Lifecycle Wages and Human Capital Investments: Selection and Missing Data

Laurent Gobillon* Thierry Magnac† Sébastien Roux‡

June 23, 2025

Abstract

We derive wage equations with individual specific coefficients from a structural model of human capital investment over the life cycle. This model allows for interruptions in labour market participation and deals with missing data and attrition problems. We propose a new framework that deals with missingness at random and is based on factor decompositions that allow for flexible control of selection. Our approach leads to an interactive effect wage specification, which we estimate using long administrative panel data on male wages in the private sector in France. A structural function approach shows that interruptions negatively affect average wages. Interestingly, they also negatively affect the inter-decile range of wages after twenty years. This is only partly due to the fact that interruptions are endogenous.

Keywords: Human capital investment, wage inequalities, factor models, missing data
JEL Codes: C38, D91, I24, J24, J31

*Paris School of Economics-CNRS, CEPR and IZA. Address: Paris School of Economics (PSE), 48 Boulevard Jourdan, 75014 Paris, France. Email: laurent.gobillon@psemail.eu.

†Toulouse School of Economics, Université Toulouse Capitole. Address: 1 Esplanade de l'Université, 31080 Toulouse Cedex 06, France. Email: thierry.magnac@tse-fr.eu

‡INSEE and CREST. INSEE, 88 Avenue Verdier, 92120 Montrouge, France. Email: sebastien.roux@ensae.fr.

1 Introduction

Recent increases in income inequality across OECD countries have spurred researchers to investigate the dynamics of earnings, wages or incomes, and the insurance mechanisms that households can use to protect themselves against earnings shocks when markets are incomplete (see Jappelli and Pistaferri, 2017, for a review). Most contributions analyse interactions between labour earnings processes along the life-cycle and consumption dynamics (Meghir and Pistaferri, 2010), or household labour supply dynamics (Keane and Wasi, 2016). A few of them more narrowly focus on the specification of earnings dynamics that can be studied using long panel survey or administrative data (Guvenen *et al.*, 2021).

Recently, there has been a few attempts to estimate earnings or wage equations à la Mincer (1974) while including lots of heterogeneity as in Browning *et al.* (2012); Polachek *et al.* (2015) or Magnac and Roux (2021). These authors individualize as much as possible earnings processes by estimating sets of individual specific parameters beyond the fixed effects that appear additively in equations commonly estimated with panel data of earnings (Heckman *et al.*, 2006). Their object of interest is the building up of inequalities over the life-cycle and their procedures lead to richer decompositions of life-cycle profiles into permanent and transitory effects. Yet, in empirical analyses, survey or administrative panel data on wages are plagued with missing data and attrition issues. The most common attitude among researchers (e.g. Abrevaya and Donald, 2017) is to select wage histories which are sufficiently long and to treat missing observations in histories as random.

In this paper, we propose a general framework to study wage histories and the evolution of wage inequalities over the life-cycle which accomodates incomplete individual wage trajectories. We address the missing data issue by considering an economically-motivated model of human capital accumulation and participation decisions that drive selection. Our approach builds upon the structural linear model proposed by Magnac *et al.* (2018) for the logarithm of wages over the life-cycle as a function of individual specific parameters.

Our first contribution is to extend the one-sector model to the case of two sectors in which some of the individual parameters become sector-specific. This setting fits empirical analyses in which wages in one sector of the labour market are observed while wages in an alternative sector, if any, are not. This provides us with a way of modelling temporary or permanent attrition in the life-cycle histories of wages. The differential structure of returns and depreciations of human capital investments across sectors creates a wedge between the accumulation processes in human capital in the two sectors (see for instance Blundell *et al.*, 2016, for part-time/full-time evidence). In particular, we expect interruptions in the career to have a sizeable effect on human capital investments (Light and Ureta, 1995). This structural model justifies the introduction of additional linear terms in wage equations reflecting the number of periods spent in the alternative sector. It provides a tractable approach with lots of individual heterogeneity and complements the literature on the impact of potential and actual experiences on wages in an homogeneous set-up (Eckstein and Wolpin, 1989; Altuğ and Miller, 1998).

