditional expectation for the residual term. Regarding year-to-year variations in firm premium at least, Engbom, Moser, and Sauermann (2023) and Lachowska et al. (2023) provide some reassurance: firm premium are mostly stable in time, even if their variance appears somewhat procyclic. We analyze this source of heterogeneity in Annex E.2 and find signs of procyclicality[13]. Still, we find strong evolutions in age effects variance and covariances, that are difficult to interpret. We cannot separately identify age, generation and year effects, and cannot exclude that specification errors interact with real changes, such as changes in firm premium with time, age composition evolution of the population of workers (linked for instance to retirement age reforms), and generational rise in human capital.

2.2. Log-wage Variance Decomposition

Following Card, Heining, and Kline (2013) and Song et al. (2019), we take $V(y) = Var(y_{it})$ as a measure of wage inequalities and observe its evolution through 3 six-year periods: 2002-2007, 2008-2013 and 2014-2019. Like them, we also assume the AKM model to correctly describe wages and use it as the basis for variance decomposition. Ignoring for simplicity of exposition the time-varying workers variables x_{it}, we can describe for each period a decomposition of $V(y)$ as a sum of the variances of θ, ψ, u, and their respective covariances, estimated over all worker-years observations:

$$V(y) = V(\theta) + V(\psi) + V(u) + 2Cov(\theta, \psi) \tag{2}$$

Song et al. further distinguish within-firms and between-firms components of wage variance, and extend the law of total variance $V(y) = E[V(y|j)] + V[E(y|j)]$ to:

[13] More generally, de Chaisemartin and d’Haultfoeuille (2020) and de Chaisemartin and D’Haultfoeuille (2023) raised an alarm about the consequences of heterogeneity in panel treatment effects or, in our case, fixed effects. The problem does apply here if the AKM model is ill-specified, and we are not aware of any existing solution in this setting.

$$(3) \qquad V(y) = \underbrace{V(\bar{y}_j)}_{\text{Between-firm component}} + \underbrace{\sum_j m_j \times V(y_i | i \in j)}_{\text{Within-firm component}}$$

$$(4) \qquad V(y) = \underbrace{V(\psi) + 2Cov(\bar{\theta}_j, \psi) + V(\bar{\theta}_j)}_{\text{Between-firm component}} + \underbrace{V(\theta_i - \bar{\theta}_j) + V(u)}_{\text{Within-firm component}}$$

With $\bar{y}_j = \bar{y}_{j(i,t)}$ and $\bar{\theta}_j = \bar{\theta}_{j(i,t)}$ the respective expectations on i, t in firm j. By hypothesis, the analogous $\bar{u}_j$ is equal to 0. All moments of the distribution of firm variables are weighted by the share m_j of each firm in the total number of observations either directly as in equation 3 or implicitely when when computing variance over all (i, t) observations as in equation 4. Our interest lies first with the evolution of the sorting component of this decomposition, $2Cov(\theta, \phi)$, which is by construction entirely contained in the between-firm component of wage variance. Following Song et al. (2019), we define segregation in this context as $V(\bar{\theta}_j)$. Segregation captures the extent to which high-wage workers tend to work with one another, and low-wage workers with one another. Like other dimensions of segregation (residential, school, etc.), this has a strong impact on social mixing, but no direct impact on overall wage inequality, since the increased between-firm variance stems from a decrease in within-firm variance, leaving the overall distribution unchanged[14].

2.3. Limited Mobility Bias

The so-called “limited mobility bias" could also be seen as an incidental parameter bias, impacting the inference if we were interested in estimating only a few parameters (such as the effect on age), or as overfitting in a statistical learning context. It is similar to the classical bias on variance estimand that one corrects with a finite sample variance correction (i.e. the $N/(N-1)$ factor). The importance of the bias in our context stems from three elements: the very high number of parameters to be estimated with a low number of observations each, our focus on quadratic transformations of these noisy estimates (variance and covariance), and the sparse, centralized and clustered network structure of the design matrix that can be close to collinearity. This last aspect is due to

[14] We also compute Song et al. (2019)’s “Segregation Index" as $Var(\bar{\theta}_j)/Var(\theta_i)$ when possible.

the network of workers' mobility and explains the name of the bias.

More intuitively, if two large groups of firms are connected only through a handful of mobile workers, the identification of the relative premiums of the two groups will rest on these few observations. The parameters' estimates are not biased, but the high estimation errors translate into a bias in our variance terms: individual and, especially, firm effects variance is overestimated, while their covariance, which measures sorting, is underestimated.

Several correction strategies have been put forward in the literature. Andrews et al. (2008) directly correct estimates with a bias correction factor derived from an estimate of the error term variance, with a hypothesis of homoscedasticity that is unrealistic because of the networked nature of the estimation error (Jochmans and Weidner 2019). Borovičková and Shimer (2017) model heterogeneity as random effects rather than fixed effects and find a much higher sorting than previous estimates, but while fixed effects models allow for further study of the distribution of this heterogeneity, it is much more difficult with random effects models. Bonhomme, Lamadon, and Manresa (2019) cluster firms based on the distance between their wage distributions, then estimate a wage model where workers' effects are treated as random effects. The clustering creates a dense mobility network with many observations per cluster, that allows for the estimation of richer models, including interaction and dynamic terms, at the price of the additional hypothesis that clusters are correctly identified. Kline, Saggio, and Sølvsten (2020) leave-one-out strategy amounts to the bias correction factor method of Andrews et al. (2008) compatible with heteroscedasticity but is complex and computationally costly on large datasets. We use an analogous but much simpler split-sampling strategy by applying only one split to our data, rather than a leave-one-out method with as many splits as observations. The idea is that fixed effects are estimated independently in each split so that covariance between firm effects estimated in one sample and workers effects estimated in the other is a debiased measure of sorting. Chanut (2018) introduces split-sampling in an identical setting with French data and shows, on a toy example, that it corrects the bias. Other authors in similar settings implement split sampling either for instrumental variable estimation or to compare various groups of workers, typically without explicitly mentioning that it corrects the limited mobility bias: Drenik et al. (2023), Goldschmidt and Schmieder (2017), Gerard et al. (2021), Sorkin (2018) and Schoefer and Ziv (2022). In 2.3.1 we add to these works by generalizing the idea and providing proof that, under reasonable hypothesis analogous to Kline, Saggio, and Sølvsten, split-sampling corrects the limited mobility bias on quadratic terms. We also implement Bonhomme, Lamadon, and Manresa (2019) cluster method in 2.3.2 (without random effects) and find the results coherent with split sampling.

Formally, we follow Kline, Saggio, and Sølvsten (2020) for the description of this bias. They provide a simple framework that neatly generalizes to any quadratic form of the estimated parameters. We start with a simplified notation of our linear model with only one parameter vector and one observation index i (replacing the (i, t) couple in the detailed model):

$$y_i = z_i'\alpha + u_i \tag{5}$$

With $\alpha = (\beta, \theta, \psi)$ our parameter vector of length $k = 2 + N + J$ and z_i the non-random regressors vector of the (worker * year) i's observation characteristics, including the indicator vector for worker and firm. We note $S_{zz} = \sum_{i=1}^{N^*} z_i z_i'$ the design matrix (with full rank when we limit the sample to the main connected set). Our objects of interest are (weighted) variances and covariances of parts of the α vector and can be described as quadratic forms $\omega = \alpha' A \alpha$ for a chosen symmetric matrix $A \in \mathbf{R}^{k \times k}$. We can choose A so as to compute the quantities studied here, weighted by the number of worker-year observations[15].

Our naive plug-in estimator for ω is thus $\hat{\omega}^{PI} = \hat{\alpha}' A \hat{\alpha}$ with $\hat{\alpha}$ an OLS estimate $\hat{\alpha} = {S_{zz}}^{-1} \sum_{i=1}^{N^*} z_i y_i = \alpha + {S_{zz}}^{-1} \sum_{i=1}^{N^*} z_i u_i$. $\hat{\alpha}$ contains $\hat{\theta}$ and $\hat{\psi}$ which we later name *WFE* (for workers fixed effects) and *FFE* (for firms fixed effects). The estimation error in $\hat{\alpha}$ will result in a systematic bias in $\hat{\omega}^{PI}$ equal to a linear combination of the unknown and possibly heteroscedastic variances σ_i^2 of the error terms u_i. From classic results on quadratic forms, Kline, Saggio, and Sølvsten deduce :

$$\mathbf{E}[\hat{\omega}^{PI}] - \omega = trace(A\mathbf{V}[\hat{\alpha}]) = \sum_{i=1}^{N^*} B_{ii} \sigma_i^2 \tag{6}$$

With $B_{ii} = z_i' S_{zz}^{-1} A S_{zz}^{-1} z_i$ representing the influence of each (squared) error term on the plug-in estimator. This bias exists for all linear models, but usually for a small parameter dimension k the S^{-2} term insures relatively fast convergence. Here however

[15] For instance, var(WFE) is computed with a matrix A filled with 0 except for the $N \times N$ square corresponding to the N WFE estimates of the parameter vector, which we fill with a generic term $-1/N$ and a diagonal term $1 - 1/N$. Covariances objects are built with analogous matrices, with for instance a $1 - m_j$ term for the $(2 + i, 2 + N + j)$ and $(2 + N + j, 2 + i)$ positions in the matrix if the worker i is in firm j and a generic $-m_j$ term if the worker i is not working in firm j.

k is large, and so is, potentially, B_{ii}. Moreover, the complex structure of the design matrix, reflecting the complex network of worker / firm connections, is present both in matrices S and A when computing $cov(\theta, \psi)$, leaving way to even stronger bias.

This expression for the bias leads Kline, Saggio, and Sølvsten to an obvious correction strategy: estimating the σ_i^2 error terms, and thus the bias itself. This can be done with a leave-one-out strategy that is computationally costly, adding a factor of the order $N^* = NT$ to the computation. Kline, Saggio, and Sølvsten provide a more tractable estimation method through a high number of random projections (in the hundreds). Bonhomme et al. (2023) still finds the method demanding and further approximates it, though they worry the succession of approximate estimations (with those usual in AKM models) might have consequences that are not well understood.

For purely computational reasons, we favor a split-sampling strategy that only demands two estimations on two half-samples, at worst doubling computing time[16].

2.3.1. Split-sampling Bias Correction

With $\hat{\omega}^{SP}$ the split-sampling estimate of any quadratic form ω of the parameters, we show in Annex D that the bias is

$$\mathbf{E}[\hat{\omega}^{SP}] - \omega = trace(AS_{zz,1}^{-1}\mathbf{E}(B)(S_{zz,0}^{-1})') \qquad (7)$$

With the indices 1 and 0 indexing the two split samples I_1 and I_0 and B the matrix with the generic term:

$$b_{lm} = \sum_{i \in I_1}^{N_1} u_i z_{l,i} \sum_{j \in I_0}^{N_0} u_j z_{m,j}$$

This term has null expectation under mild conditions identical to Kline, Saggio, and Sølvsten: 1. null conditional expectation $\mathbf{E}[u|z] = 0$ and 2. independence of $u_i, i \in I_1$ and $u_j, j \in I_0$. If the variance-covariance matrix of u is diagonal, the bias disappears whatever the matrices A and S_i might be. The second condition might be violated if for instance u_i are correlated for different years of the same employer / employee pair, which is likely. We do not derive here the complete conditions for consistency, nor the

[16] We also leveraged the lower computation time of the R package *fixest* (Bergé 2018), up to ten times faster than the R package *lfe* (Gaure 2013) or the Stata package *reghdfe* (Correia 2016), although we used all three packages depending on the setting.

convergence rate when the number of parameters rises with the number of observations. We follow in this most of the AKM decomposition literature, which implicitly relies on the big size of the data used. However, consistency and convergence depend on the variance of error terms but also on relations between the quadratic form matrix A and the design matrix S, hence on the mobility network. We refer to the proofs and discussion in Kline, Saggio, and Sølvsten in the different but related leave-one-out context.

Of course, there is an additional cost of split-sampling in increased uncertainty coming from the reduced effective sample size ($S_{zz,s}$ has half the observations of S_{zz}). With our data, we observe that this uncertainty is small compared to the size of the bias reduction effect. One can reduce this uncertainty by repeatedly estimating the quadratic form through split-sampling and averaging the results. The procedure reaches arbitrary precision, only limited by the computational cost. We also report Monte Carlo experiment results and standard deviations computed on multiple random splits in Annex D, which confirms the stability of the procedure and offers some reassurance about consistency and convergence rate.

The main limitation of split sampling is the impact of the split on the bipartite graph and its main connected set, and the sample-splitting strategy has to be considered in this regard. In each split sample, the main connected set is smaller than in the original sample and both are distinct so that the common sample of workers and firms belonging to the main connected set in both split samples is reduced and so is the corresponding parameter vector of individual effects.

The most simple split strategy is a direct random split of observations in two equally sized samples. By balancing the sampling by workers, and splitting for each worker the periods of observation, one increases the odds that each worker is present in both samples' main connected set. We dub this method "period splitting". On the contrary, by splitting individuals rather than observations, one increases the connectivity in each set (because individual careers are kept intact), but each worker's fixed effect is estimated only once: one loses the capacity to correct the $var(\theta)$ and $var(u)$ quadratic forms through split sampling. If this splitting of individuals is balanced by firm, it increases the odds that each firm's fixed effect is estimated in each sample. We dub this method "firm splitting"[17]. With firm splitting, each firm with two workers or more is present in both samples and belongs to each main connected set if it remains connected with each random half of its employees.

[17]Chanut (2018) describes an algorithm to compute this split.

Period splitting might not completely correct the limited mobility bias for the reason mentioned above: it is likely that u_i are correlated for several observations of the same employer / employee pair. This problem is attenuated by the specifics of our setting. Because we keep only observations with full-year jobs, movers are generally observed only five years or less among the six in the panel. It is less likely that after the random split of these five observations or less, an individual worker would remain a mover in both samples. Because the estimation relies exclusively on movers, a given individual residuals would generally not be correlated to errors on both sides of the split. Still, by keeping all observations of one individual on one side of the split, the firm splitting method avoids entirely this drawback. Consequently, we favor firm splitting to compute a debiased sorting effect, but use period splitting to compute the complete variance decomposition.

2.3.2. Firm Clustering

We also implemented Bonhomme, Lamadon, and Manresa (2019) strategy (Table A6). We ran a firm-clustering algorithm with 1000 clusters (explaining around 90% of the between-firm dispersion in earnings for each period) before estimating AKM on firm clusters (rather than individual firms), with the hypothesis that firms' fixed effects are discretely distributed with a small number of values. The mobility network between clusters is very dense and each cluster's fixed effect estimate has very low variance, thus correcting the limited mobility bias. The clustering algorithm is a kmeans clustering based on quantiles of the wage distribution, as the identification of clusters can not rely on firm mean wage and must use higher moments of the distribution of wages[18]. Even so, it remains plausible that the segregation of workers could bias the clustering, with firms being clustered based on some combination of their own fixed effects and their average workers' fixed effects. An AKM estimation following this procedure would then show higher sorting, and lower cluster effect variance than is really the case. Bonhomme, Lamadon, and Manresa (2019) acknowledge the risk, mention job-market models that satisfy the conditions for cluster identification,[19] and provide in-depth robustness analysis that suggests it is of limited impact in practice. More generally,

[18]Following the original Bonhomme, Lamadon, and Manresa (2019) specification, we do not add additional firm variables to feed the clustering algorithm. Such developments are possible.

[19]"In some environments without firm capacity constraints, such as Postel-Vinay and Robin (2002), the upper bound of earnings in the firm is increasing in firm productivity, so firm-specific distributions are all different and firms may be consistently classified based on their earnings distributions. It is difficult to obtain similar guarantees in models with capacity constraints" (p. 217).

Bonhomme, Lamadon, and Manresa (2022) provides theoretical conditions and convergence results for the *two-step grouped fixed-effects* (GFE) method, where the clusters are viewed as an approximation of an underlying continuous unobserved heterogeneity. The relevant condition here is that there is an injective relationship between this unobserved heterogeneity and the moments of the distributions of the observed variables used in the k-means algorithm.

Compared to split-sampling, firm-clustering comes with this additional hypothesis and the added approximation and computing cost from the clustering step but allows for more versatile models and uses the whole data[20]. And while split-sampling is fast and only relies on AKM hypotheses, it further reduces the estimation sample to individuals or firms belonging to the main connected components in each split. We view the convergence of both methods as a strong argument in favor of the robustness of our results.

3. Results

3.1. A Robust Rise in Segregation and Sorting

To conduct the baseline variance decomposition, Table 2 applies two variance decompositions (Equations 2 and 4) to the estimates from AKM, corrected using firm split sampling. The correction involves splitting the data at the individual level and balancing it by firm (see Section 2.3.1). However, our baseline correction method does not allow for the computation of corrected and distinct estimates of $Var(u)$ and $Var(\theta)$. To provide a comprehensive analysis of their evolution, we rely on several alternative specifications, which are presented in Annex A.

As shown in Figure 1, the analysis reveals an increase in between-firm inequalities and a decrease in within-firm inequalities. The overall log-hourly wage variance slightly increased from 2002-2007 to 2014-2019[21]. France's quasi-stable wage inequalities during this period are atypical among developed countries. Nevertheless, the rise in between-firm wage inequalities aligns with the findings of Song et al. for the US and Card, Heining,

[20]Following BLM, we cluster firms using the empirical distribution of earnings of stayers only. As a consequence, only firms with at least one stayer in each period are selected. This explains the discrepancy in Table A6's last row and Table 1' overall sample person/yr observations.

[21]it is important to note a qualitative discrepancy between the overall trend and the specific yearly log-hourly wage variance as depicted in Figure 1 This discrepancy can be attributed to small methodological breaks in the original data, which were corrected in the graphical evidence but could not be accounted for in the AKM estimations.

TABLE 2. Decomposition of wage variance and its evolution - Split-sampling with firm split

		2002-2007		2008-2013		2014-2019		Change from 2002-2007 to 2014-2019	
		Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Total variance	Var(y)	0.211		0.207		0.219		0.008	
	Var (θ)	**	**	**	**	**	**	**	**
	Var (ψ)	0.014	6.7	0.014	6.8	0.013	5.8	-0.001	-19.3
	Var(Xb)	0.019	9	0.025	12.1	0.040	18.2	0.021	272.5
	Var(u)	**	**	**	**	**	**	**	**
	2*Cov($\theta.\psi$)	0.025	12	0.026	12.6	0.028	13.0	0.003	40.8
	2*Cov($\theta.Xb$)	-0.007	-3.3	-0.017	-8.2	-0.039	-17.8	-0.032	-419.6
	2*Cov($\psi.Xb$)	0.002	0.8	0.001	0.7	0.001	0.6	0.000	-5.8
Between-firm variance	Var($\bar{y}$)	0.088	41.5	0.094	45.3	0.103	46.9	0.015	197.7
	Var ($\bar{\theta}$)	0.041	19.6	0.049	23.6	0.058	26.5	0.017	218.7
	Var (ψ)	0.014	6.7	0.014	6.8	0.013	5.8	-0.001	-19.3
	Var($\bar{XB}$)	0.003	1.4	0.004	1.9	0.006	2.7	0.003	39.1
	2*Cov($\bar{\theta}.\psi$)	0.025	12.0	0.026	12.6	0.028	13.0	0.003	40.9
	2*Cov($\bar{\theta}.\bar{XB}$)	0.000	0.2	-0.002	-0.9	-0.005	-2.4	-0.006	-73.8
	2*Cov($\psi.\bar{XB}$)	0.002	0.8	0.001	0.7	0.001	0.6	0.000	-5.8
Within-firm variance	Var($y - \bar{y}$)	0.124	58.5	0.113	54.7	0.116	53.1	-0.007	-97.7
	Var ($\theta - \bar{\theta}$)	**	**	**	**	**	**	**	**
	Var($Xb - \bar{Xb}$)	0.016	7.6	0.021	10.3	0.034	15.5	0.018	233.3
	Var(u)	**	**	**	**	**	**	**	**
	2*Cov($\theta - \bar{\theta}.Xb - \bar{Xb}$)	-0.007	-3.4	-0.015	-7.2	-0.033	-15.2	-0.026	-340.4
	2*Cov($\theta - \bar{\theta}, u$)	**	**	**	**	**	**	**	**
	2*Cov($Xb - \bar{Xb}$,u)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.2
N* (firms in both largest connected sets)		51,932,308		54,530,553		52,921,242			

Note: "Comp." is the component of variance. "Share" is the share (in %) of total $var(y)$, or total change in $var(y)$ in the last column. All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Estimation on firms present in both main connected components in each split sample. Decomposition based on Equations 2, 3 and 4. The split-sampling method is described in Section 2.3.1.
** : These parameters' estimates are not corrected by firm-split

and Kline for Germany.

3.1.1. Decomposition of the Rise in Between-Firm Variance

The increase in between-firm variance can be decomposed into several components (Equation 4): the rise in $Var(\bar{\theta})$ reflecting segregation, the rise in $2Cov(\theta, \psi)$ capturing sorting, the decrease in firm premium variance $Var(\psi)$, and age effects. The dynamics of age effects are uncertain and difficult to interpret, as already mentioned. Focusing on the robust findings, it can be concluded that the growing gaps between firms are primarily driven by the rise in segregation and sorting. In 2002-2007, segregation accounted for 19.6% of total log-hourly wage variance, increasing to 26.5% in 2014-2019. Sorting represented 12% of inequalities in the first period and 13% in the last.

The rise in segregation is strong and consistent across different specifications. This finding complements previous research of Godechot et al. which shows a consistent increase in workplace segregation in 12 advanced capitalist economies (including France), with the top 10 and 1 percent earners increasingly isolated from the rest of the wage distribution. The increase in sorting, while comparatively limited in size, is also robust across various checks conducted.

3.1.2. Limited Mobility Bias and Correction

The rise in sorting can be observed even before correcting for the limited mobility bias or when restricting the analysis to firms with more than 20 observations per year. Figure 2 presents the results obtained from the principal corrections discussed in Sections 2.3.1 and 2.3.2[22].

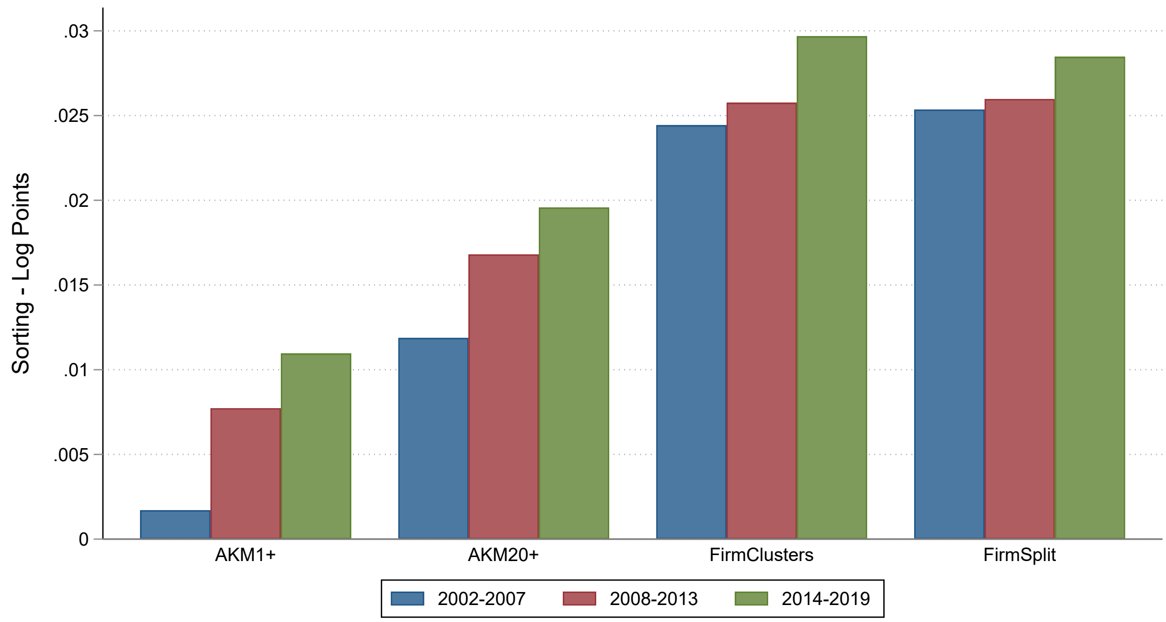
The measure of sorting is affected by a substantial limited mobility bias. While the AKM estimate attributes 1% to 5% of overall wage inequality to sorting, the baseline corrected estimates attribute 12% to 13%. The firm-clustering method produces results that closely align with the split sampling method and confirm the magnitude of the bias.

Importantly, the limited mobility bias is not neutral. The intensity of the rise in sorting is lower in the corrected estimates, suggesting an overall reduction in the limited mobility bias during the period, possibly due to increased mobility and connectivity in the firm-workers network. This reduction in bias is primarily concentrated in the smallest firms (see Table 3).

These findings highlight the importance of using corrected estimates to capture the dynamics of sorting accurately.

[22]see Tables A1, A2, 2, A5 and A6 for the full decomposition uncorrected and with the different corrections.

FIGURE 2. Sorting over time - Baseline and selected correction strategies



Notes: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. This figure reports the estimates of sorting - by period - coming from our baseline AKM estimate, an AKM estimate limited to firms with at least 20 employees, and the two bias correction strategies described in Sections 2.3.1 and 2.3.2.

TABLE 3. Sorting decomposition by firm size group and over time

		2002-2007				2014-2019			
		< 20	20 – 200	200 – 1000	> 1000	< 20	20 – 200	200 – 1000	> 1000
Split-Sampling	Overall	0.0254				0.0285			
	Between	0.0020				0.0023			
	Within	0.0233	0.0228	0.0238	0.0237	0.0245	0.0224	0.0262	0.0293
AKM	Overall	0.0017				0.0109			
	Between	0.0026				0.0027			
	Within	-0.0912	-0.0070	0.0205	0.0235	-0.0678	0.0025	0.0240	0.0296

Notes: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Split sampling with the firm split method. Firm size is computed on the observed workers employed by the firm in the panel. Sorting is computed as elsewhere as $2Cov(\theta, \psi)$, while between sorting is $2Cov(\bar{\theta}, \bar{\psi})$ with the average computed for each size group, and within is sorting computed separately for each size group.

3.1.3. Historical Trends

To provide further insights, we document the chronology of the rise in sorting and segregation from 1976 onward using long-term series (see Figure A1) and yearly decompositions from rolling panels in our main period of interest (see Figures A7A and A9A). Although long-term series are imperfect, they provide a clear indication of past trends.

The rise in sorting began in the mid-nineties, predating the period under study. The trend of increased segregation dates back even further, manifesting itself from the early eighties. The rise in sorting exhibits signs of pro-cyclicality in the most recent period, linked to the pro-cyclicality of the variance of firm premium. The variance of firm effects, measured for the same years on subsequent, overlapping panels, diminishes around 2008 and again in 2012. This decline likely reflects changes in firms' pay policies during and after the financial crisis and the Eurozone crisis (see Figure A8A). The decrease in premium variance explains the decline in estimated sorting for a given year when measured in subsequent, overlapping panels around the time of the financial crisis. In contrast, segregation has been steadily increasing without breaks around the financial crisis (see Figure A9A). These findings suggest that the rise in sorting resulting from pure matching and composition effects is likely even stronger than measured.

3.2. Explaining the Rise in Between-Firm Inequalities

In this section, we test which factors are behind the between-firm empirical trends we find. We move beyond variance and covariance and study other, non-quadratic distributional statistics of fixed effects[23]. For this reason, we use uncorrected estimates[24]. We examine several empirical arguments to sort through potential economic mechanisms. We find that the rise in segregation and sorting is mostly accounted for by composition effects in the population of firms and shifts in their occupational structure.

3.2.1. Firm Demography and Composition Effects

We get a first look at composition effects with a simple partition of all firms in the first and third periods into four categories. Fixed effects estimates are centered on 0 for each

[23]We also experimented with panel regressions of firm-level workers and firms' fixed effects on firm-level variables, followed by period-specific covariance decomposition. The method is complex and generates numerous, hard-to-interpret interaction terms. Even so, most of the rise in sorting was accounted for by changes in the size and population of firms with given fixed effects, rather than in firm-specific changes in fixed effects. We focus here on simpler results conveying the same message.

[24]Other correction strategies might be available in this context, such as Jochmans and Weidner (2022). The split-sampling method could also possibly be adapted. We leave these corrections for future research.

period. We partition firms depending on the sign of their estimated fixed effect (averaged over the two periods for firms present in both periods) and of the mean workers' fixed effect of their employees (similarly averaged). The partition distinguishes four "quadrants": covariance-increasing quadrants (high-wage firms with high-wage workers and low-wage firms with low-wage workers) and covariance-decreasing quadrants (low-wage firms with high-wage workers and high-wage firms with low-wage workers). Table 4 additionally differentiates between firms that are present in both periods and firms present either in 2002-2007 or in 2014-2019. The first column captures the number of workers where both average worker (W) and firm fixed effects (F) are above zero (F>0, W>0). Other columns map the rest of the possible combinations for F and W values.

TABLE 4. Composition effects: firm demographics and mean fixed effects by sorting quadrant

Firms present in both periods (2-period)												
	Person-Year Observations				**Mean Worker Fixed Effects**				**Mean Firm Fixed Effects**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	F>0, W>0	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*	*F>0, W>0*	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*	*F>0, W>0*	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*
2002-2007	14,003,660	8,892,099	13,122,069	5,748,544	0.1834	-0.1235	-0.1419	0.1528	0.1166	0.0727	-0.1028	-0.1103
2014-2019	15,379,346	9,252,809	15,582,805	6,931,931	0.2433	-0.1389	-0.1817	0.1171	0.1105	0.0732	-0.0824	-0.0951
Diff	1,375,686	360,710	2,460,736	1,183,387	0.0599	-0.0153	-0.0398	-0.0357	-0.0061	0.0005	0.0204	0.0152

Firms present only in one period (1-period)												
	Person-Year Observations				**Mean Worker Fixed Effects**				**Mean Firm Fixed Effects**			
	F>0, W>0	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*	*F>0, W>0*	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*	*F>0, W>0*	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*
2002-2007	3,609,701	4,516,577	5,585,022	3,188,645	0.2319	-0.2046	-0.1757	0.2022	0.1300	0.1222	-0.2065	-0.1179
2014-2019	2,612,327	2,636,805	4,357,807	2,796,457	0.2701	-0.2265	-0.1999	0.2138	0.1283	0.1205	-0.1960	-0.1236
Diff	-997,374	-1,879,772	-1,227,215	-392,188	0.0382	-0.0218	-0.0242	0.0116	-0.0017	-0.0017	0.0105	-0.0057

Notes: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. The largest connected set firms present in both periods and present either in 2002-2007 or in 2014-2019 are analyzed separately. Firms are further divided into four quadrants according to the value of the estimated worker (W) and firm fixed effects (F) (see Equation 1), averaged over both periods for staying firms. Worker and firm fixed effects are normalized so they are comparable across periods. Columns 1, 5, and 9 refer to the number of workers where both worker (W) and firm fixed effects (F) are above zero (F>0, W>0). The other columns map the rest of the possible combinations for F and W values. For firms present in both periods, the (not employment-weighted) 2-period average of firm and worker fixed effects are considered in the allocation to the four quadrants. Entries in columns 5-8 and 9-12 are the average worker and firm fixed effects, respectively, by quadrant.

Stayers firms grew in size,[25] remarkably so in the areas where both FEs are both positive and negative. In other words, even if fixed effects in every firm were the same

[25]by construction, the number of 2-period firms is the same in both periods.

in both periods, sorting would have increased simply because the size of stayers firms increased more in the sorting groups than in the non-sorting groups[26]. Notably, the groups of firms that experienced significant growth in size were already characterized by relatively higher employment (Table A9). This observation is further supported by the split-sampling findings in Table 3.

1-period firms in 2014-2019 have fewer person-year observations than those in 2002-2007. Firm age is likely to play a role here: 1-period firms in the 2014-2019 period are young and small, whereas 1-period firms in 2002-2017 are larger but shrinking before facing dissolution. Indeed, 1-period firms in 2014-2019 are on average smaller but more numerous, remarkably so in the low-F (F<0) quadrants (Table A9).

The evolution of the distribution of jobs over the four firms' quadrants can be summarized as follows: low-premium firms grew in share of employment, especially low-W ones, increasing sorting. Meanwhile, the drop in high-premium firms entirely occurred among low-W ones, especially one-period firms, again increasing sorting.

Table 4 also uses the four quadrants to describe the evolution of fixed effects for a given group of firms between the periods. The change in the mean workers' fixed effects for a firm between 2002 and 2019 might come from workers' turnover, or from a change in the WFE estimation of a given worker in two different panels, like the change in FFE for a given firm. Columns 5 to 12 in Table 4 show a notable increase in the average worker fixed effects in positive-sorting, positive-FEs stayer-firms while a decrease in the positive-sorting, negative-FEs stayer-firms. On the contrary, firm effects demonstrate remarkable stability in high-F (F>0) firms and exhibit a slight upward trend in low-F (F<0) firms.

Table 5 presents simple descriptive data that help to investigate the role of occupations. We use the fixed effect quadrants already mentioned and map the evolution of the occupational structure for the two groups with the most important growth in size: surviving firms with either high-premium and high-wage workers or low-premium and low-wage workers. The color gradient allows for a visual representation of the magnitude and direction of the changes observed, highlighting the areas of significant shifts in the data: darker shades of red represent more negative changes and darker shades of green represent more positive changes. The group of high-wage, stayer firms have more managers and engineers in period 3 than in period 1, and fewer blue-collar workers and associate professionals. Low-wage, stayer firms, instead, have a very stable

[26] This is not due to a general change in size distribution. A significant shift in the size distribution over time could explain the rise in between-firm inequalities. Figure A2 plots the fraction of firms below a given size, by period (2002-2007 vs 2014-2019). The two distributions overlap almost perfectly.

occupational structure.

TABLE 5. Change in the occupational structure by type of firm

		2-period, F>0, W>0			2-period, F<0, W<0		
Occupation	CS	2002-2007	2014-2019	Diff	2002-2007	2014-2019	Diff
Entrepreneurs	20	0.0050	0.0055	0.0005	0.0080	0.0059	-0.0022
Professionals	32	0.0076	0.0116	0.0040	0.0076	0.0196	0.0119
Artists and media professionals	35	0.0087	0.0093	0.0006	0.0008	0.0011	0.0003
Managers	37	0.1634	0.2163	0.0528	0.0356	0.0385	0.0028
Engineers	38	0.1460	0.2084	0.0624	0.0145	0.0169	0.0024
Primary school teachers	42	0.0041	0.0047	0.0006	0.0106	0.0186	0.0080
Health and social workers	43	0.0062	0.0075	0.0013	0.0663	0.0742	0.0080
Public administration intermediates	45	0.0005	0.0010	0.0005	0.0001	0.0003	0.0002
Business administration intermediates	46	0.1572	0.1229	-0.0343	0.0541	0.0576	0.0035
Technicians	47	0.0979	0.1001	0.0022	0.0201	0.0248	0.0047
Intermediate supervisors	48	0.0444	0.0328	-0.0116	0.0233	0.0200	-0.0033
Public administration clerks	52	0.0029	0.0033	0.0004	0.0631	0.0756	0.0126
Security agents	53	0.0030	0.0023	-0.0007	0.0174	0.0196	0.0022
Business administration clerks	54	0.1151	0.1031	-0.0120	0.0860	0.0827	-0.0033
Retail salespersons	55	0.0100	0.0123	0.0023	0.1317	0.1242	-0.0075
Personal service employees	56	0.0053	0.0049	-0.0004	0.0999	0.0857	-0.0142
Skilled manufacturing workers	62	0.1335	0.0843	-0.0492	0.0870	0.0650	-0.0220
Skilled artisans	63	0.0191	0.0189	-0.0002	0.0571	0.0565	-0.0006
Drivers	64	0.0059	0.0044	-0.0014	0.0740	0.0785	0.0045
Handling, transport skilled workers	65	0.0220	0.0189	-0.0031	0.0281	0.0261	-0.0019
Unskilled manufacturing workers	67	0.0388	0.0231	-0.0157	0.0565	0.0472	-0.0093
Unskilled artisans	68	0.0028	0.0040	0.0012	0.0544	0.0559	0.0015
Farm workers	69	0.0004	0.0004	0.0000	0.0038	0.0056	0.0018

Notes: All firms and individuals in firms with at least 2 employees are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Firms are divided into types as described in Table 4. We consider only the types that experienced a remarkable change in the person-year observations across time (see Table 4). We report the incidence of a certain 2-digits occupation in 2002-2007 vs 2014-2019, and the difference (in percentage points), by type. The color code in the visualization indicates the intensity of the change, with darker shades of red representing more negative changes and darker shades of green representing more positive changes. The second column displays the two-digit code corresponding to the *catégories socioprofessionnelles*, which can be further explored in detail here.

We extend our analysis by replicating the same investigation on 1-period firms, as presented in Table A10. By considering these additional findings, we uncover several significant trends associated with de-industrialization, the rise of low-end services, labor market polarization, and privatization of public services. In the low-W (W<0), 1-period firms quadrants, firms with a high share of manufacturing workers have been replaced by firms with more personal services employees. Furthermore, we observe a decrease in business administration occupations within the high-W quadrants, while there is an increase in the presence of managers and engineers within the same quadrants. Additionally, there is observable growth in the number of health and education workers employed by low-F, 1-period firms.

This descriptive evidence hints towards a structural evolution in the division of work between firms. Goldschmidt and Schmieder (2017) show for Germany that the rise in sorting is explained by the outsourcing of low productivity activities. Bilal and Lhuillier (2021)[27] find evidence for France that contractor firms pay lower wages and hire less from employment, while firms that outsource are on average more productive. Further, they document that a decline in the cost of outsourcing leads firms to scale up and increase labor demand. The structural estimation of a framework that rationalizes these findings highlights the link between outsourcing and labor market sorting. Additionally, Godechot et al. (2023) demonstrate that restructuring events in France, such as outsourcing, layoffs, offshoring, and subcontracting, contribute to increased segregation. Specifically, workplaces implementing these restructuring actions show a concentration of the top 10 percent of earners.

Our composition results are very compatible with such mechanisms, and more generally with any dynamics of firm specialization taking place through firm growth, creation, and destruction, and shifts in labor demand rather than changes in firm premiums.

3.2.2. Potential Alternative Channels

The economic literature has dealt extensively with skill-biased technological change (Acemoglu and Autor 2011). A rise in skill premium could explain the increase in sorting and segregation we document. Indeed, worker fixed effects dispersion Var(θ) diverges significantly during the study period (Annex A), entirely driven by its between-firm com-

[27] Both papers define contractor firms and outsourcing events from detailed firms' industry, occupations, and joint mobility of clusters of workers. Bilal and Lhuillier (2021) additionally use a firm survey (EAE) on intermediate inputs to measure expenditure on external workers.

ponent Var($\bar{\theta}$). However, an increase in the skill premium would likely impact the link between occupations and worker fixed effects[28]. The relationship between occupations and worker fixed effects is very stable between the periods and does not show evidence of an increase in occupational effects variance. [29] However, occupations associated with higher or lower worker fixed effects grew more than occupations with average fixed effect. Overall, the findings suggest that the increase in worker unobserved heterogeneity dispersion is not about a change in skill returns but rather a transformation in the occupational structure associated with increased occupational segregation in specific categories of firms.

Firm pay premium's variance has been slightly decreasing[30]. To further investigate this trend, we analyze the employment-weighted relationships between value added per worker and estimated firm fixed effects at the firm level across different periods, as illustrated in Figure A3. Our findings reveal that, for a given level of period-demeaned productivity, the distribution of shared rents exhibits a slight increase at the bottom during the most recent period compared to the previous one, while converging towards the top. These results are consistent with the evidence presented in columns 9 to 12 of Table 4. This suggests that the overall decrease in the variance of firm effects can be attributed to an increase in rents at the bottom end of the firm rent distribution. However, the period fitting lines' slopes are visually very similar and statistically non-different from each other. We conclude that the decrease in the variance of firm effects, and the relatively small increases in rents at the bottom end of the distribution do not significantly contribute to explaining the observed increase in sorting. On the contrary, their influence might be in the other direction.

3.3. Explaining the Decline in Within-Firm Inequalities

To investigate the decrease in within-firm inequalities, we examine the average change in earnings for employees across different percentiles, ranging from the top 1%-paid employee down to the employees in the first percentile (Figure A4). This simple statistic provides valuable insights into the dynamics of wage disparities within firms over time. In contrast to findings in the United States, particularly in mega-firms (Song

[28]to be distinguished from the evolution of the distribution of occupations in firms, discussed in the previous section.

[29]Table A13 in Annex F gives the results.

[30]This decline is more pronounced without bias correction, which corroborates the idea that the bias decreased with time. The decline of premium variance is compatible with an "eclipse of rent-sharing" as recently documented by Acemoglu, He, and le Maire (2022) for Denmark and the US.

et al. 2019), our analysis reveals that wages at the bottom percentile in France have not stagnated. Instead, they have experienced a higher growth rate compared to the rest of the distribution across almost all firm-size classes. Figure A5 shows how this pace of growth is strongly correlated with hourly minimum wage growth, which experienced a steep increase in the first year of our panel following the need for harmonization after the 35-hours-a-week Reform. This suggestive evidence points towards a potential role for labor market institutions in the French context. The link between P1 and hourly minimum wage is partly mechanical though for two main reasons. First, it is mandatory to pay at least the SMIC in most sectors and for most employees. Second, our population selection is partially linked to the hourly minimum wage as explained in Section 1.2.

On a similar note, Bozio, Breda, and Guillot (2020) shows that since the 1970's the redistributive effects of payroll taxation have regularly increased. Considered before tax, labour cost inequalities have increased in France at a comparable rate as in other countries.

Kramarz et al. (2022) concludes as well that the early 2000s' significant increase in the minimum wage—followed by the workweek's reduction to 35 hours and the elimination of all employer-paid payroll taxes around the minimum wage—translates into a significant rise in the bottom percentiles of the earnings distribution in the 2000s. They observe that this mix of policies was particularly beneficial for women, young people, and workers in rural and remote municipalities.

Finally, Bilal and Lhuillier (2021) show that in the context of their model, an increase in the minimum wage can mitigate earnings losses arising from substantial worker reallocation across firms.

4. Discussion

Despite the stable log-wage variance in France throughout 2002-2019, our results demonstrate that both sorting and segregation increased, similar to the trends observed in the USA and Germany. The rise in between-firm inequalities in France is consistent with a structural evolution in the division of work between firms. Alternative explanations related to increasing returns to skills and changes in firms' rent-sharing behavior were not supported by our analysis. In parallel, wages at the bottom percentiles have experienced faster growth compared to the middle and top percentiles. This phenomenon is closely linked to the nature of French labor market institutions, which aligns with recent findings in the literature. France experienced the same mechanisms raising inequalities elsewhere, only to be contained by increasing redistribution.

The AKM model has provided valuable insights into wage inequalities. The model's estimation is subject to the well-known limitation of the limited mobility bias, which results in the underestimation of sorting. We show that the bias also affects the measurement of the evolution of sorting, likely because mobility intensity and patterns do evolve over time. These findings underscore the importance of employing corrected estimates to capture the precise dynamics of sorting.

There are additional substantive and methodological questions that warrant further investigation. On the substantive side, there is more to explore regarding the interplay between French institutional features and segregation and sorting. Wage inequalities in France have been effectively controlled, primarily through an increase in the redistributive power of payroll taxes and periodic increments in the minimum wage. These phenomena inevitably impact wage distribution and likely interact with firm pay policies, the job market, and sorting. While our focus has been on wages rather than labor costs, as it provides better insights into workers' incentives and observed inequalities, replicating our analysis on labor costs in addition to wages could enhance our understanding of the French exception and its underlying dynamics. Finally, the increased concentration of specific workers in particular firms raises questions about firm specialization, externalization, and job polarization, which merit further exploration.

Methodologically, there are limitations that require a deeper understanding, such as the non-fixed nature of fixed effects. The implications of this specification error are complex, but they could potentially affect the measurement of sorting and other components of the decomposition, particularly during periods of significant short-term economic fluctuations. Another question pertains to accurately measuring the age and experience components of wages when disentangling yearly effects, age, and individual effects. Fluctuations in the interactions between these terms suggest possible estimation artifacts that need to be addressed through alternative specifications.