Our second contribution is an original empirical strategy that deals with selection issues under a specific and testable missing at random (MAR) assumption. We posit a factor structure (e.g. Aakvik *et al.*, 2005) for the residual structural processes of human capital prices, depreciations, and sectoral preferences, and use the structural restrictions on the wage and participation equations. The factor structure implies conditional independence between the wage and sectoral choice equations when conditioning on histories, unobserved factors and factor loadings. Econometric moment restrictions are further supported by a "flat spot" approach introduced heuristically by Heckman *et al.* (1998). Our setting allows a new formalization of this approach that enables the distinction between volumes and prices of human capital, and the separate identification of time, cohort and age effects.

In the empirical analysis, we resort to a long administrative panel dataset collected in France for social security purposes, and which is typical of administrative datasets that can be found in many countries. We study the building up of inequalities of wages in the private sector for cohorts of males who entered in that sector between 1985 and 1992 and were potentially followed until 2011. The other sector gathers all other employment options, e.g. either public or self-employment, as well as non-employment alternatives for which we do not have information on earnings.

Our econometric procedure aims at estimating the reduced form wage equation derived from the structural human capital model. Observed variables in this wage equation comprise a level, trend and curvature in potential experience, as well as the years of interruptions in participation and its associated curvature term. As parameters of those variables are individual specific, the wage equation is a random coefficient model that we estimate with a fixed-effect approach. Additional unobserved factors and factor loadings are introduced to control for selection in a way that preserves the structure of the estimated wage equation. We estimate various specifications adapting Bai (2009)'s least-square method.

To understand the building up of inequalities over the life-cycle, we estimate summaries of the distribution of predicted wage profiles. Those summaries depend on individual specific parameters which converge at rate $\sqrt{T}$, and the incidental parameter issue makes most summary statistics asymptotically biased when N and T tend to infinity (e.g. Fernández-Val and Weidner, 2018). We correct biases using methods proposed by Jochmans and Weidner (2024), and we investigate the small sample properties of these methods in Monte Carlo experiments. We show that variances are not well estimated even when T is greater than 20, and we prefer to measure the dispersion of wages with robust inter-decile ranges.

Results based on our original empirical strategy constitute our third contribution. We first show that returns to experience are significantly biased downwards after 20 years when omitting interruptions and unobserved factors. Second, we demonstrate that most of this bias comes from the influence of interruptions on human capital accumulation. Third, we evaluate counterfactual average structural functions (Blundell and Powell, 2003) obtained by manipulating interruptions. Accordingly, we estimate the causal impact of the duration and timing of interruptions. We show that interruptions have a significant negative effect on average wages. Interestingly, interruptions also have a negative effect on the dispersion of wages after 20 years. In other words, inequalities are lower than in a world in which there would have been no interruptions. We show that this key result comes from the negative correlation between two components of our wage equation: (1) the effects of interruptions and (2) the effects of potential experience.

The outline of the paper is the following. We start with a brief literature review in Section 2. Section 3 describes empirical evidence about the panel data on wages that we use. Section 4 sets up the structural model, and Section 5 presents the identifying restrictions of the econometric model and our estimation strategy. Section 6 defines our counterfactual objects of interest. Finally, Section 7 motivates our specification choice with results for different specifications and Section 8 discusses the results obtained for different counterfactuals and provides robustness checks. Supplementary elements (Tables, Figures and other developments) are relegated into the Supplementary Appendix.

2 Literature review

Earnings dynamics We first discuss the very extensive empirical literature on earnings dynamics. An important part of this literature aims at fitting the empirical covariance structure of (log) earnings over the life-cycle using competing specifications. Broadly speaking most studies assume that data are missing at random while we adopt a conditional-on-factor version of this assumption. Our paper also relates to the estimation of the traditional homogenous Mincer equation (Lagakos *et al.*, 2018). This literature has mostly remained in a linear framework but there has been non-linear alternatives (see e.g. Arellano *et al.*, 2017).

There is also a more economically oriented literature trying to distinguish theories of wage growth, namely human capital, job search or learning by doing (see surveys in Rubinstein and Weiss, 2006; Magnac and Roux, 2021). Our approach is different, since we rely on a human capital model in line with Polachek *et al.* (2015) and Magnac *et al.* (2018) who study earnings or wages over the life-cycle à la Ben-Porath (1967) under different guises. Their models, as well as ours, are with perfect information, whereas a growing literature on learning uses imperfect information setups (see e.g. Bunting *et al.*, 2024). In their specifications, individual specific parameters governing wage equations have a structural economic interpretation, and they can be related to the abilities of individuals to learn and to earn (Browning *et al.*, 1999; Rubinstein and Weiss, 2006). Few papers study multidimensional human capital as surveyed by Deming (2023). We develop below the conditions under which our wage equation of interest can be derived in a multidimensional setting.