References

Abowd, John M, Francis Kramarz, Paul Lengermann, and Sébastien Pérez-Duarte. 2004. "Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States." Unpublished Manuscript.

Abowd, John M, Francis Kramarz, and David N Margolis. 1999. "High wage workers and high wage firms." *Econometrica* 67 (2): 251–333.

Abowd, John M, Francis Kramarz, and Sebastien Roux. 2006. "Wages, mobility and firm performance: Advantages and insights from using matched worker–firm data." *The Economic Journal* 116 (512): F245–F285.

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

Acemoglu, Daron, Alex He, and Daniel le Maire. 2022. "Eclipse of Rent-Sharing: The Effects of Managers' Business Education on Wages and the Labor Share in the US and Denmark." NBER Working Paper No. 29874.

Andrews, Martyn J, Len Gill, Thorsten Schank, and Richard Upward. 2008. "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171 (3): 673–697.

Bergé, Laurent. 2018. "Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm." Crea Discussion Paper No. 13.

Bergeaud, Antonin, Clément Malgouyres, Clément Mazet-Sonilhac, and Sara Signorelli. 2021. "Technological Change and Domestic Outsourcing." IZA Discussion Paper No. 14603.

Bilal, Adrien, and Hugo Lhuillier. 2021. "Outsourcing, inequality and aggregate output." NBER Working Paper No. 29348.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019. "A distributional framework for matched employer employee data." *Econometrica* 87 (3): 699–739.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2022. "Discretizing unobserved heterogeneity." *Econometrica* 90 (2): 625–643.

Bonhomme, Stéphane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler. 2023. "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?." *Journal of Labor Economics* 41 (2): 291–322.

Borovičková, Katarína, and Robert Shimer. 2017. "High wage workers work for high wage firms." NBER Working Paper No. 24074.

Bozio, Antoine, Thomas Breda, and Malka Guillot. 2020. "The Contribution of Payroll Taxation to Wage Inequality in France." IZA Discussion Paper No. 13317.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and labor market inequality: Evidence and some theory." *Journal of Labor Economics* 36 (S1): S13–S70.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace heterogeneity and the rise of West German wage inequality." *The Quarterly journal of economics* 128 (3): 967–1015.

de Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review* 110 (9): 2964–96.

de Chaisemartin, Clément, and Xavier D'Haultfoeuille. 2023. "Two-way Fixed Effects and

Differences-in-Differences Estimators with Several Treatments." *Journal of Econometrics*.
Chanut, Nicolas. 2018. "Distinguishing Between Signal and Noise in the Measurement of the Firm Wage Premium." Working Paper.
Correia, Sergio. 2016. "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator." Working Paper.
Coudin, Elise, Sophie Maillard, and Maxime Tô. 2018. "Family, firms and the gender wage gap in France." IFS Working Papers W18/01.
Di Addario, Sabrina L, Patrick M Kline, Raffaele Saggio, and Mikkel Sølvsten. 2023. "It Ain't Where You're From, It's Where You're At: Hiring Origins, Firm Heterogeneity, and Wages.." *Journal of Econometrics* 233 (2): 340–374.
Dorn, David, Johannes Schmieder, and James Spletzer. 2018. "Domestic Outsourcing in the United States." Working Paper.
Drenik, Andres, Simon Jäger, Miguel Pascuel Plotkin, and Benjamin Schoefer. 2023. "Paying Outsourced Labor: Direct Evidence from Linked Temp Agency-Worker-Client Data." *The Review of Economics and Statistics* 105 (1): 206–216.
Engbom, Niklas, Christian Moser, and Jan Sauermann. 2023. "Firm pay dynamics." *Journal of Econometrics* 233 (2): 396–423.
Gaure, Simen. 2013. "lfe: Linear group fixed effects." *The R Journal* 5 (2): 104–117.
Gerard, François, Lorenzo Lagos, Edson Severnini, and David Card. 2021. "Assortative Matching or Exclusionary Hiring? The Impact of Employment and Pay Policies on Racial Wage Differences in Brazil." *American Economic Review* 111 (10): 3418–57.
Godechot, Olivier, Paula Apascaritei, István Boza, Lasse Folke Henriksen, Are Skeie Hermansen, Feng Hou, Naomi Kodama, Alena Křížková, Jiwook Jung, Marta M Elvira et al. 2020. "The great separation: Top earner segregation at work in high-income countries." MaxPo Discussion Paper.
Godechot, Olivier, Donald Tomaskovic-Devey, István Boza, Lasse Henriksen, Are Skeie Hermansen, Feng Hou, Jiwook Jung, Naomi Kodama, Alena Křížková, Zoltán Lippényi, Silvia Maja Melzer, Eunmi Mun, Halil Sabanci, Max Thaning et al. 2023. "The great separation (Reloaded): Top Earner Segregation at Work in Advanced Capitalist Economies." Working Paper.
Goldschmidt, Deborah, and Johannes F Schmieder. 2017. "The rise of domestic outsourcing and the evolution of the German wage structure." *The Quarterly Journal of Economics* 132 (3): 1165–1217.
Jochmans, Koen, and Martin Weidner. 2019. "Fixed-Effect Regressions on Network Data." *Econometrica* 87 (5): 1543–1560.
Jochmans, Koen, and Martin Weidner. 2022. "Inference on a distribution from noisy draws." *Econometric Theory*: 1–38.
Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-out estimation of variance components." *Econometrica* 88 (5): 1859–1898.
Kramarz, Francis, Elio Nimier-David, Thomas Delemotte et al. 2022. "Inequality and earnings dynamics in France: National policies and local consequences." *Quantitative Economics* 13 (4): 1527–1591.

Lachowska, Marta, Alexandre Mas, Raffaele Saggio, and Stephen A. Woodbury. 2023. "Do firm effects drift? Evidence from Washington administrative data." *Journal of Econometrics* 233 (2): 375–395.

Magnac, Thierry, and Sébastien Roux. 2021. "Heterogeneity and wage inequalities over the life cycle." *European Economic Review*: 103715.

OECD. 2021. *The Role of Firms in Wage Inequality: Policy Lessons from a Large Scale Cross-Country Study*. Éditions OCDE, Paris.

Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. "Equilibrium wage dispersion with worker and employer heterogeneity." *Econometrica* 70 (6): 2295–2350.

Schoefer, Benjamin, and Oren Ziv. 2022. "Productivity, Place, and Plants." *The Review of Economics and Statistics*: 1–46.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. 2019. "Firming up inequality." *The Quarterly journal of economics* 134 (1): 1–50.

Sorkin, Isaac. 2018. "Ranking Firms Using Revealed Preference." *The Quarterly Journal of Economics* 133 (3): 1331–1393.

Tomaskovic-Devey, Donald, Anthony Rainey, Dustin Avent-Holt, Nina Bandelj, István Boza, David Cort, Olivier Godechot, Gergely Hajdu, Martin Hällsten, Lasse Folke Henriksen et al. 2020. "Rising between-workplace inequalities in high-income countries." *Proceedings of the National Academy of Sciences* 117 (17): 9277–9283.

ANNEXES

Appendix A. Appendix Tables

TABLE A1. Decomposition of wage variance and its evolution - Firms with 1+ employees (uncorrected AKM)

		2002-2007		2008-2013		2014-2019		Change from 2002-2007 to 2014-2019	
		Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Total variance	Var(y)	0.210		0.204		0.216		0.006	
	Var (θ)	0.157	74.6	0.153	74.8	0.169	78.5	0.013	217.9
	Var (ψ)	0.028	13.4	0.024	11.9	0.023	10.7	-0.005	-86.4
	Var(Xb)	0.020	9.4	0.025	12.3	0.040	18.6	0.020	346.0
	Var(u)	0.010	4.7	0.010	4.7	0.010	4.8	0.000	7.3
	2*Cov($\theta.\psi$)	0.002	0.8	0.008	3.8	0.011	5.1	0.009	157.4
	2*Cov($\theta.Xb$)	-0.008	-3.8	-0.017	-8.3	-0.040	-18.4	-0.032	-539.0
	2*Cov($\psi.Xb$)	0.002	0.9	0.002	0.9	0.002	0.8	0.000	-3.1
Between-firm variance	Var($\bar{y}$)	0.089	42.2	0.094	46.1	0.103	48.0	0.015	252.3
	Var ($\bar{\theta}$)	0.054	25.6	0.058	28.5	0.067	31.1	0.013	224.9
	Var (ψ)	0.028	13.4	0.024	11.9	0.023	10.7	-0.005	-86.4
	Var($\bar{XB}$)	0.003	1.6	0.004	2.1	0.007	3.1	0.003	57.0
	2*Cov($\bar{\theta}.\psi$)	0.002	0.8	0.008	3.8	0.011	5.1	0.009	157.4
	2*Cov($\bar{\theta}.\bar{XB}$)	0.000	0.0	-0.002	-1.1	-0.006	-2.7	-0.006	-97.5
	2*Cov($\psi.\bar{XB}$)	0.002	0.9	0.002	0.9	0.002	0.8	0.000	-3.1
Within-firm variance	Var($y - \bar{y}$)	0.121	57.8	0.110	53.9	0.112	52.0	-0.009	-152.3
	Var ($\theta - \bar{\theta}$)	0.103	49	0.094	46.3	0.102	47.4	0.000	-6.9
	Var($Xb - \bar{Xb}$)	0.016	7.9	0.021	10.2	0.033	15.5	0.017	289.0
	Var(u)	0.010	4.7	0.010	4.7	0.010	4.8	0.000	7.3
	2*Cov($\theta - \bar{\theta}.Xb - \bar{Xb}$)	-0.008	-3.7	-0.015	-7.3	-0.034	-15.7	-0.026	-441.5
	2*Cov($\theta - \bar{\theta}, u$)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	-0.1
	2*Cov($Xb - \bar{Xb}, u$)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	-0.1
Segregation Index	$\frac{Var(\bar{\theta}_j)}{Var(\theta_i)}$	0.344		0.381		0.396		0.052	
N* (largest connected set)		58,666,316		61,413,372		59,550,288			

Notes: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Decomposition based on Equations 2, 3 and 4, weighted by worker-year observations. The largest connected set entails the group of firms connected by worker mobility.

TABLE A2. Decomposition of wage variance and its evolution - Firms with 20+ employees

		2002-2007		2008-2013		2014-2019		Change from 2002-2007 to 2014-2019	
		Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Total variance	Var(y)	0.209		0.205		0.217		0.008	
	Var (θ)	0.153	73.3	0.152	73.9	0.169	77.6	0.015	185.4
	Var (ψ)	0.021	10.1	0.018	9.0	0.017	8.0	-0.004	-45.5
	Var(Xb)	0.018	8.6	0.025	12.1	0.039	17.7	0.021	250.2
	Var(u)	0.010	4.6	0.009	4.6	0.010	4.7	0.001	8.2
	2*Cov($\theta.\psi$)	0.012	5.7	0.017	8.2	0.020	9.0	0.008	93.4
	2*Cov($\theta.Xb$)	-0.006	-3	-0.017	-8.4	-0.038	-17.4	-0.032	-384
	2*Cov($\psi.Xb$)	0.002	0.7	0.001	0.6	0.001	0.4	-0.001	-7.6
Between-firm variance	Var($\bar{y}$)	0.086	41.0	0.092	44.7	0.101	46.5	0.015	184.9
	Var ($\bar{\theta}$)	0.048	23.1	0.054	26.2	0.063	28.9	0.015	178
	Var (ψ)	0.021	10.1	0.018	9.0	0.017	8.0	-0.004	-45.5
	Var($\bar{XB}$)	0.003	1.3	0.004	1.8	0.006	2.5	0.003	34
	2*Cov($\bar{\theta}.\psi$)	0.012	5.7	0.017	8.2	0.020	9.0	0.008	93.4
	2*Cov($\bar{\theta}.\bar{XB}$)	0.000	0.2	-0.002	-1.0	-0.005	-2.4	-0.006	-67.5
	2*Cov($\psi.\bar{XB}$)	0.002	0.7	0.001	0.6	0.001	0.4	-0.001	-7.6
Within-firm variance	Var($y - \bar{y}$)	0.123	59.0	0.113	55.3	0.116	53.5	-0.007	-84.9
	Var ($\theta - \bar{\theta}$)	0.105	50.3	0.098	47.7	0.106	48.6	0.001	7.4
	Var($Xb - \bar{Xb}$)	0.015	7.3	0.021	10.3	0.033	15.2	0.018	216.2
	Var(u)	0.010	4.6	0.009	4.6	0.010	4.7	0.001	8.2
	2*Cov($\theta - \bar{\theta}.Xb - \bar{Xb}$)	-0.007	-3.1	-0.015	-7.4	-0.033	-15.0	-0.026	-316.6
	2*Cov($\theta - \bar{\theta}$,u)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0
	2*Cov($Xb - \bar{Xb}$,u)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0
Segregation Index	$\frac{Var(\bar{\theta}_j)}{Var(\theta_i)}$	0.315		0.354		0.373		0.052	
N* (largest connected set)		53,378,336		55,173,311		54,723,634			

Note: In each period, all firms and individuals in firms with at least 20 different employees over the period, and in the main connected component, are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Decomposition based on Equations 2, 3 and 4.

Table A3. Decomposition of wage variance and its evolution - Establishments with 20+ employees

		2002-2007		2008-2013		2014-2019		Change from 2002-2007 to 2014-2019	
		Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Total variance	Var(y)	0.212		0.209		0.22		0.009	
	Var (θ)	0.152	71.7	0.152	72.6	0.169	76.8	0.017	202.4
	Var (ψ)	0.027	12.6	0.023	11.1	0.022	9.8	-0.005	-59.4
	Var(Xb)	0.017	7.9	0.024	11.4	0.037	17.0	0.021	240.0
	Var(u)	0.009	4.3	0.009	4.4	0.01	4.5	0.001	8.7
	2*Cov($\theta.\psi$)	0.011	5.2	0.017	8.1	0.019	8.5	0.008	90.2
	2*Cov($\theta.Xb$)	-0.005	-2.5	-0.017	-8.1	-0.037	-16.9	-0.032	-371.3
	2*Cov($\psi.Xb$)	0.001	0.7	0.001	0.5	0.001	0.2	-0.001	-10.6
Between-firm variance	Var($\bar{y}$)	0.098	46.5	0.105	50.4	0.113	51.4	0.015	173.5
	Var ($\bar{\theta}$)	0.056	26.6	0.063	30.2	0.073	33.0	0.016	190.3
	Var (ψ)	0.027	12.6	0.023	11.1	0.022	9.8	-0.005	-59.4
	Var($\bar{XB}$)	0.003	1.3	0.004	1.9	0.006	2.8	0.003	39.2
	2*Cov($\bar{\theta}.\psi$)	0.011	5.2	0.017	8.1	0.019	8.5	0.008	90.2
	2*Cov($\bar{\theta}.\bar{XB}$)	0.000	0.1	-0.003	-1.3	-0.006	-2.9	-0.007	-76.2
	2*Cov($\psi.\bar{XB}$)	0.001	0.7	0.001	0.5	0.001	0.2	-0.001	-10.6
Within-firm variance	Var($y - \bar{y}$)	0.113	53.5	0.104	49.6	0.107	48.6	-0.006	-73.5
	Var ($\theta - \bar{\theta}$)	0.096	45.2	0.089	42.5	0.097	43.9	0.001	12.1
	Var($Xb - \bar{Xb}$)	0.014	6.6	0.02	9.5	0.031	14.1	0.017	200.8
	Var(u)	0.009	4.3	0.009	4.4	0.01	4.5	0.001	8.7
	2*Cov($\theta - \bar{\theta}.Xb - \bar{Xb}$)	-0.005	-2.6	-0.014	-6.8	-0.031	-14.0	-0.025	-295.1
	2*Cov($\theta - \bar{\theta}, u$)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.1
	2*Cov($Xb - \bar{Xb}, u$)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0
Segregation Index	$\frac{Var(\bar{\theta}_j)}{Var(\theta_i)}$	0.371		0.415		0.429		0.059	
N* (largest connected set)		49,378,243		50,278,039		49,994,151			

Note: All establishments and individuals in establishments with at least 20 employees are included. Only individuals employed for at least 360 days by the same establishment during the year are included for a given year. Individuals and establishments in public administration are not included. Decomposition based on Equations 2, 3 and 4.

TABLE A4. Decomposition of wage variance and its evolution - Split-sampling correction with firm split, with an approximation for missing estimates

		2002-2007		2008-2013		2014-2019		Change from 2002-2007 to 2014-2019	
		Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Total variance	Var(y)	0.211		0.207		0.219		0.008	
	** Var (θ)	0.142	67.2	0.143	69.1	0.159	72.8	0.017	227.8
	Var (ψ)	0.014	6.7	0.014	6.8	0.013	5.8	-0.001	-19.3
	Var(Xb)	0.019	9	0.025	12.1	0.040	18.2	0.021	272.5
	** Var(u)	0.016	7.5	0.014	7	0.016	7.4	0.000	4.3
	2*Cov(θ.ψ)	0.025	12	0.026	12.6	0.028	13.0	0.003	40.8
	2*Cov(θ.Xb)	-0.007	-3.3	-0.017	-8.2	-0.039	-17.8	-0.032	-419.6
	2*Cov(ψ.Xb)	0.002	0.8	0.001	0.7	0.001	0.6	0.000	-5.8
Between-firm variance	Var($\bar{y}$)	0.088	41.5	0.094	45.3	0.103	46.9	0.015	197.7
	Var ($\bar{\theta}$)	0.041	19.6	0.049	23.6	0.058	26.5	0.017	218.7
	Var (ψ)	0.014	6.7	0.014	6.8	0.013	5.8	-0.001	-19.3
	Var($\bar{XB}$)	0.003	1.4	0.004	1.9	0.006	2.7	0.003	39.1
	2*Cov($\bar{\theta}$.ψ)	0.025	12.0	0.026	12.6	0.028	13.0	0.003	40.9
	2*Cov($\bar{\theta}$.$\bar{XB}$)	0.000	0.2	-0.002	-0.9	-0.005	-2.4	-0.006	-73.8
	2*Cov(ψ.$\bar{XB}$)	0.002	0.8	0.001	0.7	0.001	0.6	0.000	-5.8
Within-firm variance	Var($y - \bar{y}$)	0.124	58.5	0.113	54.7	0.116	53.1	-0.007	-97.7
	** Var ($\theta - \bar{\theta}$)	0.101	47.6	0.094	45.5	0.101	46.3	0.001	9
	Var($Xb - \bar{Xb}$)	0.016	7.6	0.021	10.3	0.034	15.5	0.018	233.3
	** Var(u)	0.016	7.5	0.014	7	0.016	7.4	0.000	4.3
	2*Cov($\theta - \bar{\theta}$.$Xb - \bar{Xb}$)	-0.007	-3.4	-0.015	-7.2	-0.033	-15.2	-0.026	-340.4
	2*Cov($\theta - \bar{\theta}, u$) **	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0
	2*Cov($Xb - \bar{Xb}$,u)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.2
Segregation Index	** $\frac{Var(\bar{\theta}_j)}{Var(\theta_i)}$	0.291		0.341		0.364		0.073	
N* (largest connected set)		51,932,308		54,530,553		52,921,242			

Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Estimation on firms present in both main connected components in each split sample. Decomposition based on Equations 2, 3 and 4. The split-sampling method is described in Section 2.3.1.
** : These parameters' estimates are NOT directly corrected by firm split. $Var(u)$ is approximated from the within-worker variance of wages, net from age and firm effects estimated on the other split. $Var(\theta)$ and Var ($\theta - \bar{\theta}$) are then deduced by substraction, assuming $cov(u, \theta) = 0$

TABLE A5. Decomposition of wage variance and its evolution - Split-sampling correction with period split

		2002-2007		2008-2013		2014-2019		Change from 2002-2007 to 2014-2019	
		Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Total variance	Var(y)	0.212		0.209		0.221		0.009	
	Var (θ)	0.144	67.9	0.147	70.4	0.162	73.1	0.018	196.1
	Var (ψ)	0.014	6.5	0.013	6.4	0.011	5.1	-0.002	-27.5
	Var(Xb)	0.018	8.3	0.025	12.0	0.038	17.1	0.020	223.3
	Var(u)	**	**	**	**	**	**	**	**
	2*Cov(θ.ψ)	0.025	12.0	0.025	12.1	0.027	12.2	0.002	17.8
	2*Cov(θ.Xb)	-0.005	-2.5	-0.017	-8.1	-0.035	-16	-0.030	-332.4
	2*Cov(ψ.Xb)	0.001	0.7	0.001	0.6	0.001	0.4	-0.001	-7
Between-firm variance	Var($\bar{y}$)	0.087	41.3	0.094	44.9	0.102	46.3	0.015	165.3
	Var ($\bar{\theta}$)	0.044	20.5	0.052	25.1	0.062	28.3	0.019	209.4
	Var (ψ)	0.014	6.5	0.013	6.4	0.011	5.1	-0.002	-27.5
	Var($\bar{XB}$)	0.003	1.2	0.004	1.7	0.005	2.4	0.003	29.2
	2*Cov($\bar{\theta}$.ψ)	0.026	12.0	0.025	12.1	0.027	12.3	0.002	17.5
	2*Cov($\bar{\theta}$.$\bar{XB}$)	0.001	0.4	-0.002	-0.9	-0.005	-2.1	-0.005	-58.9
	2*Cov(ψ.$\bar{XB}$)	0.001	0.7	0.001	0.6	0.001	0.4	-0.001	-7
Within-firm variance	Var($y - \bar{y}$)	0.125	58.7	0.115	55.1	0.119	53.7	-0.006	-65.3
	Var ($\theta - \bar{\theta}$)	0.100	47.3	0.094	45.2	0.100	45.1	-0.001	-7.8
	Var($Xb - \bar{Xb}$)	0.015	7.1	0.021	10.3	0.033	14.7	0.018	194.1
	Var(u)	**	**	**	**	**	**	**	**
	2*Cov($\theta - \bar{\theta}$.$Xb - \bar{Xb}$)	-0.006	-2.7	-0.015	-7.1	-0.031	-13.8	-0.025	-273.6
	2*Cov($\theta - \bar{\theta}$,u)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0
	2*Cov($Xb - \bar{Xb}$,u)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0
Segregation Index	$\frac{Var(\bar{\theta}_j)}{Var(\theta_i)}$	0.302		0.356		0.387		0.084	
N* (largest connected set)		47,529,687		50,395,656		49,395,499			

Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Estimation on firms and, when necessary, individuals present in both main connected components in each split sample. Decomposition based on Equations 2, 3 and 4. The split-sampling method is described in Section 2.3.1.
** : These parameters' estimates are not corrected by period-split

Table A6. Decomposition of wage variance and its evolution - Firm Clustering

		2002-2007		2008-2013		2014-2019		Change from 2002-2007 to 2014-2019	
		Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Total variance	Var(y)	0.208		0.201		0.211		0.003	
	Var (θ)	0.156	74.7	0.152	75.2	0.165	77.8	0.009	280.7
	Var (ψ)	0.005	2.4	0.005	2.4	0.005	2.5	0.000	7.6
	Var(Xb)	0.021	9.9	0.024	12.1	0.038	18.1	0.018	553.6
	Var(u)	0.011	5.1	0.010	5.1	0.011	5.2	0.000	7.7
	2*Cov($\theta.\psi$)	0.024	11.7	0.026	12.8	0.030	14	0.005	164.6
	2*Cov($\theta.Xb$)	-0.010	-4.7	-0.017	-8.4	-0.039	-18.4	-0.029	-914.6
	2*Cov($\psi.Xb$)	0.002	0.8	0.001	0.7	0.002	0.8	0.000	0.4
Between-firm variance	Var($\bar{y}$)	0.078	37.6	0.082	40.9	0.093	44.2	0.015	472.6
	Var ($\bar{\theta}$)	0.042	20.4	0.046	22.8	0.052	24.7	0.010	307.4
	Var (ψ)	0.005	2.4	0.005	2.4	0.005	2.5	0.000	7.6
	Var($\bar{XB}$)	0.001	0.3	0.001	0.3	0.001	0.4	0.000	10.2
	2*Cov($\bar{\theta}.\psi$)	0.024	11.7	0.026	12.8	0.030	14	0.005	164.6
	2*Cov($\bar{\theta}.\bar{XB}$)	0.004	2.1	0.004	1.8	0.004	1.8	-0.001	-17.7
	2*Cov($\psi.\bar{XB}$)	0.002	0.8	0.001	0.7	0.002	0.8	0.000	0.4
Within-firm variance	Var($y - \bar{y}$)	0.130	62.4	0.119	59.1	0.118	55.8	-0.012	-372.6
	Var ($\theta - \bar{\theta}$)	0.113	54.3	0.105	52.4	0.112	53.1	-0.001	-26.7
	Var($Xb - \bar{Xb}$)	0.020	9.7	0.024	11.9	0.038	17.7	0.017	543.4
	Var(u)	0.011	5.1	0.010	5.1	0.011	5.2	0.000	7.7
	2*Cov($\theta - \bar{\theta}.Xb - \bar{Xb}$)	-0.014	-6.8	-0.021	-10.2	-0.043	-20.2	-0.029	-896.9
	2*Cov($\theta - \bar{\theta}, u$)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.8
	2*Cov($Xb - \bar{Xb}, u$)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	-0.4
Segregation Index	$\frac{Var(\bar{\theta}_j)}{Var(\theta_i)}$	0.273		0.304		0.317		0.045	
N*		61,925,099		66,706,199		65,410,886			

Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Decomposition based on Equations 2, 3 and 4. Firm clustering method is described in Section 2.3.2.

Table A7. Decomposition of wage variance and its evolution - Year fixed effects - Firm-split correction

		2002-2007		2008-2013		2014-2019		Change from 2002-2007 to 2014-2019	
		Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Total variance	Var(y)	0.213		0.208		0.220		0.006	
	Var (θ)	**	**	**	**	**	**	**	**
	Var (ψ)	0.014	6.5	0.014	6.7	0.013	5.8	-0.001	-18.2
	Var(Xb)	0.003	1.4	0.002	1.1	0.003	1.1	-0.001	-8.2
	Var(u)	**	**	**	**	**	**	**	**
	2*Cov(θ.ψ)	0.026	12.4	0.027	12.8	0.029	13.2	0.002	37.9
	2*Cov(θ.Xb)	0.000	0.0	0.000	0.2	-0.002	-0.8	-0.002	-25.5
	2*Cov(ψ.Xb)	0.001	0.3	0.001	0.3	0.001	0.3	0.000	0.1
Between-firm variance	Var($\bar{y}$)	0.088	41.1	0.094	45.0	0.103	46.7	0.015	233
	Var ($\bar{\theta}$)	0.043	20.4	0.05	23.8	0.057	26.1	0.014	213.2
	Var (ψ)	0.014	6.5	0.014	6.7	0.013	5.8	-0.001	-18.2
	Var($\bar{XB}$)	0.000	0.1	0.000	0.1	0.000	0.1	0.000	-0.9
	2*Cov($\bar{\theta}$.ψ)	0.026	12.4	0.027	12.8	0.029	13.1	0.002	37.1
	2*Cov($\bar{\theta}$.$\bar{XB}$)	0.001	0.6	0.001	0.6	0.001	0.7	0.000	3.9
	2*Cov(ψ.$\bar{XB}$)	0.001	0.3	0.001	0.3	0.001	0.3	0.000	0.1
Within-firm variance	Var($y - \bar{y}$)	0.126	58.9	0.114	55.0	0.117	53.3	-0.009	-133
	Var (θ − $\bar{\theta}$)	**	**	**	**	**	**	**	**
	Var($Xb - \bar{Xb}$)	0.003	1.3	0.002	1.0	0.002	1.1	0.000	-7.3
	Var(u)	**	**	**	**	**	**	**	**
	2*Cov(θ − $\bar{\theta}$.$Xb - \bar{Xb}$)	-0.001	-0.5	-0.001	-0.4	-0.003	-1.4	-0.002	-29.3
	2*Cov(θ − $\bar{\theta}$, u)	**	**	**	**	**	**	**	**
	2*Cov($Xb - \bar{Xb}$,u)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0
N* (largest connected set)		51,932,308		54,530,553		52,921,242			

Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Estimation on firms present in both main connected components in each split sample. Decomposition based on Equations 2, 3 and 4. The split-sampling method is described in Section 2.3.1.
AKM estimation with fixed effects for years and a cubic function of age flat at 40. No selection for age. Bias corrected with firm split.
** : These parameters' estimates are not corrected by firm-split

TABLE A8. Decomposition of wage variance and its evolution - Split-sampling correction with firm split - 90+ days worked in a year

		2002-2007		2008-2013		2014-2019		Change from 2002-2007 to 2014-2019	
		Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
Total variance	Var(y)	0.210		0.200		0.211		0.001	
	Var (θ)	**	**	**	**	**	**	**	**
	Var (ψ)	0.015	7.1	0.014	7.2	0.013	5.9	-0.002	-291.8
	Var(Xb)	0.040	19.3	0.043	21.5	0.068	32.1	0.027	3190.6
	Var(u)	0.025	12	0.023	11.5	0.026	12.3	0.001	105.6
	2*Cov($\theta.\psi$)	0.024	11.5	0.024	12.2	0.026	12.5	0.002	258.8
	2*Cov($\theta.Xb$)	-0.027	-12.6	-0.034	-17.2	-0.073	-34.4	-0.046	-5384
	2*Cov($\psi.Xb$)	0.004	1.7	0.004	2.1	0.004	2	0.001	79.3
Between-firm variance	Var($\bar{y}$)	0.089	42.5	0.092	46.0	0.101	48.0	0.012	1388.8
	Var ($\bar{\theta}$)	0.037	17.6	0.041	20.4	0.05	23.6	0.013	1486.4
	Var (ψ)	0.015	7.1	0.014	7.2	0.013	5.9	-0.002	-291.8
	Var($\bar{XB}$)	0.008	3.8	0.008	4.2	0.014	6.4	0.006	647.3
	2*Cov($\bar{\theta}.\psi$)	0.024	11.5	0.024	12.2	0.026	12.5	0.002	250.2
	2*Cov($\bar{\theta}.\bar{XB}$)	0	-0.2	-0.002	-1.1	-0.007	-3.4	-0.007	-797.1
	2*Cov($\psi.\bar{XB}$)	0.004	1.7	0.004	2.1	0.004	2	0.001	79.3
Within-firm variance	Var($y - \bar{y}$)	0.121	57.5	0.108	54.0	0.110	52.0	-0.011	-1288.8
	Var ($\theta - \bar{\theta}$)	**	**	**	**	**	**	**	**
	Var($Xb - \bar{Xb}$)	0.032	15.5	0.034	17.3	0.054	25.7	0.022	2543.2
	Var(u)	**	**	**	**	**	**	**	**
	2*Cov($\theta - \bar{\theta}.Xb - \bar{Xb}$)	-0.026	-12.3	-0.031	-15.7	-0.064	-30.4	-0.038	-4478.0
	2*Cov($\theta - \bar{\theta}, u$)	**	**	**	**	**	**	**	**
	2*Cov($Xb - \bar{Xb}, u$)	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.4
N* (largest connected set)		84,772,337		89,060,613		91,320,821			

Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 90 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Estimation on firms present in both main connected components in each split sample. Decomposition based on Equations 2, 3 and 4. The split-sampling method is described in Section 2.3.1.

** : These parameters' estimates are not corrected by firm-split

Table A9. Composition effects: number of firms and average size by sorting quadrant

Firms present in both periods (2-period)

	Person-Year Observations				Number of Firms				Average Size			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	F>0, W>0	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*	*F>0, W>0*	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*	*F>0, W>0*	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*
2002-2007	14,003,660	8,892,099	13,122,069	5,748,544	29,114	41,287	79,037	51,230	480.99	215.37	166.02	112.21
2014-2019	15,379,346	9,252,809	15,582,805	6,931,931	29,114	41,287	79,037	51,230	528.25	224.11	197.16	135.31
Diff	1,375,686	360,710	2,460,736	1,183,387	0	0	0	0	47.25	8.74	31.13	23.10

Firms present only in one period (1-period)

	Person-Year Observations				Number of Firms				Average Size			
	F>0, W>0	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*	*F>0, W>0*	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*	*F>0, W>0*	*F>0, W<0*	*F<0, W<0*	*F<0, W>0*
2002-2007	3,609,701	4,516,577	5,585,022	3,188,645	34,278	82,244	120,218	99,406	105.31	54.92	46.46	32.08
2014-2019	2,612,327	2,636,805	4,357,807	2,796,457	37,580	81,401	131,086	113,378	69.51	32.39	33.24	24.66
Diff	-997,374	-1,879,772	-1,227,215	-392,188	3,302	-843	10,868	13,972	-35.79	-22.52	-13.21	-7.41

Notes: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. The largest connected set firms present in both periods and present either in 2002-2007 or in 2014-2019 are analyzed separately. Firms are further divided into four quadrants according to the value of the estimated worker (W) and firm fixed effects (F) (see Equation 1), averaged over both periods for staying firms. Worker and firm fixed effects are normalized so they are comparable across periods. Columns 1, 5, and 9 refer to the number of workers where both worker (W) and firm fixed effects (F) are above zero (F>0, W>0). The other columns map the rest of the possible combinations for F and W values. For firms present in both periods, the (not employment-weighted) 2-period average of firm and worker fixed effects are considered in the allocation to the four quadrants. Entries in columns 5-8 and 9-12 are the average worker and firm fixed effects, respectively, by quadrant.

TABLE A10. Change in the occupational structure by type of firm – 1-period firms

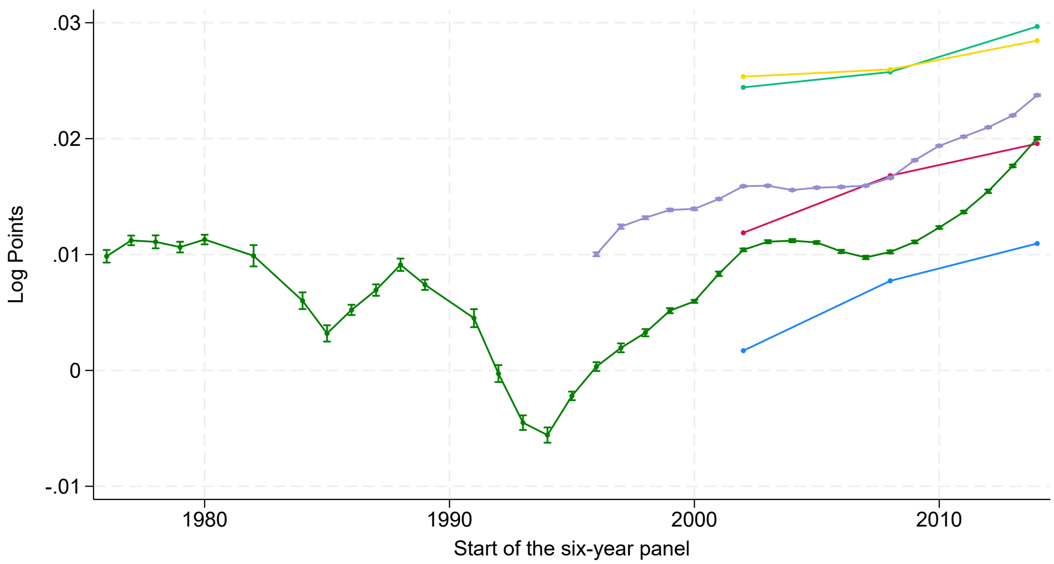
		1-period, F>0, W>0			1-period, F>0, W<0			1-period, F<0, W<0			1-period, F<0, W>0		
Occupation	CS	2002-2007	2014-2019	Diff	2002-2007	2014-2019	Diff	2002-2007	2014-2019	Diff	2002-2007	2014-2019	Diff
Entrepreneurs	20	0.0103	0.0163	0.0060	0.0112	0.0163	0.0051	0.0112	0.0130	0.0018	0.0277	0.0283	0.0006
Professionals	32	0.0083	0.0104	0.0021	0.0044	0.0067	0.0022	0.0028	0.0078	0.0050	0.0136	0.0295	0.0159
Artists and media professionals	35	0.0175	0.0081	-0.0095	0.0028	0.0032	0.0004	0.0008	0.0012	0.0004	0.0087	0.0071	-0.0016
Managers	37	0.1676	0.1965	0.0289	0.0479	0.0544	0.0065	0.0319	0.0322	0.0003	0.1053	0.1165	0.0112
Engineers	38	0.1986	0.2255	0.0269	0.0325	0.0342	0.0017	0.0160	0.0114	-0.0046	0.0645	0.0886	0.0241
Primary school teachers	42	0.0047	0.0042	-0.0005	0.0040	0.0113	0.0072	0.0042	0.0210	0.0168	0.0153	0.0293	0.0140
Health and social workers	43	0.0091	0.0094	0.0002	0.0207	0.0253	0.0046	0.0229	0.0399	0.0170	0.0421	0.0477	0.0056
Public administration intermediates	45	0.0014	0.0003	-0.0012	0.0002	0.0002	0.0000	0.0005	0.0001	-0.0004	0.0012	0.0004	-0.0008
Business administration intermediates	46	0.1308	0.0906	-0.0402	0.0561	0.0618	0.0057	0.0505	0.0442	-0.0063	0.1055	0.0871	-0.0184
Technicians	47	0.1041	0.1017	-0.0024	0.0445	0.0456	0.0011	0.0233	0.0195	-0.0038	0.0657	0.0667	0.0010
Intermediate supervisors	48	0.0310	0.0506	0.0197	0.0410	0.0301	-0.0108	0.0299	0.0151	-0.0148	0.0219	0.0153	-0.0066
Public administration clerks	52	0.0053	0.0024	-0.0029	0.0174	0.0185	0.0011	0.0281	0.0502	0.0221	0.0151	0.0185	0.0034
Security agents	53	0.0027	0.0022	-0.0004	0.0108	0.0066	-0.0042	0.0258	0.0259	0.0001	0.0020	0.0026	0.0006
Business administration clerks	54	0.1365	0.1055	-0.0311	0.1096	0.1187	0.0091	0.0975	0.0993	0.0018	0.2034	0.1721	-0.0313
Retail salespersons	55	0.0058	0.0143	0.0085	0.0775	0.0708	-0.0066	0.1154	0.1160	0.0006	0.0470	0.0505	0.0035
Personal service employees	56	0.0052	0.0108	0.0056	0.0343	0.0697	0.0354	0.0888	0.1514	0.0626	0.0393	0.0585	0.0192
Skilled manufacturing workers	62	0.1051	0.0949	-0.0101	0.2063	0.1146	-0.0917	0.1237	0.0442	-0.0795	0.0807	0.0345	-0.0462
Skilled artisans	63	0.0185	0.0198	0.0014	0.0785	0.1130	0.0344	0.0648	0.0840	0.0192	0.0651	0.0761	0.0110
Drivers	64	0.0051	0.0045	-0.0006	0.0461	0.0539	0.0079	0.0804	0.0756	-0.0048	0.0191	0.0135	-0.0056
Handling, transport skilled workers	65	0.0162	0.0123	-0.0040	0.0456	0.0426	-0.0030	0.0278	0.0172	-0.0106	0.0201	0.0110	-0.0091
Unskilled manufacturing workers	67	0.0125	0.0128	0.0002	0.0903	0.0584	-0.0320	0.0728	0.0362	-0.0366	0.0210	0.0157	-0.0053
Unskilled artisans	68	0.0029	0.0050	0.0022	0.0162	0.0289	0.0127	0.0768	0.0774	0.0006	0.0132	0.0170	0.0038
Farm workers	69	0.0007	0.0019	0.0012	0.0020	0.0152	0.0133	0.0042	0.0172	0.0130	0.0024	0.0132	0.0108

Notes: All firms and individuals in firms with at least 2 employees are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Firms are divided into types as described in Table 4. We report the incidence of a certain 2-digits occupation in 2002-2007 vs 2014-2019, and the difference (in percentage points), by type. The color code in the visualization indicates the intensity of the change, with darker shades of red representing more negative changes and darker shades of green representing more positive changes. The second column displays the two-digit code corresponding to the *catégories socioprofessionnelles*, which can be further explored in detail here.

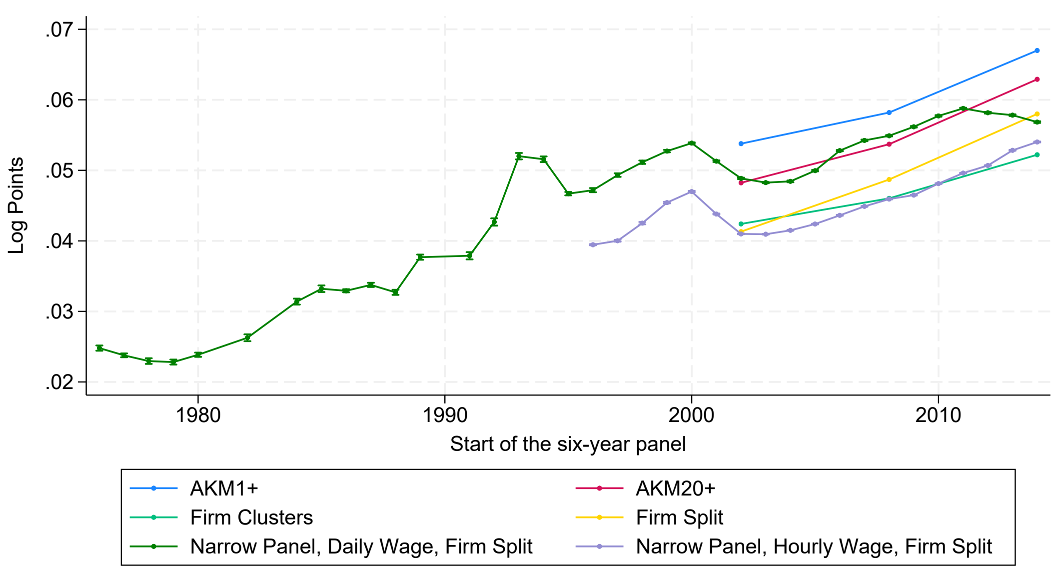
Appendix B. Appendix Figures

FIGURE A1. Historical Series

A. Sorting

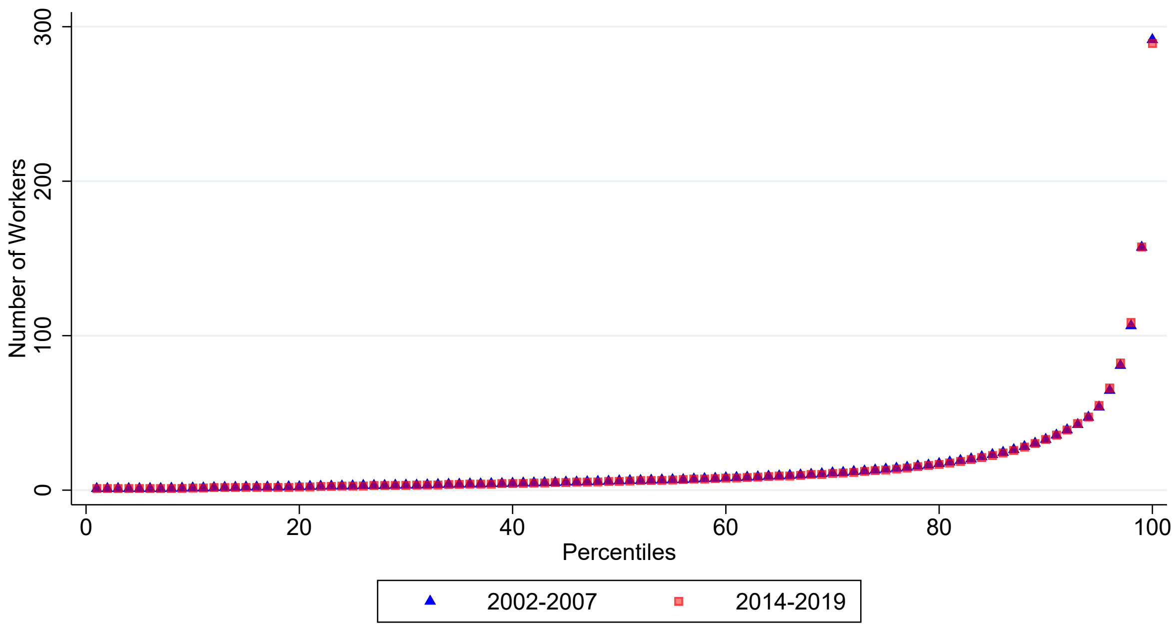


B. Segregation



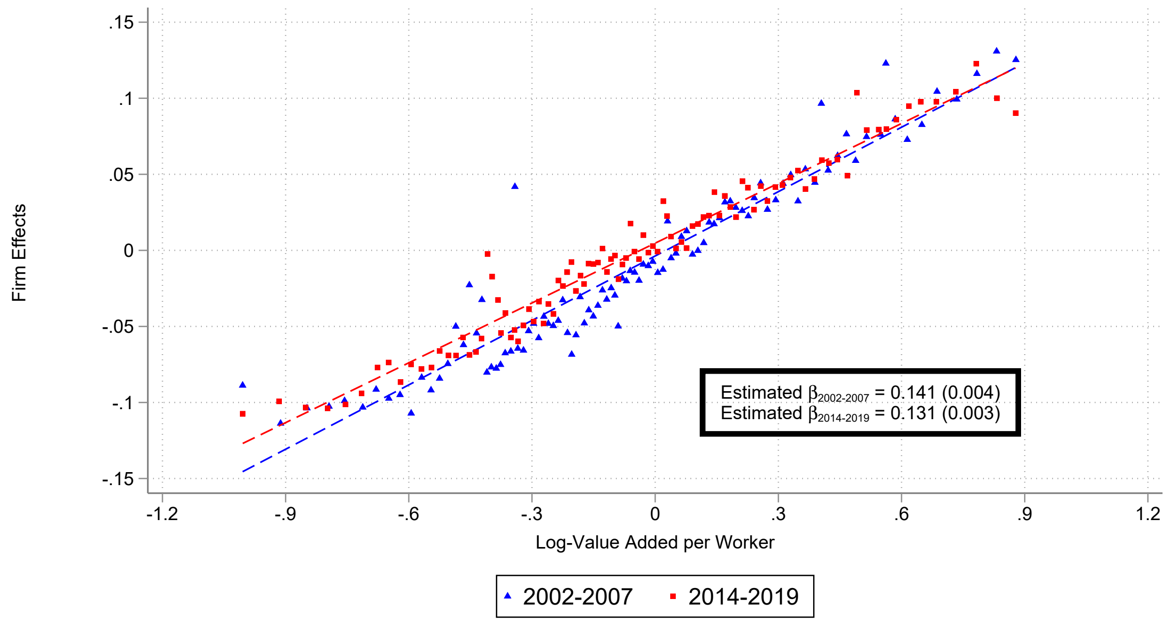
Note: Only individuals employed for at least 360 days for the exhaustive panels, and 90 days for the narrow panel, by the same firm during the year are included for a given year. The wider selection on the narrow panel aims at increasing connectivity. Individuals and firms in public administration are not included. Sorting – $2 * Cov(\theta, \psi)$ – and Segregation – $Var(\bar{\theta})$ – estimates by six-year periods. Long-term series are computed on the narrow panel, on rolling six-year periods, corrected by firm-splitting, with mean estimate and confidence intervals computed on repeated (split) sampling with 20 repetitions, reflecting only the noise stemming from the randomness of the split. Estimates on the narrow panel are particularly affected by sampling error and the selection of bigger and more connected firms. The years 1981, 1983, and 1990 are missing. There have been several changes in scope and variable definition since 1976.

FIGURE A2. Cumulative firm size distribution



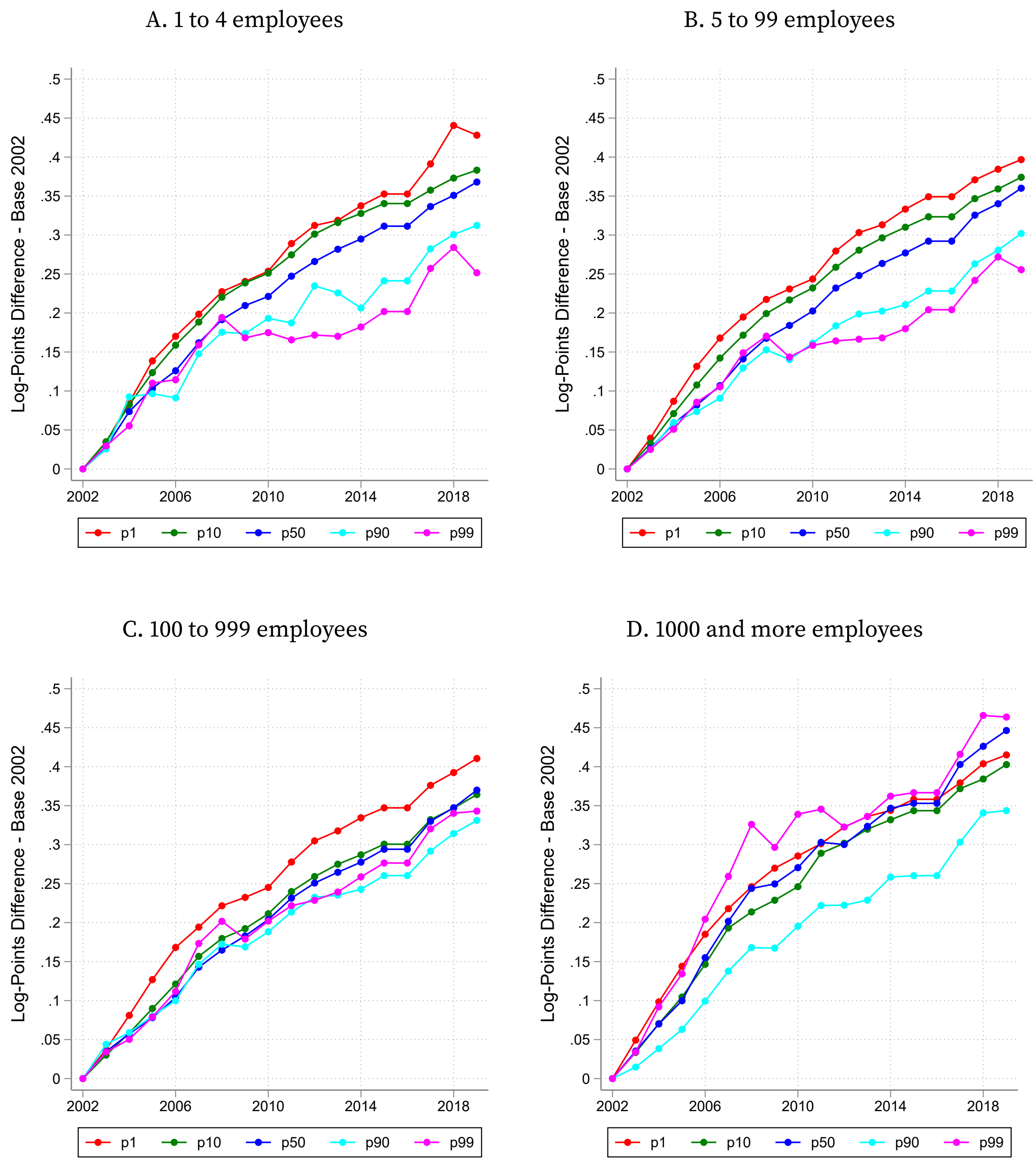
Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. This figure shows the fraction of firms below a given size, by period.

FIGURE A3. Rent-Sharing - Firm Fixed Effects vs Log Value Added/Worker



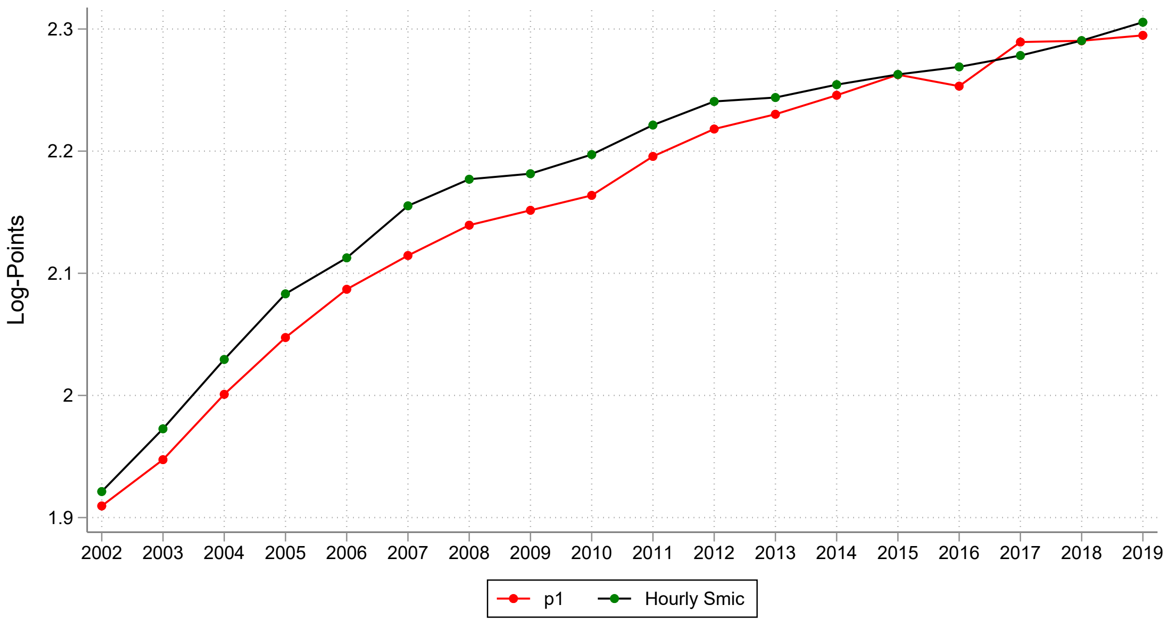
Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Points shown represent mean estimated firm fixed effects from AKM models, averaged across firms in 100 percentile bins of the period-demeaned log value added per worker. Period best-fitting lines coming from employment-weighted OLS are reported.

FIGURE A4. Change in Percentiles of Real Earnings Relative to 2002 - By Firm Size Class



Note: This figure illustrates the average change in log-earnings for employees across different percentiles, relative to the year 2002. The data is categorized into four distinct firm size classes, based on measurements taken on December 31st of each year. The percentiles range from the top 1% of earners to the employees in the first percentile.

FIGURE A5. Evolution of P1 of real earnings and of real hourly minimum wage



Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. The figure plots the evolution over time of the first percentile of real earnings and of the real hourly minimum wage (in log points).

Appendix C. Constructing a BTS full panel

C.1. Chaining the yearfiles

The French BTS is not a proper panel dataset as there are no individual IDs before 2002 and after 2002 the individual IDs are specific to each yearfile. However, each yearfile y contains information both on the current year t and the preceding year $t-1$ (variables for the year $t-1$ end with "_1"). We therefore take advantage of this overlap to build a pseudo-panel based on common information between year t of yearfile $y-1$ and year $t-1$ of yearfile y.

We obtained Insee's authorization to chain the BTS yearly files in order to create a full panel of the wage-earning population between 1994 and 2020. Between 1994 and 2001, as there are no individual IDs enabling to track mobilities within the yearfiles, we can only match the "stayers", employees as long as they stay in the same establishment: for instance, an individual who moves from workplace j to workplace k will be given two different IDs. After 2002, individual IDs in the original yearfiles, even if yearfile specific, enables us to match both stayers and movers.

In order to conduct the match, we used the following variables: sex (`SEXE`), firm ID (`SIREN`), establishment ID (`NIC`), number of hours (`NBHEUR` or `NBHEUR_1`), starting day of the job during the year (`DATDEB` or `DATDEB_1`), ending day of the job during the year (`DATFIN` or `DATFIN_1`), number of days between starting and ending day (`DUREE` or `DUREE_1`), municipality of residence (`COMR` or `COMR_1`), municipality of work (`COMT` or `COMT_1`), being part of the sample used for the DADS panel (`SONDE` or `SONDE_1`), and gross wage (`S_BRUT` or `S_BRUT_1`) and age (`AGE`).

We run the match with a SAS script at the regional level, using the BTS regional files[31]. Within the regional file, we keep the job for which a worker i has the highest pay.

We create the following keys for the year t of yearfile $y-1$:

```
pseudoid=COMPRESS(SEXE!!"#"!!SIREN!!"#"!!NIC!!"#"!!ROUND(NBHEUR,1)!!"#"!!
DATDEB!!"#"!!DATFIN!!"#"!!DUREE !! "#" !!COMR!!"#"!!COMT !! "#" !! SONDE);
```

and the following for the year $t-1$ of yearfile y:

```
pseudoid_b=COMPRESS(SEXE!!"#"!!SIREN!!"#" !!NIC!!"#"!!ROUND(NBHEUR_1,1)!!"#"!!
DATDEB_1!!"#"!!DATFIN_1!!"#"!!DUREE_1!! "#"!! COMR_1!!"#"!!COMT_1!!"#" !! SONDE_1);
```

[31]In the current project, we restricted the match to mainland France and excluded overseas departments (DOM).

However, as there are some discrepancies in the ages and the wages reported for the same year in yearfile $y-1$ and y, we do not use them directly in the matching key. We use the HAVING property of the SQL procedure, in order to select the match with the minimal difference between the two wages and an absolute age difference below two years.

```
PROC SQL;
    CREATE TABLE ab (DROP=pseudoid pseudoid_b S_BRUT S_BRUT_1 AGE)
        AS SELECT * FROM a1 (KEEP=pseudoid s_brut IDENT_S ID2 REGT
        AGE NBHEUR) AS aa
    FULL JOIN b1 (keep=pseudoid_b s_brut_1 IDENT_S ID2_B AGE
        DEP_NAISS NBHEUR_1 rename=(IDENT_S=IDENT_S_B AGE=AGE_B))
        AS bb
    ON aa.pseudoid=bb.pseudoid_B
    GROUP BY aa.S_BRUT,aa.PSEUDOID
    HAVING ABS(aa.s_brut-bb.s_brut_1)=MIN(ABS(aa.s_brut-bb.s_brut_1))
    AND (0<=bb.AGE_B-aa.AGE<2 or AGE_B=. or AGE=.)
    ORDER BY aa.PSEUDOID, bb.s_brut_1;
QUIT;
```

This code was adapted to account for fileyear specificity.

- For years before 2002 ($y < 2002$ and $y-1 < 2001$), we create an individual ID based on the initial row numbers in each regional file, to which we add at the end the regional code. For instance: the ID for the 10th observation of the Paris region (code: 11) will be 1011.
- In 2013 ($y = 2013$ and $y-1 = 2012$), the `SONDE` variable leads to some mismatch and is excluded from the pseudoid key.
- After 2013 ($y > 2013$ and $y-1 > 2012$), we found that the number of hours for the same year differed between yearfile $y-1$ and y. We thus excluded the number of hours from the matching key and we added the minimal difference in the number of hours in the having clause.

We count the number of matches based on the procedure and we attribute the same ID only to workers with a single match. Finally, we chain the different IDs starting from the first year of the DADS (1994). The ID files (*PSID*_1994 to *PSID*_2020) contain the ID of

the year (IDENT_S) and a permanent ID (IDENT_ALL), which is based on the initial ID of an employee when she first appears in the DADS, to which we add the year of the first appearance on two digits. The full SAS script pseudo_id.sas is available at the following address:
http://olivier.godechot.free.fr/hopfichiers/pseudo_id.zip

It comes with three additional SAS scripts for creating DADS files with the identifier IDENT_ALL included, for creating and adding seniority variables, and for correcting information on workers' location of birth and citizenship.

C.2. Quality of the identification

To avoid false identification, we opted for a conservative procedure in order to identify two individuals as the same person, by using the maximal available overlapping information. When the procedure leads to multiple matches, we do not impute any identification. However, these duplicates remain rare, around 0.4% of the observations. Most matching failures are due to observations for which we don't find any match.

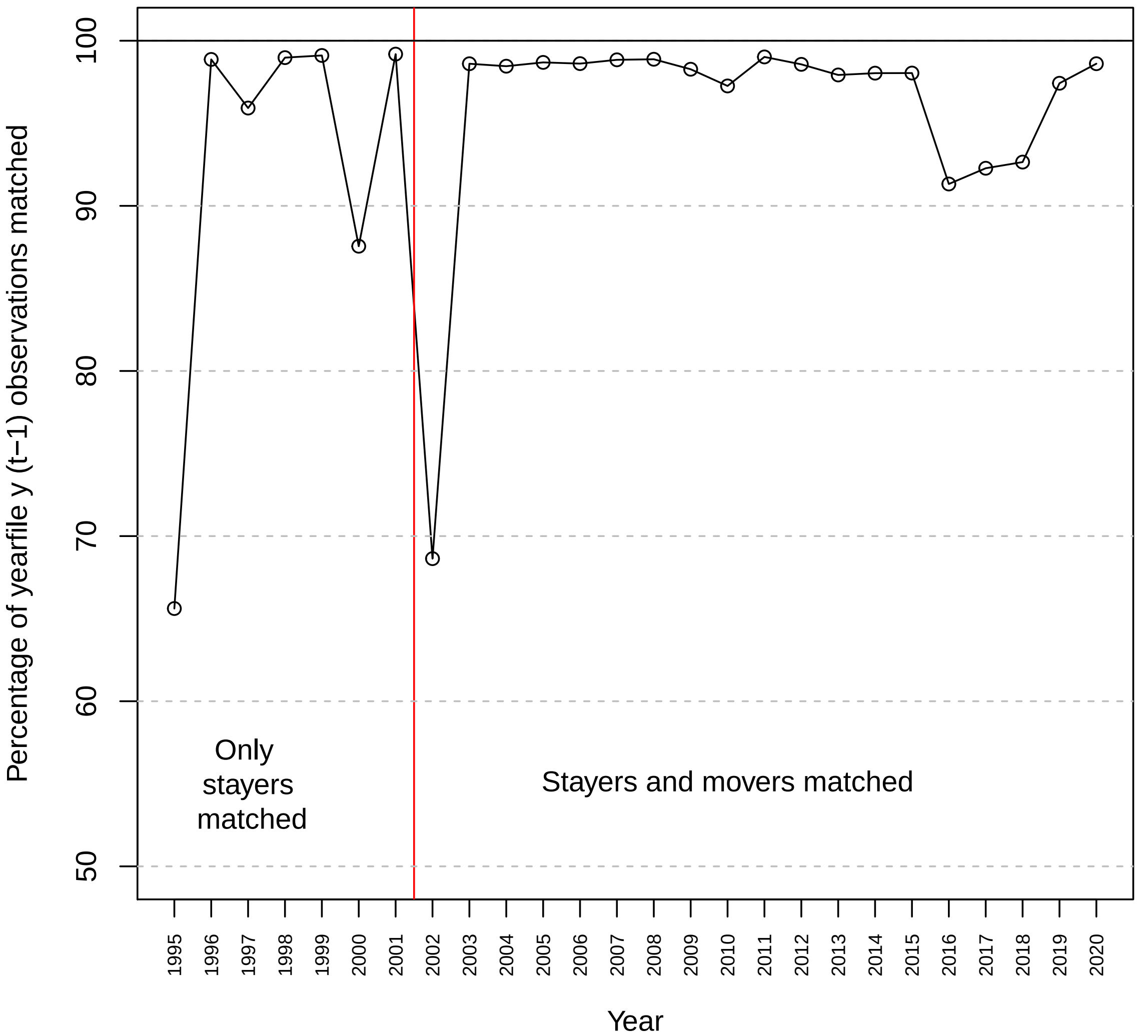
Figure A6 gives a first proxy of the quality of the matching. Generally, we find a single match for 98% of the observations of the overlapping years of two yearfiles. The quality of the matching declines between 2016 and 2018, dropping to 91-93% and resumes back to 97% in 2019, probably as a result of the switch from DADS to the DSN[32]. With the existing procedure, the match is poor for yearfile 2002 (and similarly in 1995), as the consequence of the major transformation of the BTS between the 1994-2001 series and the 2002-2020 series[33].

Despite a high level of matches, the matching has some limitations. We must bear in mind, that a false positive is still possible (but unlikely). Second, in order to be identified as the same person, an employee needs to be present each year as a wage earner in the BTS. This means that we cannot link the initial ID of an employee who was either further unemployed, self-employed or civil servant (before 2009) for more than a calendar year to her subsequent employment periods. Overall, the quality of the match seems sufficiently good to run AKM panel regressions.

[32]The "déclaration sociale nominative" is a new monthly administrative source replacing the "déclarations annuelles de données sociales". INSEE produces from the DSN a yearly datafile on the BTS format for series continuity.

[33]One could probably improve the match for these years by eliminating variables in the matching key which are incomplete or coded differently. We already dropped the number of hours for 2002 and this increased the matching rate from 60 to 68%.

FIGURE A6. Quality of the match



Note: The figure present the proportion of observations from the year $t-1$ of yearfile y for which we found a single match in the year t of yearfile $y-1$. Before 2002, the lack of individual ID in the initial dataset makes it impossible to follow the movers. In 2002 and after, we can match both stayers and movers.

C.3. How to use the ID files

Hence, in order to add the permanent ID to a given datafile (for instance a file `b2010` for the year 2010), the procedure is as follows:[34]

```
PROC SQL;
 CREATE TABLE b2010b
 AS SELECT * FROM b2010 AS aa
```

[34] The script pseudo_id_use.sas also provides a macro program to run these steps automatically.

```
 LEFT JOIN psid.psid_2010 AS bb
 ON aa.ident_s=bb.ident_s;
QUIT;

data b2010c; set b2010b;
 if Missing(ident_all) then ident_all=ident_s*100+substr(AN,3,4);
run;
```

Before 2002, in order to get permanent IDs necessary for the match with the `PSID_yyyy` files, one needs to create an ID in each regional file (prior to any selection) as follows (for instance for Paris Region in 1997):

```
DATA b1197; SET po1997.post1197;
 ident_s=_N_*100+REG;
RUN;
```

Appendix D. Proof of split-sampling bias correction

We suppose that observations are randomly split into two half samples and that each sample retains the same connectivity as the full sample. In practice, we reduce the data to firms and individuals (for period-split, or only firms for firm-split) that belong to both main connected components in each split sample. This condition is analogous to the Kline, Saggio, and Sølvsten (2020) leave-one-out condition that the main connected sample is not disconnected when any one observation is removed. We start again with a simplified notation of the AKM model in equation 5:

$$y_i = z_i'\alpha + u_i \tag{A1}$$

With $\alpha = (\beta, \theta, \psi)$ our parameter vector of length $k = 2 + N + J$ and z_i the non-random regressors vector of the (worker * year) i's observation characteristics, including the indicator vector for worker and firm. For a given symetric matrix A corresponding to a given quadratic form ω of interest, our split-sampling plug-in estimator becomes $\hat{\omega}^{SP} = \hat{\alpha}_0' A \hat{\alpha}_1$ with $\hat{\alpha}_s$ an OLS estimate in the sample I_s, $s = 0, 1$ of size N_s : $\hat{\alpha}_s = S_{zz,s}^{-1} \sum_{i\in s}^{N_s} z_i y_i = \alpha + S_{zz,s}^{-1} \sum_{i\in s}^{N_s} z_i u_i$, with $S_{zz,s} = \sum_{i\in s}^{N_s} z_i z_i'$ the split sample design matrix (with full rank when we limit the sample to the split sample main connected set). We can express $\hat{\alpha}_s$ as

$\hat{\alpha}_s = \alpha + \epsilon_s$. We calculate the bias by expressing a scalar as the trace of a (1,1) matrix, as in the classic demonstration of the expectation of quadratic forms

$$
\begin{aligned}
\mathbf{E}[\hat{\omega}^{SP}] &= \mathbf{E}[\hat{\alpha}_0' A \hat{\alpha}_1] = \mathbf{E}[trace(\hat{\alpha}_0' A \hat{\alpha}_1)] \\
&= \mathbf{E}[trace(\hat{\alpha}_1 \hat{\alpha}_0' A)] && \text{by propriety of trace()} \\
&= \mathbf{E}[trace(A \hat{\alpha}_1 \hat{\alpha}_0')] && \text{idem} \\
&= trace(A \mathbf{E}[\hat{\alpha}_1 \hat{\alpha}_0{}']) && \text{A non-random, trace() is linear}
\end{aligned}
$$

We further have:

$$
\begin{aligned}
\mathbf{E}[\hat{\alpha}_1 \hat{\alpha}_0{}'] &= \mathbf{E}[(\alpha + \epsilon_1)(\alpha + \epsilon_0)'] \\
&= \alpha\alpha' + \mathbf{E}[\epsilon_1 \epsilon_0']
\end{aligned}
$$

And: $trace(A\alpha\alpha') = \omega$

The bias is thus equal to:

$$
\begin{aligned}
\mathbf{E}[\hat{\omega}^{SP}] - \omega &= trace(A\mathbf{E}[(S_{zz,1}^{-1} \sum_{i \in I_1}^{N_1} z_i u_i)(S_{zz,0}^{-1} \sum_{j \in I_0}^{N_0} z_j u_j)']) \\
\mathbf{E}[\hat{\omega}^{SP}] - \omega &= trace(A\mathbf{E}[(S_{zz,1}^{-1} \sum_{i \in I_1}^{N_1} u_i z_i)(S_{zz,0}^{-1} \sum_{j \in I_0}^{N_0} u_j z_j)']) \\
&= trace(A S_{zz,1}^{-1} \underbrace{\mathbf{E}[(\sum_{i \in I_1}^{N_1} u_i z_i)(\sum_{j \in I_0}^{N_0} u_j z_j)']}_{\text{matrix } (b_{lm})} (S_{zz,0}^{-1})')
\end{aligned}
$$

with generic term:

$$
b_{lm} = \sum_{i \in I_1}^{N_1} u_i z_{l,i} \sum_{j \in I_0}^{N_0} u_j z_{m,j}
$$

The variance of the split sampling estimator of a quadratic form stems from the random errors u_i, but also from the randomness of the split. We can abstract from this last source by considering only a given split, and the corresponding sample of firms and individuals connected in both split samples. The size of the variance then depends on the quadratic form matrix A and the design matrices $S_{zz,s}$ and their relation, so additional hypotheses are needed for this variance to be of finite value. Kline, Saggio, and Sølvsten (2020) discuss these conditions in the context of leave-one-out. When introducing the randomness of the split, however, A and $S_{zz,s}$ become random matrices.

Finally, to study the consistency and convergence of our estimator, we need to consider the series of these random matrices when some index n of the number of observations grows. We leave the study of the precise conditions for consistency and convergence to further research. Instead, we checked that the estimator showed reasonable stability over multiple random splits, and recovered known values in Monte Carlo experiments.

D.1. Simulations

We generated simulated workers' and firms' fixed effects with sorting and noise, calibrated on measured distributions, to get simulated wages on the observed match. It is important to keep the real mobility network, on which the bias depends. In Table A11, split sample correction recovers the true value on the sample of firms belonging to a connected component in both splits, which slightly differs from the full sample true values because firms are bigger and there is fewer firm variance in fixed effects and mean workers effects. This kind of simulation cannot however reproduce potential selection effects in the sample reduction.

TABLE A11. Simulated wage: true fixed effects and estimations

	(1)	(2)	(3)	(4)
Var(θ)	0.1494	0.1543	0.1488	
Var(ψ)	0.0138	0.0164	0.0136	0,0137
2Cov(θ,ψ)	0.0248	0.0199	0.0243	0.0243
N of obs	58,666,317	58,666,317	52,157,735	52,157,735

Simulation on 2002-2007 data, corrected estimates with period split method. First column: ground truth on AKM estimation sample (true quadratic terms computed on simulated fixed effects). Second column: AKM estimates (on simulated fixed effects). Third column: ground truth on the split-sampling estimation sample. Fourth column: slip-sampling estimates.

D.2. Multiple random splits

We checked the stability of the (firm split) split sample estimators with multiple random splits on two different data sets. First on the long-term historical series, which are computed on the smaller and less connected "narrow panel" and show more variability due to splitting. We plot the means of 20 split sample estimations and a confidence interval on this mean in Figure A1A.

On our main estimates, we limited the multiple random split experiments to the firm-split in the first and third periods for computational reasons. In Table A12 we report the mean and standard deviation of 20 estimations.

TABLE A12. Decomposition of wage variance - Mean and standard deviation over 20 firm split estimations

		2002-2007		2014-2019	
		Mean	SD	Mean	SD
Total variance	Var(y)	0.211	0.00003	0.219	0.00003
	Var(ψ)	0.014	0.00004	0.013	0.00005
	Var(Xb)	0.019	0.00001	0.040	0.00002
	2Cov(θ,ψ)	0.026	0.00008	0.029	0.00006
	2Cov(θ,Xb)	-0.007	0.00002	-0.039	0.00003
	2Cov(ψ,Xb)	0.002	0.00001	0.001	0.00002
Between-firm variance	Var($\bar{y}$)	0.087	0.00002	0.103	0.00003
	Var($\bar{\theta}$)	0.041	0.00007	0.058	0.00006
	Var($\bar{XB}$)	0.003	0.00000	0.006	0.00000
	2Cov($\bar{\theta}$,ψ)	0.026	0.00008	0.029	0.00007
	2Cov($\bar{\theta}$,$\bar{XB}$)	0.000	0.00001	-0.005	0.00002
	2Cov(ψ,$\bar{XB}$)	0.002	0.00001	0.001	0.00002
Within-firm variance	Var($y - \bar{y}$)	0.124	0.00002	0.116	0.00001
	Var($Xb - \bar{Xb}$)	0.016	0.00001	0.034	0.00002
	2Cov($\theta - \bar{\theta}$,$Xb - \bar{Xb}$)	-0.007	0.00001	-0.033	0.00003
N of obs		52,149,132	12,361.51	53,142,360	8,563.32

Note: Mean and standard deviations computed on 20 estimations similar to table 2, on firms belonging to both main connected components

The standard deviations are very small relative to the estimates, the sizes of the bias correction, and the evolution between periods. Our split sampling corrected results do not stem from random split noise. This exercise can also be interpreted as a bootstrap, indicative more generally of the high stability of AKM decomposition statistics in our large dataset.

Appendix E. Log-wage variance decomposition from alternate specifications

We conduct multiple robustness experiments with alternate specifications.

E.1. Model with year fixed effects and a cubic function of age

Card et al. (2018)'s AKM model specification differs from ours. While we demean wages per year, they use year fixed effects. To avoid collinearity between year effects, workers effects, and age, they suppress the linear term in age and keep only a quadratic and a cubic term in (age-40). We implement this specification and present results in Table A7, with split sampling bias correction with the firm-split method (variance of workers' effects and residual are not corrected).

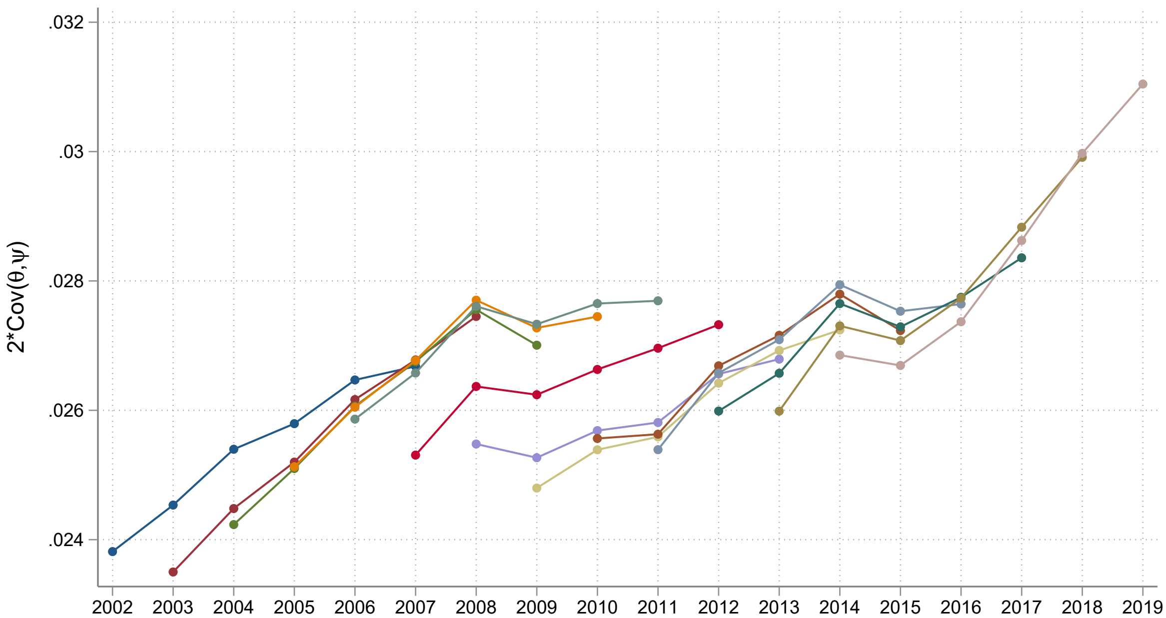
With this specification, the covariance between workers' effects and age functions is close to zero. The rise in sorting is still visible but less important.

E.2. Rolling panels and yearly log-wage variance decomposition

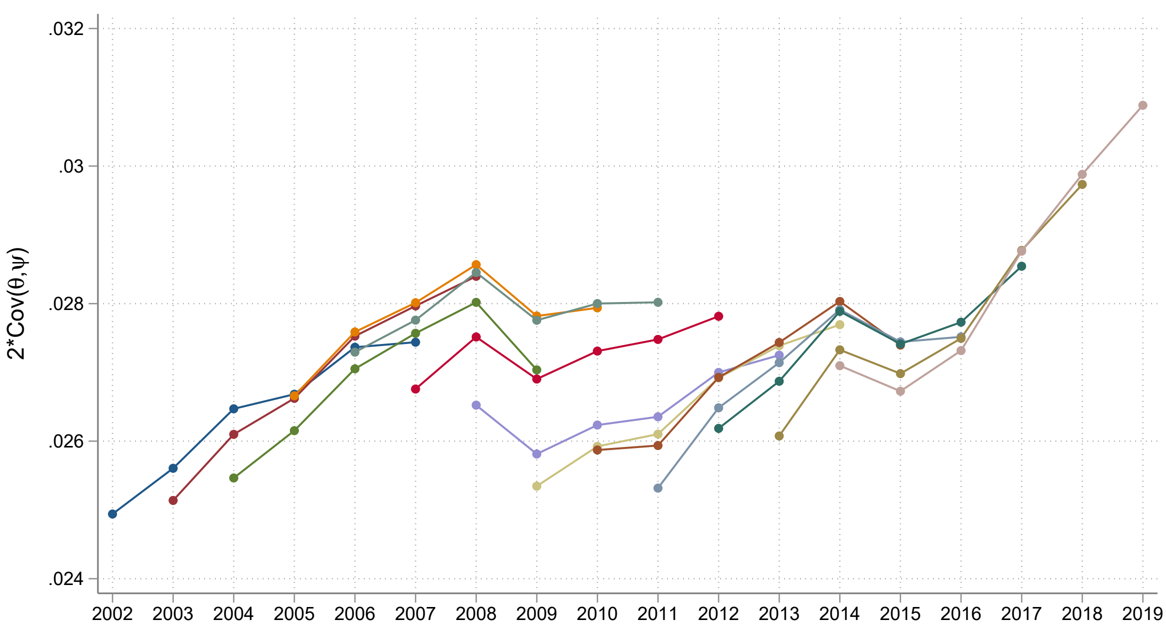
Are fixed effects really fixed? To better understand their stability or lack thereof, we implement the rolling AKM (R-AKM) method of Lachowska et al. (2023) by estimating firm-split corrected AKM on six-year panels for every possible starting year in our data. We then apply log-wage variance decomposition separately for each year of each panel estimation. Within each six-year panel, year-to-year variations in sorting reflect composition effects (year-to-year changes in the populations of firms and workers) and pure sorting (changes in worker-to-firm matching). If fixed effects were fixed, and the AKM estimation was correct, the yearly results would be the same whatever the panel used for estimation. On the contrary, we observe systematic shifts between different panels' estimates, for a given year (Figure A7). Sampling differences between panels due to the main connected component selection might explain some of the shift, but it is most likely due to changes in the fixed effects of a given firm or a given worker, estimated in different, overlapping panels. This, in turn, could be a mechanical consequence of the evolution of individuals' fixed effects themselves, from year to year.

FIGURE A7. Sorting - Rolling Panels

A. Baseline



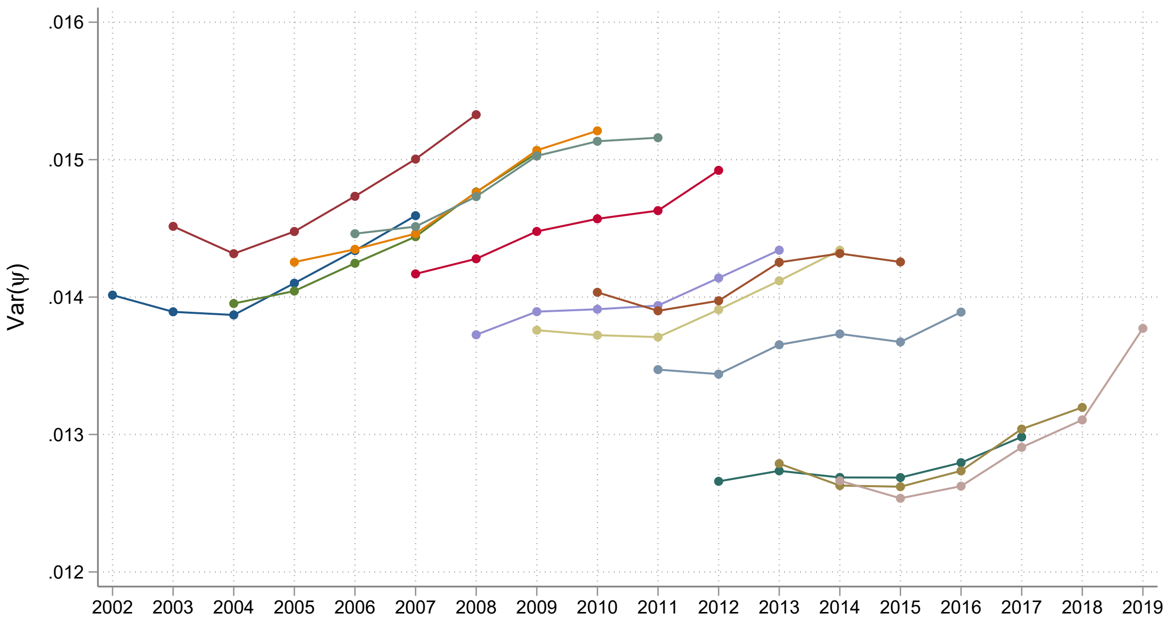
B. Year fixed effects and cubic function of (age-40)



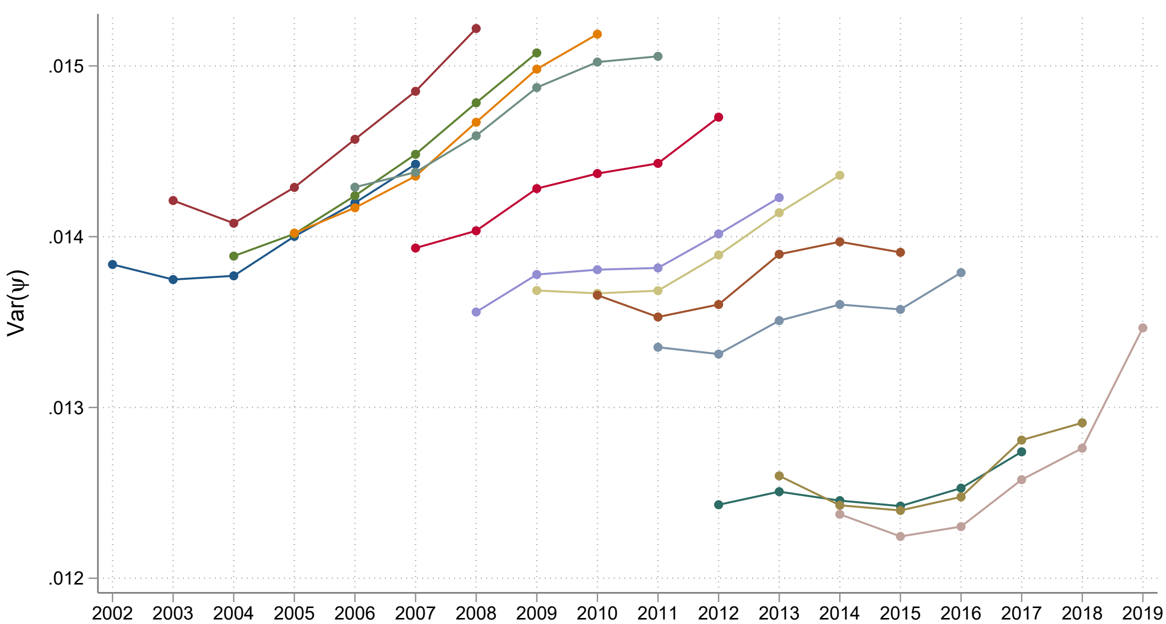
Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Fixed effects estimates by a panel of six-year periods split sample bias correction with firm splitting (Panel a). In Panel b, we further include year fixed effects and cubic function of (age-40). For each panel, log wage variance decomposition is computed separately for each year.

FIGURE A8. Variance of Firm Effects - Rolling Panels

A. Baseline



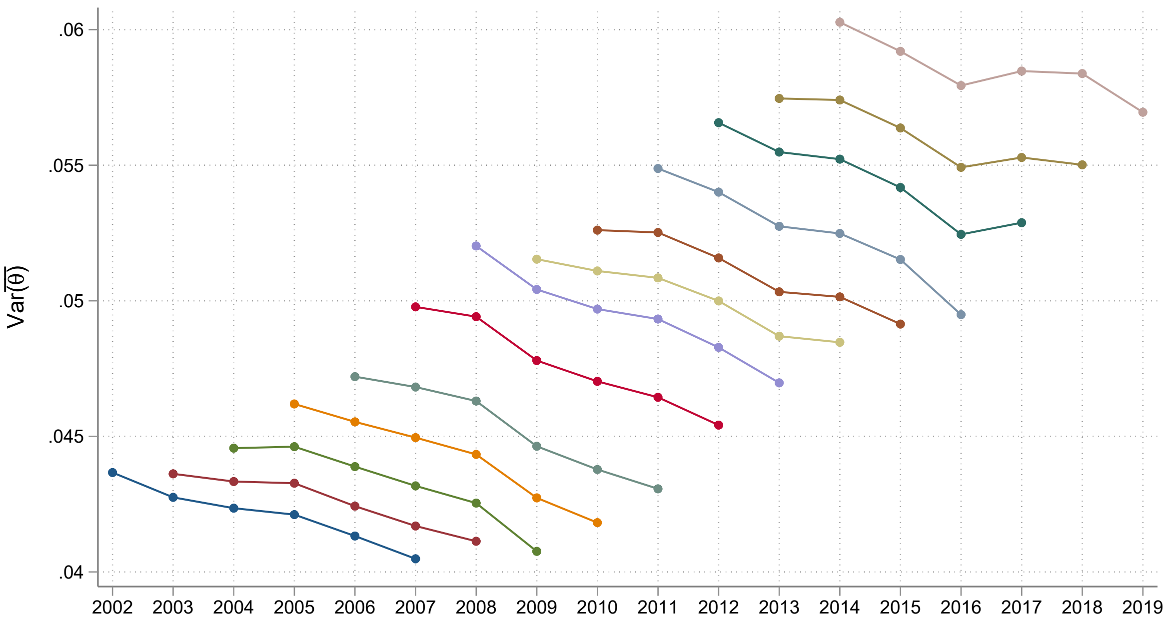
B. Year fixed effects and cubic function of (age-40)



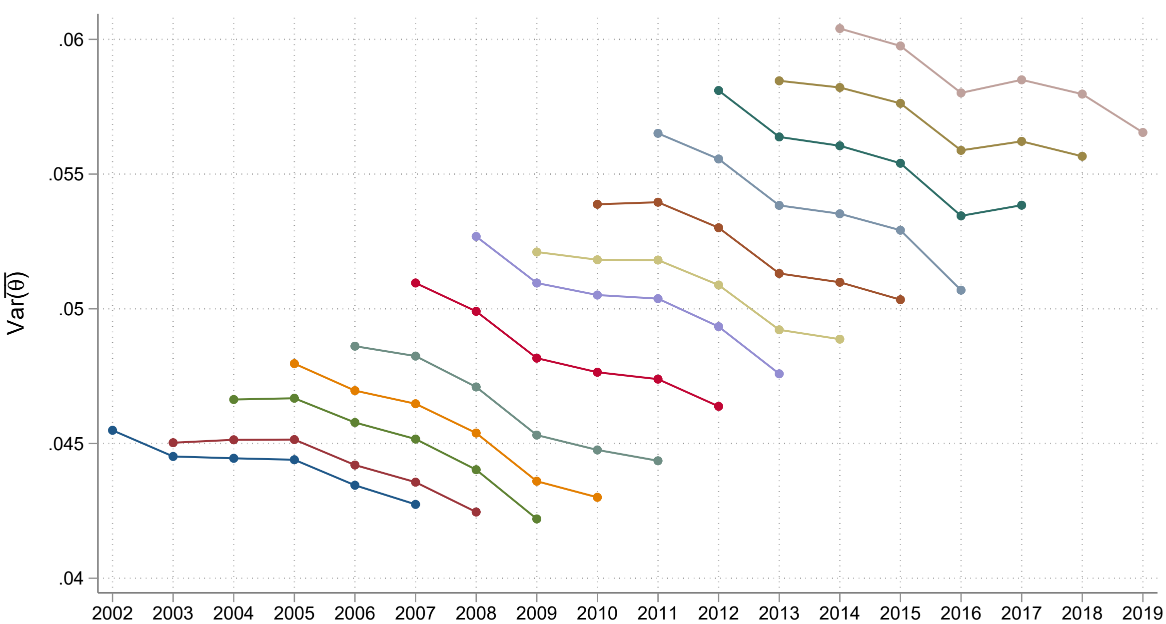
Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Fixed effects estimates by a panel of six-year periods split sample bias correction with firm splitting (Panel a). In Panel b, we further include year fixed effects and cubic function of (age-40). For each panel, log wage variance decomposition is computed separately for each year.

FIGURE A9. Segregation - Rolling Panels

A. Baseline



B. Year fixed effects and cubic function of (age-40)



Note: All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included. Fixed effects estimates by a panel of six-year periods split sample bias correction with firm splitting (Panel a). In Panel b, we further include year fixed effects and cubic function of (age-40). For each panel, log wage variance decomposition is computed separately for each year.

Appendix F. Fixed effects regressions

We regressed at the observation level the estimated firms and workers fixed effects (without correction) on occupation, sex, and industry (Table A13). This is similar to a firm-level regression of FFE on industry, occupational shares, and proportion of women, weighted by worker-year observations, or an individual-level regression of WFE on occupation, sex, and industry shares (in a given year), weighted by worker-year observations. Estimates are relatively stable from period 1 to period 3. We do not observe strong patterns where a change in the association between an occupation, industry, or sex and each fixed effect would imply an increase in sorting, though these results are mostly descriptive. They do not account for interactions between terms (if, for instance, women's shares in industries changed over the period) and compositional effects (for instance, because sex is associated with sorting, an increase in the share of the minority would increase sorting).

Table A13. Fixed effects regressions on occupation, sex and industry

		2002-2007							2014-2019						
		freq. (%)	WFE	SE		FFE	SE		freq. (%)	WFE	SE		FFE	SE	
Occupation															
20	Entrepreneurs	0.91	0.962	0.008	***	0.001	0.008		0.90	0.794	0.008	***	-0.010	0.007	
32	Professionals	0.83	0.802	0.009	***	0.076	0.009	***	1.68	0.628	0.022	***	0.013	0.010	
35	Artists and media professionals	0.48	0.516	0.010	***	0.121	0.010	***	0.44	0.456	0.015	***	0.071	0.010	***
37	Managers	8.93	0.708	0.010	***	0.091	0.008	***	10.71	0.647	0.008	***	0.082	0.007	***
38	Engineers	7.23	0.646	0.008	***	0.098	0.008	***	9.24	0.589	0.009	***	0.080	0.007	***
42	Primary school teachers	0.85	0.277	0.011	***	0.018	0.010	*	1.39	0.279	0.008	***	-0.014	0.007	*
43	Health and social workers	3.24	0.312	0.006	***	0.043	0.008	***	3.98	0.308	0.008	***	0.024	0.007	***
45	Public admin. intermediates	0.04	0.230	0.020	***	-0.007	0.010		0.05	0.197	0.030	***	0.074	0.012	***
46	Business admin. intermediates	9.32	0.261	0.006	***	0.073	0.008	***	8.28	0.240	0.007	***	0.047	0.007	***
47	Technicians	5.75	0.193	0.007	***	0.078	0.008	***	6.08	0.178	0.008	***	0.051	0.007	***
48	Intermediate supervisors	3.37	0.197	0.006	***	0.058	0.008	***	2.76	0.171	0.008	***	0.048	0.007	***
52	Public admin. clerks	2.57	0.006	0.006		0.045	0.008	***	3.28	0.013	0.008	*	0.031	0.007	***
53	Security agents	0.92	-0.073	0.011	***	-0.005	0.013		0.96	-0.063	0.010	***	-0.032	0.011	***
54	Business admin. clerks	11.76	0.105	0.006	***	0.047	0.007	***	10.78	0.114	0.007	***	0.020	0.007	***
55	Retail salespersons	6.48	0	ref		0	ref		6.87	0	ref		0	ref	
56	Personal service employees	4.38	-0.057	0.005	***	-0.014	0.008	*	4.86	-0.034	0.007	***	-0.007	0.007	
62	Skilled manufacturing workers	12.93	-0.035	0.008	***	0.049	0.008	***	8.63	-0.028	0.007	***	0.041	0.007	***
63	Skilled artisans	4.80	-0.040	0.006	***	0.011	0.007		5.22	-0.017	0.007	**	-0.006	0.007	
64	Drivers	3.98	-0.108	0.011	***	-0.022	0.009	**	4.08	-0.149	0.014	***	-0.035	0.008	***
65	Handling, transport skilled workers	2.92	-0.092	0.007	***	0.055	0.008	***	2.56	-0.085	0.008	***	0.041	0.007	***
67	Unskilled manufacturing workers	5.58	-0.133	0.008	***	0.032	0.009	***	3.93	-0.122	0.008	***	0.018	0.007	**
68	Unkilled artisans	2.55	-0.162	0.006	***	-0.040	0.010	***	2.80	-0.172	0.008	***	-0.038	0.007	***
69	Farm workers	0.20	-0.093	0.010	***	-0.065	0.010	***	0.51	-0.058	0.008	***	-0.072	0.008	***
Industry															
AC	Farming and industry	29.57	0.020	0.006	***	0.057	0.004	***	20.95	0.037	0.004	***	0.041	0.003	***
DE	Utilities	2.52	0.056	0.011	***	0.086	0.005	***	2.41	0.089	0.027	***	0.093	0.010	***
F	Construction	6.14	0.022	0.003	***	0.066	0.003	***	6.80	0.027	0.003	***	0.064	0.003	***
G	Commerce	17.31	0	ref		0	ref		17.54	0	ref		0	ref	
H	Transport	5.92	0.039	0.015	***	0.025	0.007	***	6.02	0.039	0.017	**	0.025	0.005	***
I	Hotels, tourism, catering	2.98	0.004	0.005		-0.020	0.005	***	3.56	-0.023	0.006	***	-0.025	0.004	***
J	Media	3.65	0.006	0.007		0.078	0.006	***	4.90	0.019	0.015		0.027	0.007	***
K	Financial services	7.37	0.037	0.010	***	0.115	0.009	***	5.71	0.072	0.007	***	0.099	0.006	***
LM	Real estate, professional services	5.69	0.022	0.004	***	0.016	0.005	***	8.46	0.082	0.005	***	0.004	0.004	
N	Administrative services	4.36	-0.006	0.004		0.002	0.014		6.56	-0.020	0.004	***	-0.005	0.007	
OPQ	Health, educ. and public admin.	12.49	-0.012	0.003	***	-0.030	0.003	***	14.28	-0.058	0.005	***	-0.076	0.003	***
R	Arts and recreation	0.51	0.034	0.014	**	0.018	0.012		0.99	0.009	0.008		-0.008	0.007	
STU	Other	1.48	-0.023	0.005	***	-0.032	0.005	***	1.81	-0.032	0.006	***	-0.054	0.005	***
Sex															
	Women	40.94	-0.140	0.002	***	-0.004	0.001	***	43.17	-0.120	0.001	***	-0.001	0.000	*

Note: Significance levels : $p < 0.01$ ***; $p < 0.05$ **; $p < 0.10$ *
All firms and individuals in firms with at least 1 employee are included. Only individuals employed for at least 360 days by the same firm during the year are included for a given year. Individuals and firms in public administration are not included.

Liste des documents de travail de la Direction des Études et Synthèses Économiques

G 9001 J. FAYOLLE et M. FLEURBAEY
Accumulation, profitabilité et endettement des entreprises

G 9002 H. ROUSSE
Détection et effets de la multicolinéarité dans les modèles linéaires ordinaires - Un prolongement de la réflexion de BELSLEY, KUH et WELSCH

G 9003 P. RALLE et J. TOUJAS-BERNATE
Indexation des salaires : la rupture de 1983

G 9004 D. GUELLEC et P. RALLE
Compétitivité, croissance et innovation de produit

G 9005 P. RALLE et J. TOUJAS-BERNATE
Les conséquences de la désindexation. Analyse dans une maquette prix-salaires

G 9101 Équipe AMADEUS
Le modèle AMADEUS - Première partie - Présentation générale

G 9102 J.L. BRILLET
Le modèle AMADEUS - Deuxième partie - Propriétés variantielles

G 9103 D. GUELLEC et P. RALLE
Endogenous growth and product innovation

G 9104 H. ROUSSE
Le modèle AMADEUS - Troisième partie - Le commerce extérieur et l'environnement international

G 9105 H. ROUSSE
Effets de demande et d'offre dans les résultats du commerce extérieur manufacturé de la France au cours des deux dernières décennies

G 9106 B. CREPON
Innovation, taille et concentration : causalités et dynamiques

G 9107 B. AMABLE et D. GUELLEC
Un panorama des théories de la croissance endogène

G 9108 M. GLAUDE et M. MOUTARDIER
Une évaluation du coût direct de l'enfant de 1979 à 1989

G 9109 P. RALLE et alii
France - Allemagne : performances économiques comparées

G 9110 J.L. BRILLET
Micro-DMS ***NON PARU***

G 9111 A. MAGNIER
Effets accélérateur et multiplicateur en France depuis 1970 : quelques résultats empiriques

G 9112 B. CREPON et G. DUREAU
Investissement en recherche-développement : analyse de causalités dans un modèle d'accélérateur généralisé

G 9113 J.L. BRILLET, H. ERKEL-ROUSSE, J. TOUJAS-BERNATE
"France-Allemagne Couplées" - Deux économies vues par une maquette macro-économétrique

G 9201 W.J. ADAMS, B. CREPON, D. ENCAOUA
Choix technologiques et stratégies de dissuasion d'entrée

G 9202 J. OLIVEIRA-MARTINS, J. TOUJAS-BERNATE
Macro-economic import functions with imperfect competition - An application to the E.C. Trade

G 9203 I. STAPIC
Les échanges internationaux de services de la France dans le cadre des négociations multilatérales du GATT
Juin 1992 (1ère version)
Novembre 1992 (version finale)

G 9204 P. SEVESTRE
L'économétrie sur données individuelles-temporelles. Une note introductive

G 9205 H. ERKEL-ROUSSE
Le commerce extérieur et l'environnement international dans le modèle AMADEUS (réestimation 1992)

G 9206 N. GREENAN et D. GUELLEC
Coordination within the firm and endogenous growth

G 9207 A. MAGNIER et J. TOUJAS-BERNATE
Technology and trade: empirical evidences for the major five industrialized countries

G 9208 B. CREPON, E. DUGUET, D. ENCAOUA et P. MOHNEN
Cooperative, non cooperative R & D and optimal patent life

G 9209 B. CREPON et E. DUGUET
Research and development, competition and innovation: an application of pseudo maximum likelihood methods to Poisson models with heterogeneity

G 9301 J. TOUJAS-BERNATE
Commerce international et concurrence imparfaite : développements récents et implications pour la politique commerciale

G 9302 Ch. CASES
Durées de chômage et comportements d'offre de travail : une revue de la littérature

G 9303 H. ERKEL-ROUSSE
Union économique et monétaire : le débat économique

G 9304 N. GREENAN - D. GUELLEC / G. BROUSSAUDIER - L. MIOTTI
Innovation organisationnelle, dynamisme technologique et performances des entreprises

G 9305 P. JAILLARD
Le traité de Maastricht : présentation juridique et historique

G 9306 J.L. BRILLET
Micro-DMS : présentation et propriétés

G 9307 J.L. BRILLET
Micro-DMS - variantes : les tableaux

G 9308 S. JACOBZONE
Les grands réseaux publics français dans une perspective européenne

G 9309 L. BLOCH - B. CŒURÉ
Profitabilité de l'investissement productif et transmission des chocs financiers

G 9310 J. BOURDIEU - B. COLIN-SEDILLOT
Les théories sur la structure optimale du capital : quelques points de repère

G 9311 J. BOURDIEU - B. COLIN-SEDILLOT
Les décisions de financement des entreprises françaises : une évaluation empirique des théories de la structure optimale du capital

G 9312 L. BLOCH - B. CŒURÉ
Q de Tobin marginal et transmission des chocs financiers

G 9313 Équipes Amadeus (INSEE), Banque de France, Métric (DP)
Présentation des propriétés des principaux modèles macroéconomiques du Service Public

G 9314 B. CREPON - E. DUGUET
Research & Development, competition and innovation

G 9315 B. DORMONT
Quelle est l'influence du coût du travail sur l'emploi ?

G 9316 D. BLANCHET - C. BROUSSE
Deux études sur l'âge de la retraite

G 9317 D. BLANCHET
Répartition du travail dans une population hétérogène : deux notes

G 9318 D. EYSSARTIER - N. PONTY
AMADEUS - an annual macro-economic model for the medium and long term

G 9319 G. CETTE - Ph. CUNÉO - D. EYSSARTIER - J. GAUTIÉ
Les effets sur l'emploi d'un abaissement du coût du travail des jeunes

G 9401 D. BLANCHET
Les structures par âge importent-elles ?

G 9402 J. GAUTIÉ
Le chômage des jeunes en France : problème de formation ou phénomène de file d'attente ? Quelques éléments du débat

G 9403 P. QUIRION
Les déchets en France : éléments statistiques et économiques

G 9404 D. LADIRAY - M. GRUN-REHOMME
Lissage par moyennes mobiles - Le problème des extrémités de série

G 9405 V. MAILLARD
Théorie et pratique de la correction des effets de jours ouvrables

G 9406 F. ROSENWALD
La décision d'investir

G 9407 S. JACOBZONE
Les apports de l'économie industrielle pour définir la stratégie économique de l'hôpital public

G 9408 L. BLOCH, J. BOURDIEU, B. COLIN-SEDILLOT, G. LONGUEVILLE
Du défaut de paiement au dépôt de bilan : les banquiers face aux PME en difficulté

G 9409 D. EYSSARTIER, P. MAIRE
Impacts macro-économiques de mesures d'aide au logement - quelques éléments d'évaluation

G 9410 F. ROSENWALD
Suivi conjoncturel de l'investissement

G 9411 C. DEFEUILLEY - Ph. QUIRION
Les déchets d'emballages ménagers : une analyse économique des politiques française et allemande

G 9412 J. BOURDIEU - B. CŒURÉ - B. COLIN-SEDILLOT
Investissement, incertitude et irréversibilité
Quelques développements récents de la théorie de l'investissement

G 9413 B. DORMONT - M. PAUCHET
L'évaluation de l'élasticité emploi-salaire dépend-elle des structures de qualification ?

G 9414 I. KABLA
Le Choix de breveter une invention

G 9501 J. BOURDIEU - B. CŒURÉ - B. SEDILLOT
Irreversible Investment and Uncertainty: When is there a Value of Waiting?

G 9502 L. BLOCH - B. CŒURÉ
Imperfections du marché du crédit, investissement des entreprises et cycle économique

G 9503 D. GOUX - E. MAURIN
Les transformations de la demande de travail par qualification en France
Une étude sur la période 1970-1993

G 9504 N. GREENAN
Technologie, changement organisationnel, qualifications et emploi : une étude empirique sur l'industrie manufacturière

G 9505 D. GOUX - E. MAURIN
Persistance des hiérarchies sectorielles de salaires: un réexamen sur données françaises

G 9505 Bis D. GOUX - E. MAURIN
Persistence of inter-industry wages differentials: a reexamination on matched worker-firm panel data

G 9506 S. JACOBZONE
Les liens entre RMI et chômage, une mise en perspective
NON PARU - article sorti dans Économie et Prévision n° 122 (1996) - pages 95 à 113

G 9507 G. CETTE - S. MAHFOUZ
Le partage primaire du revenu
Constat descriptif sur longue période

G 9601 Banque de France - CEPREMAP - Direction de la Prévision - Érasme - INSEE - OFCE
Structures et propriétés de cinq modèles macro-économiques français

G 9602 Rapport d’activité de la DESE de l’année 1995

G 9603 J. BOURDIEU - A. DRAZNIEKS
L’octroi de crédit aux PME : une analyse à partir d’informations bancaires

G 9604 A. TOPIOL-BENSAÏD
Les implantations japonaises en France

G 9605 P. GENIER - S. JACOBZONE
Comportements de prévention, consommation d’alcool et tabagie : peut-on parler d’une gestion globale du capital santé ?
Une modélisation microéconométrique empirique

G 9606 C. DOZ - F. LENGLART
Factor analysis and unobserved component models: an application to the study of French business surveys

G2005/08 C. L'ANGEVIN - N. LAÏB
Éducation et croissance en France et dans un panel de 21 pays de l'OCDE

G2005/09 N. FERRARI
Prévoir l'investissement des entreprises
Un indicateur des révisions dans l'enquête de conjoncture sur les investissements dans l'industrie.

G2005/10 P.-O. BEFFY - C. L'ANGEVIN
Chômage et boucle prix-salaires :
apport d'un modèle « qualifiés/peu qualifiés »

G2005/11 B. HEITZ
A two-states Markov-switching model of inflation in France and the USA: credible target VS inflation spiral

G2005/12 O. BIAU - H. ERKEL-ROUSSE - N. FERRARI
Réponses individuelles aux enquêtes de conjoncture et prévision macroéconomiques : Exemple de la prévision de la production manufacturière

G2005/13 P. AUBERT - D. BLANCHET - D. BLAU
The labour market after age 50: some elements of a Franco-American comparison

G2005/14 D. BLANCHET - T. DEBRAND - P. DOURGNON - P. POLLET
L'enquête SHARE : présentation et premiers résultats de l'édition française

G2005/15 M. DUÉE
La modélisation des comportements démographiques dans le modèle de microsimulation DESTINIE

G2005/16 H. RAOUI - S. ROUX
Étude de simulation sur la participation versée aux salariés par les entreprises

G2006/01 C. BONNET - S. BUFFETEAU - P. GODEFROY
Disparités de retraite de droit direct entre hommes et femmes : quelles évolutions ?

G2006/02 C. PICART
Les gazelles en France

G2006/03 P. AUBERT - B. CRÉPON -P. ZAMORA
Le rendement apparent de la formation continue dans les entreprises : effets sur la productivité et les salaires

G2006/04 J.-F. OUVRARD - R. RATHELOT
Demographic change and unemployment:
what do macroeconometric models predict?

G2006/05 D. BLANCHET - J.-F. OUVRARD
Indicateurs d'engagements implicites des systèmes de retraite : chiffrages, propriétés analytiques et réactions à des chocs démographiques types

G2006/06 G. BIAU - O. BIAU - L. ROUVIERE
Nonparametric Forecasting of the Manufacturing Output Growth with Firm-level Survey Data

G2006/07 C. AFSA - P. GIVORD
Le rôle des conditions de travail dans les absences pour maladie

G2006/08 P. SILLARD - C. L'ANGEVIN - S. SERRAVALLE
Performances comparées à l'exportation de la France et de ses principaux partenaires
Une analyse structurelle sur 12 ans

G2006/09 X. BOUTIN - S. QUANTIN
Une méthodologie d'évaluation comptable du coût du capital des entreprises françaises : 1984-2002

G2006/10 C. AFSA
L'estimation d'un coût implicite de la pénibilité du travail chez les travailleurs âgés

G2006/11 C. LELARGE
Les entreprises (industrielles) françaises sont-elles à la frontière technologique ?

G2006/12 O. BIAU - N. FERRARI
Théorie de l'opinion
Faut-il pondérer les réponses individuelles ?

G2006/13 A. KOUBI - S. ROUX
Une réinterprétation de la relation entre productivité et inégalités salariales dans les entreprises

G2006/14 R. RATHELOT - P. SILLARD
The impact of local taxes on plants location decision

G2006/15 L. GONZALEZ - C. PICART
Diversification, recentrage et poids des activités de support dans les groupes (1993-2000)

G2007/01 D. SRAER
Allègements de cotisations patronales et dynamique salariale

G2007/02 V. ALBOUY - L. LEQUIEN
Les rendements non monétaires de l'éducation : le cas de la santé

G2007/03 D. BLANCHET - T. DEBRAND
Aspiration à la retraite, santé et satisfaction au travail : une comparaison européenne

G2007/04 M. BARLET - L. CRUSSON
Quel impact des variations du prix du pétrole sur la croissance française ?

G2007/05 C. PICART
Flux d'emploi et de main-d'œuvre en France : un réexamen

G2007/06 V. ALBOUY - C. TAVAN
Massification et démocratisation de l'enseignement supérieur en France

G2007/07 T. LE BARBANCHON
The Changing response to oil price shocks in France: a DSGE type approach

G2007/08 T. CHANEY - D. SRAER - D. THESMAR
Collateral Value and Corporate Investment
Evidence from the French Real Estate Market

G2007/09 J. BOISSINOT
Consumption over the Life Cycle: Facts for France

G2007/10 C. AFSA
Interpréter les variables de satisfaction : l'exemple de la durée du travail

G2007/11 R. RATHELOT - P. SILLARD
Zones Franches Urbaines : quels effets sur l'emploi salarié et les créations d'établissements ?

G2007/12 V. ALBOUY - B. CRÉPON
Aléa moral en santé : une évaluation dans le cadre du modèle causal de Rubin

G2008/01 C. PICART
Les PME françaises : rentables mais peu dynamiques

G2008/02 P. BISCOURP - X. BOUTIN - T. VERGÉ
The Effects of Retail Regulations on Prices
Evidence form the Loi Galland

G2008/03 Y. BARBESOL - A. BRIANT
Économies d'agglomération et productivité des entreprises : estimation sur données individuelles françaises

G2008/04 D. BLANCHET - F. LE GALLO
Les projections démographiques : principaux mécanismes et retour sur l'expérience française

G2008/05 D. BLANCHET - F. TOUTLEMONDE
Évolutions démographiques et déformation du cycle de vie active : quelles relations ?

G2008/06 M. BARLET - D. BLANCHET - L. CRUSSON
Internationalisation et flux d'emplois : que dit une approche comptable ?

G2008/07 C. LELARGE - D. SRAER - D. THESMAR
Entrepreneurship and Credit Constraints - Evidence from a French Loan Guarantee Program

G2008/08 X. BOUTIN - L. JANIN
Are Prices Really Affected by Mergers?

G2008/09 M. BARLET - A. BRIANT - L. CRUSSON
Concentration géographique dans l'industrie manufacturière et dans les services en France : une approche par un indicateur en continu

G2008/10 M. BEFFY - É. COUDIN - R. RATHELOT
Who is confronted to insecure labor market histories? Some evidence based on the French labor market transition

G2008/11 M. ROGER - E. WALRAET
Social Security and Well-Being of the Elderly: the Case of France

G2008/12 C. AFSA
Analyser les composantes du bien-être et de son évolution
Une approche empirique sur données individuelles

G2008/13 M. BARLET - D. BLANCHET - T. LE BARBANCHON
Microsimuler le marché du travail : un prototype

G2009/01 P.-A. PIONNIER
Le partage de la valeur ajoutée en France, 1949-2007

G2009/02 Laurent CLAVEL - Christelle MINODIER
A Monthly Indicator of the French Business Climate

G2009/03 H. ERKEL-ROUSSE - C. MINODIER
Do Business Tendency Surveys in Industry and Services Help in Forecasting GDP Growth?
A Real-Time Analysis on French Data

G2009/04 P. GIVORD - L. WILNER
Les contrats temporaires : trappe ou marchepied vers l'emploi stable ?

G2009/05 G. LALANNE - P.-A. PIONNIER - O. SIMON
Le partage des fruits de la croissance de 1950 à 2008 : une approche par les comptes de surplus

G2009/06 L. DAVEZIES - X. D'HAULTFOEUILLE
Faut-il pondérer ?… Ou l'éternelle question de l'économètre confronté à des données d'enquête

G2009/07 S. QUANTIN - S. RASPILLER - S. SERRAVALLE
Commerce intragroupe, fiscalité et prix de transferts : une analyse sur données françaises

G2009/08 M. CLERC - V. MARCUS
Élasticités-prix des consommations énergétiques des ménages

G2009/09 G. LALANNE - E. POULIQUEN - O. SIMON
Prix du pétrole et croissance potentielle à long terme

G2009/10 D. BLANCHET - J. LE CACHEUX - V. MARCUS
Adjusted net savings and other approaches to sustainability: some theoretical background

G2009/11 V. BELLAMY - G. CONSALES - M. FESSEAU - S. LE LAIDIER - É. RAYNAUD
Une décomposition du compte des ménages de la comptabilité nationale par catégorie de ménage en 2003

G2009/12 J. BARDAJI - F. TALLET
Detecting Economic Regimes in France: a Qualitative Markov-Switching Indicator Using Mixed Frequency Data

G2009/13 R. AEBERHARDT - D. FOUGÈRE - R. RATHELOT
Discrimination à l'embauche : comment exploiter les procédures de *testing* ?

G2009/14 Y. BARBESOL - P. GIVORD - S. QUANTIN
Partage de la valeur ajoutée, approche par données microéconomiques

G2009/15 I. BUONO - G. LALANNE
The Effect of the Uruguay round on the Intensive and Extensive Margins of Trade

G2010/01 C. MINODIER
Avantages comparés des séries des premières valeurs publiées et des séries des valeurs révisées - Un exercice de prévision en temps réel de la croissance trimestrielle du PIB en France

G2010/02 V. ALBOUY - L. DAVEZIES - T. DEBRAND
Health Expenditure Models: a Comparison of Five Specifications using Panel Data

G2010/03 C. KLEIN - O. SIMON
Le modèle MÉSANGE réestimé en base 2000
Tome 1 – Version avec volumes à prix constants

G2010/04 M.-É. CLERC - É. COUDIN
L'IPC, miroir de l'évolution du coût de la vie en France ? Ce qu'apporte l'analyse des courbes d'Engel

G2010/05 N. CECI-RENAUD - P.-A. CHEVALIER
Les seuils de 10, 20 et 50 salariés : impact sur la taille des entreprises françaises

G2010/06 R. AEBERHARDT - J. POUGET
National Origin Differences in Wages and Hierarchical Positions - Evidence on French Full-Time Male Workers from a matched Employer-Employee Dataset

G2010/07 S. BLASCO - P. GIVORD
Les trajectoires professionnelles en début de vie active : quel impact des contrats temporaires ?

G2010/08 P. GIVORD
Méthodes économétriques pour l'évaluation de politiques publiques

G2010/09 P.-Y. CABANNES - V. LAPÈGUE - E. POULIQUEN - M. BEFFY - M. GAINI
Quelle croissance de moyen terme après la crise ?

G2010/10 I. BUONO - G. LALANNE
La réaction des entreprises françaises à la baisse des tarifs douaniers étrangers

G2010/11 R. RATHELOT - P. SILLARD
L'apport des méthodes à noyaux pour mesurer la concentration géographique - Application à la concentration des immigrés en France de 1968 à 1999

G2010/12 M. BARATON - M. BEFFY - D. FOUGÈRE
Une évaluation de l'effet de la réforme de 2003 sur les départs en retraite - Le cas des enseignants du second degré public

G2010/13 D. BLANCHET - S. BUFFETEAU - E. CRENNER S. LE MINEZ
Le modèle de microsimulation Destinie 2 : principales caractéristiques et premiers résultats

G2010/14 D. BLANCHET - E. CRENNER
Le bloc retraites du modèle Destinie 2 : guide de l'utilisateur

G2010/15 M. BARLET - L. CRUSSON - S. DUPUCH - F. PUECH
Des services échangés aux services échangeables : une application sur données françaises

G2010/16 M. BEFFY - T. KAMIONKA
Public-private wage gaps: is civil-servant human capital sector-specific?

G2010/17 P.-Y. CABANNES - H. ERKEL-ROUSSE - G. LALANNE - O. MONSO - E. POULIQUEN
Le modèle Mésange réestimé en base 2000
Tome 2 - Version avec volumes à prix chaînés

G2010/18 R. AEBERHARDT - L. DAVEZIES
Conditional Logit with one Binary Covariate: Link between the Static and Dynamic Cases

G2011/01 T. LE BARBANCHON - B. OURLIAC - O. SIMON
Les marchés du travail français et américain face aux chocs conjoncturels des années 1986 à 2007 : une modélisation DSGE

G2011/02 C. MARBOT
Une évaluation de la réduction d'impôt pour l'emploi de salariés à domicile

G2011/03 L. DAVEZIES
Modèles à effets fixes, à effets aléatoires, modèles mixtes ou multi-niveaux : propriétés et mises en œuvre des modélisations de l'hétérogénéité dans le cas de données groupées

G2011/04 M. ROGER - M. WASMER
Heterogeneity matters: labour productivity differentiated by age and skills

G2011/05 J.-C. BRICONGNE - J.-M. FOURNIER V. LAPÈGUE - O. MONSO
De la crise financière à la crise économique L'impact des perturbations financières de 2007 et 2008 sur la croissance de sept pays industrialisés

G2011/06 P. CHARNOZ - É. COUDIN - M. GAINI
Wage inequalities in France 1976-2004: a quantile regression analysis

G2011/07 M. CLERC - M. GAINI - D. BLANCHET
Recommendations of the Stiglitz-Sen-Fitoussi Report: A few illustrations

G2011/08 M. BACHELET - M. BEFFY - D. BLANCHET
Projeter l'impact des réformes des retraites sur l'activité des 55 ans et plus : une comparaison de trois modèles

G2011/09 C. LOUVOT-RUNAVOT
L'évaluation de l'activité dissimulée des entreprises sur la base des contrôles fiscaux et son insertion dans les comptes nationaux

G2011/10 A. SCHREIBER - A. VICARD
La tertiarisation de l'économie française et le ralentissement de la productivité entre 1978 et 2008

G2011/11 M.-É. CLERC - O. MONSO - E. POULIQUEN
Les inégalités entre générations depuis le baby-boom

G2011/12 C. MARBOT - D. ROY
Évaluation de la transformation de la réduction d'impôt en crédit d'impôt pour l'emploi de salariés à domicile en 2007

G2011/13 P. GIVORD - R. RATHELOT - P. SILLARD
Place-based tax exemptions and displacement effects: An evaluation of the Zones Franches Urbaines program

G2011/14 X. D'HAULTFOEUILLE - P. GIVORD - X. BOUTIN
The Environmental Effect of Green Taxation: the Case of the French "Bonus/Malus"

G2011/15 M. BARLET - M. CLERC - M. GARNEO - V. LAPÈGUE - V. MARCUS
La nouvelle version du modèle MZE, modèle macroéconométrique pour la zone euro

G2011/16 R. AEBERHARDT - I. BUONO - H. FADINGER
Learning, Incomplete Contracts and Export Dynamics: Theory and Evidence form French Firms

G2011/17 C. KERDRAIN - V. LAPÈGUE
Restrictive Fiscal Policies in Europe: What are the Likely Effects?

G2012/01 P. GIVORD - S. QUANTIN - C. TREVIEN
A Long-Term Evaluation of the First Generation of the French Urban Enterprise Zones

G2012/02 N. CECI-RENAUD - V. COTTET
Politique salariale et performance des entreprises

G2012/03 P. FÉVRIER - L. WILNER
Do Consumers Correctly Expect Price Reductions? Testing Dynamic Behavior

G2012/04 M. GAINI - A. LEDUC - A. VICARD
School as a shelter? School leaving-age and the business cycle in France

G2012/05 M. GAINI - A. LEDUC - A. VICARD
A scarred generation? French evidence on young people entering into a tough labour market

G2012/06 P. AUBERT - M. BACHELET
Disparités de montant de pension et redistribution dans le système de retraite français

G2012/07 R. AEBERHARDT - P GIVORD - C. MARBOT
Spillover Effect of the Minimum Wage in France: An Unconditional Quantile Regression Approach

G2012/08 A. EIDELMAN - F. LANGUMIER - A. VICARD
Prélèvements obligatoires reposant sur les ménages : des canaux redistributifs différents en 1990 et 2010

G2012/09 O. BARGAIN - A. VICARD
Le RMI et son successeur le RSA découragent-ils certains jeunes de travailler ? Une analyse sur les jeunes autour de 25 ans

G2012/10 C. MARBOT - D. ROY
Projections du coût de l'APA et des caractéristiques de ses bénéficiaires à l'horizon 2040 à l'aide du modèle Destinie

G2012/11 A. MAUROUX
Le crédit d'impôt dédié au développement durable : une évaluation économétrique

G2012/12 V. COTTET - S. QUANTIN - V. RÉGNIER
Coût du travail et allègements de charges : une estimation au niveau établissement de 1996 à 2008

G2012/13 X. D'HAULTFOEUILLE - P. FÉVRIER - L. WILNER
Demand Estimation in the Presence of Revenue Management

G2012/14 D. BLANCHET - S. LE MINEZ
Joint macro/micro evaluations of accrued-to-date pension liabilities: an application to French reforms

G2013/01-F1301 T. DEROYON - A. MONTAUT - P-A PIONNIER
Utilisation rétrospective de l'enquête Emploi à une fréquence mensuelle : apport d'une modélisation espace-état

G2013/02-F1302 C. TREVIEN
Habiter en HLM : quel avantage monétaire et quel impact sur les conditions de logement ?

G2013/03 A. POISSONNIER
Temporal disaggregation of stock variables - The Chow-Lin method extended to dynamic models

G2013/04 P. GIVORD - C. MARBOT
Does the cost of child care affect female labor market participation? An evaluation of a French reform of childcare subsidies

G2013/05 G. LAME - M. LEQUIEN - P.-A. PIONNIER
Interpretation and limits of sustainability tests in public finance

G2013/06 C. BELLEGO - V. DORTET-BERNADET
La participation aux pôles de compétitivité : quelle incidence sur les dépenses de R&D et l'activité des PME et ETI ?

G2013/07 P.-Y. CABANNES - A. MONTAUT - P.-A. PIONNIER
Évaluer la productivité globale des facteurs en France : l'apport d'une mesure de la qualité du capital et du travail

G2013/08 R. AEBERHARDT - C. MARBOT
Evolution of Instability on the French Labour Market During the Last Thirty Years

G2013/09 J-B. BERNARD - G. CLÉAUD
Oil price: the nature of the shocks and the impact on the French economy

G2013/10 G. LAME
Was there a « Greenspan Conundrum » in the Euro area?

G2013/11 P. CHONÉ - F. EVAIN - L. WILNER - E. YILMAZ
Introducing activity-based payment in the hospital industry : Evidence from French data

G2013/12 C. GRISLAIN-LETRÉMY
Natural Disasters: Exposure and Underinsurance

G2013/13 P.-Y. CABANNES - V. COTTET - Y. DUBOIS - C. LELARGE - M. SICSIC
French Firms in the Face of the 2008/2009 Crisis

G2013/14 A. POISSONNIER - D. ROY
Households Satellite Account for France in 2010. Methodological issues on the assessment of domestic production

G2013/15 G. CLÉAUD - M. LEMOINE - P.-A. PIONNIER
Which size and evolution of the government expenditure multiplier in France (1980-2010)?

G2014/01 M. BACHELET - A. LEDUC - A. MARINO
Les biographies du modèle Destinie II : rebasage et projection

G2014/02 B. GARBINTI
L'achat de la résidence principale et la création d'entreprises sont-ils favorisés par les donations et héritages ?

G2014/03 N. CECI-RENAUD - P. CHARNOZ - M. GAINI
Évolution de la volatilité des revenus salariaux du secteur privé en France depuis 1968

G2014/04 P. AUBERT
Modalités d'application des réformes des retraites et prévisibilité du montant de pension

G2014/05 C. GRISLAIN-LETRÉMY - A. KATOSSKY
The Impact of Hazardous Industrial Facilities on Housing Prices: A Comparison of Parametric and Semiparametric Hedonic Price Models

G2014//06 J.-M. DAUSSIN-BENICHOU - A. MAUROUX
Turning the heat up. How sensitive are households to fiscal incentives on energy efficiency investments?

G2014/07 C. LABONNE - G. LAMÉ
Credit Growth and Capital Requirements: Binding or Not?

G2014/08 C. GRISLAIN-LETRÉMY et C. TREVIEN
The Impact of Housing Subsidies on the Rental Sector: the French Example

G2014/09 M. LEQUIEN et A. MONTAUT
Croissance potentielle en France et en zone euro : un tour d'horizon des méthodes d'estimation

G2014/10 B. GARBINTI - P. LAMARCHE
Les hauts revenus épargnent-ils davantage ?

G2014/11 D. AUDENAERT - J. BARDAJI - R. LARDEUX - M. ORAND - M. SICSIC
Wage Resilience in France since the Great Recession

G2014/12 F. ARNAUD - J. BOUSSARD - A. POISSONNIER - H. SOUAL

Imposition des couples et des familles : effets budgétaires et redistributifs de l'impôt sur le revenu

G2019/11 K. MOHKAM – O. SIMON
L'empreinte matière de l'économie française : une analyse par matière et catégorie de produits

G2019/12 S. BUNEL – B. HADJIBEYLI
Évaluation du crédit d'impôt innovation

G2019/13 C. BONNET – F. GODET – A. SOLAZ
Gendered economic determinants of couple formation over 50 in France

G2019/14 P. GIVORD – M. SUAREZ CASTILLO
Excellence for all? Heterogeneity in high schools' value-added

G2019/15 G. CETTE – L. KOEHL – T. PHILIPPON
Labor Share in Some Advanced Countries

G2020/01 J. SILHOL – B. VENTELOU
Les zones d'intervention prioritaire reflètent-elles des écarts de pratiques des médecins généralistes ?

G2020/02 B. BOUCHTENIK – R. LARDEUX
The Take-Up of Unemployment Benefit Extensions

G2020/03 J-M. GERMAIN
A Welfare Based Estimate of "Real Feel GDP" for Europe and the USA

G2020/04 J. BOUSSARD – R. LEE
Competition, Profit Share and Concentration

G2020/05 P. PORA
Keep Working and Spend Less?
Collective Childcare and Parental Earnings in France

G2020/06 R. MONIN – M. SUAREZ CASTILLO
Product Switching, market power and distance to core competency

G2020/07 L. WILNER
How do citizens perceive centralization reforms? Evidence from the merger of French regions

G2020/08 L. WILNER
The persistence of subjective well-being: permanent happiness, transitory misery?

G2020/09 J-M. GERMAIN – T. LELLOUCH
Comptabilité économique de la soutenabilité climatique

G2020/10 A. BAUER – M. ROTEMBERG
Tax avoidance in French Firms:
Evidence from the Introduction of a Tax Notch

G2020/11 P. AGHION – A. BERGEAUD
M. LEQUIEN – M. J. MELITZ
The Heterogeneous Impact of Market Size on Innovation: Evidence from French Firm-Level Exports

G2020/12 L. GALIANA – B. SAKAROVITCH
F. SÉMÉCURBE – Z. SMOREDA
Residential segregation, daytime segregation and spatial frictions : an analysis

G2020/13 A. BAUER – J. BOUSSARD
Market Power and Labor Share

G2020/14 A. BAUER – J. BOUSSARD – D. LASHKARI
Information Technology and Returns to Scale

G2020/15 V. LIN – O. MESLIN
Hausse des prix immobiliers et mesure du niveau de vie

G2021/01 N. BECHICHI – G. THEBAULT
Students' Preferences, Capacity Constraints and Post-Secondary Achievements in a Non-Selective System

G2021/02 B. BUREAU – A. DUQUERROY
J. GIORGI – M. LÉ – S. SCOTT – F. VINAS
Une année de crise COVID : impact sur la dynamique des entreprises en France
Une évaluation sur données individuelles

G2021/03 B. BUREAU – A. DUQUERROY
J. GIORGI – M. LÉ – S. SCOTT – F. VINAS
L'impact de la crise sanitaire sur la situation financière des entreprises en 2020 : une analyse sur données individuelles

G2021/04 A. GODZINSKI – M.SUAREZ CASTILLO
Disentangling the effects of air pollutants with many instruments

2021/01 S.QUANTIN – S.BUNEL – C.LENOIR
Évaluation du dispositif Jeune entreprise innovante (JEI) Un exemple d'application du modèle d'analyse de sensibilité de Rosenbaum

2021/02 L. GOBILLON – D. MEURS – S. ROUX
Differences in positions along a hierarchy : Counterfactuals based on an assignment model

2021/03 N. BECHICHI – J. GRENET – G. THEBAULT
Ségrégation à l'entrée des études supérieures en France et en région parisienne : quels effets du passage à Parcoursup ?

2021/04 M. ANDRÉ – O. MESLIN
Et pour quelques appartements de plus :Étude de la propriété immobilière des ménages et du profil redistributif de la taxe foncière

2022/05 S. GEORGES-KOT – D. GOUX – E. MAURIN
The value of leisure synchronozation

2022/06 N. BECHICHI – M. FABRE – T. OLIVIA
Projections de la population active à l'horizon 2080

2022/11 S. QUANTIN – C. WELTER-MÉDÉE
Estimation des montants manquants de versements de TVA :exploitation des données du contrôle fiscal

2022/14 L. BLOCH – B. FAVETTO – A. LAGOUGE – F. SÉDILLOT
Inégalités de rendements et de patrimoine en France en 2017

2022/15 H. GENIN – S. SCOTT
Un portrait de la r´etention de main-d'oeuvre dans l'industrie française : analyse à partir des enquêtes mensuelles de conjoncture

2023/02 D. GOUX – E. MAURIN
On the seventh day you shall do not any work : the winners and losers of Sunday work deregulation

2023/03 P. AGHION – A. BERGEAUD – M. LEQUIEN – M. MELITZ – T. ZUBER
Opposing firm-level responses to the China shock: Output competition versus input supply

2023/04 L. GALIANA – L. WILNER
Private Wealth over the Life-Cycle:
A Meeting between Microsimulation and Structural Approaches

2023/06 M. HILLION – E. MAUGENDRE
L'effet du « jour de carence » sur les absences pour maladie ordinaire, la santé perçue et le recours aux soins à court- terme

2023/07 M.ANDRÉ- J-M GERMAIN – M. SICSIC
'Do I get my money back?': A Broader Approach to Inequality and Redistribution in France With a Monetary Valuation of Public Services

2023/08 R. LAFROGNE-JOUSSIER – J. MARTIN – I. MEJEAN
Supply shocks in supply chains: Evidence from the early lockdown in China

2023/09 J. SILHOL – L. WILNER
Teachers' desired mobility to disadvantaged schools: Do financial incentives matter?

2023/12 P. AGHION - C. ANTONIN - S. BUNEL - X. JARAVEL
Modern Manufacturing Capital, Labor Demand, and Product Market Dynamics: Evidence from France

2023/13 R. LAFROGNE-JOUSSIER – J. MARTIN – I. MEJEAN
Cost Pass-Through and the Rise of Inflation

2023/14 A. BOURGEOIS – F. GERVOIS - R. LAFROGNE-JOUSSIER
Forces et fragilités des tableaux internationaux entrées-sorties pour le calcul de l'empreinte carbone

2023/17 M. ADAM – O. BONNET – E. FIZE
T.LOISEL – M. RAULT – L. WILNER
How does fuel demand respond to price changes?
Quasi-experimental evidence based on high-frequency data

2023/18 S. QUANTIN – C. WELTER-MÉDÉE
Estimation de l'activité dissimulée pour le changement de base 2020 des comptes nationaux

2023/19 T. LOISEL – M. SICSIC
La mobilité des individus le long de l'échelle des revenus en France sur la période 2003-2020