Another strand of the literature is interested in consumption smoothing over time, and studies the resulting joint dynamics of wages and consumption (see Alan *et al.*, 2018; Eika *et al.*, 2020; Arellano *et al.*, 2017). To our knowledge, there is no wage and consumption panel data long enough to deal with lots of heterogeneity as we pursue in our setting with many individual effects.

Finally, some papers study the impact on wages of interruptions in participation. Light and Ureta (1995) show that career breaks, in particular their timing, matter empirically, and we obtain similar results. Biewen *et al.* (2018) find that interruptions affect wage inequalities in Germany using a reduced-form approach.

Missing data The pattern of missingness considered in our paper is due to missing outcomes since wages are irregularly observed over time because of interruptions in private sector participation. We adopt a specific MAR assumption based on a factor structure – i.e. Missing At Random Conditional On Factors (MARCOF) – that relies on our economic model and accounts for selection. Assumptions of selection on observables (e.g. Little and Rubin, 2019) are then replaced with ones on both observables and time-invariant unobservables.

Our setting follows a long history in econometrics that started in the 1980s (Heckman and Hotz, 1989) in which selection was controlled for by using fixed effects or random growth models. Carneiro *et al.* (2003) suggested using factor structures, as well as more recently Eisenhauer *et al.* (2015) and Gobillon and Magnac (2016), but their models and extension to a general framework (Fernández-Val *et al.*, 2021) remain static with no life-cycle perspective.

Our approach is an alternative to recent developments away from MAR which rather model the missingness using exclusion restrictions to correct for selection (Arellano and Bonhomme, 2012; Sasaki, 2015). In the literature on sensitivity (e.g. Kline and Santos, 2013), an intermediate "breakdown" solution between MAR and worst case bounds à la Manski is sought so that substantial results remain (just) significant. We cannot use selection-correction methods due to the absence of credible exclusion restrictions in our administrative data. We cannot rely either on sensitivity analyses since the worst case bounds are infinite because our outcomes of interest are unbounded.

Factor models The development of factor models for panel data started with Holtz-Eakin *et al.* (1988), Ahn *et al.* (2001) and Pesaran (2006). We follow Bai (2009) who proposes to minimize a sum of squares objective function, and uses principal component methods and asymptotics in both N and T. More specifically, asymptotic properties of our estimation method are derived by Song (2013) who extends Bai (2009) to the case of individual specific coefficients. We complete the proof of Song (2013) in which a step was missing. Furthermore, recent advances on the estimation of the interactive effect model includes Moon and Weidner (2019) and Beyhum and Gautier (2022) who propose to use an objective function which is convex in contrast with Bai's. We experiment with their objective function that ensures convergence (Hsiao, 2018), and find the same minimizers as with Bai's algorithm.

Because of interactive effects and missing data, we rely on an Expectation Maximization (EM) algorithm for the estimation. Its use has a long tradition in the statistical literature and its properties have been studied by Heyde and Morton (1996) in the case of pseudo-likelihood maximization. Convergence issues of the EM algorithm and conditions that make our algorithm a contractive mapping have been studied by Dominitz and Sherman (2005) and Balakrishnan *et al.* (2017).

3 Empirical Evidence on Wage Profiles

3.1 The data

The data are constructed from the *2011 DADS Grand Format-EDP* panel dataset which merges two different sources, *DADS* (*Déclaration Annuelles des Données Sociales)* on social security and tax records and *EDP (Echantillon Démographique Permanent*) extracted from censuses. We follow over time all individuals born in the first four days of October of an even year, and use information on their jobs between 1985 and 2011 except in 1990 since it is missing in the original data. The information on spells in the public sector, self-employment, unemployment, and non-employment is incomplete, and we focus on job spells in the private sector.

The data record job characteristics, and in particular whether jobs are full-time or part-time as well as earnings and days of work. For every individual, we aggregate earnings and days of work for all full-time jobs within a year, and construct the individual daily wage deflated by the Consumer Price Index (CPI)[1] in every year. Education as measured by diploma is recovered from *EDP* and censuses, and the highest education level is used to group individuals into four categories: high-school drop outs, high-school graduates, some college – two years or less – and college graduates including top engineering schools.

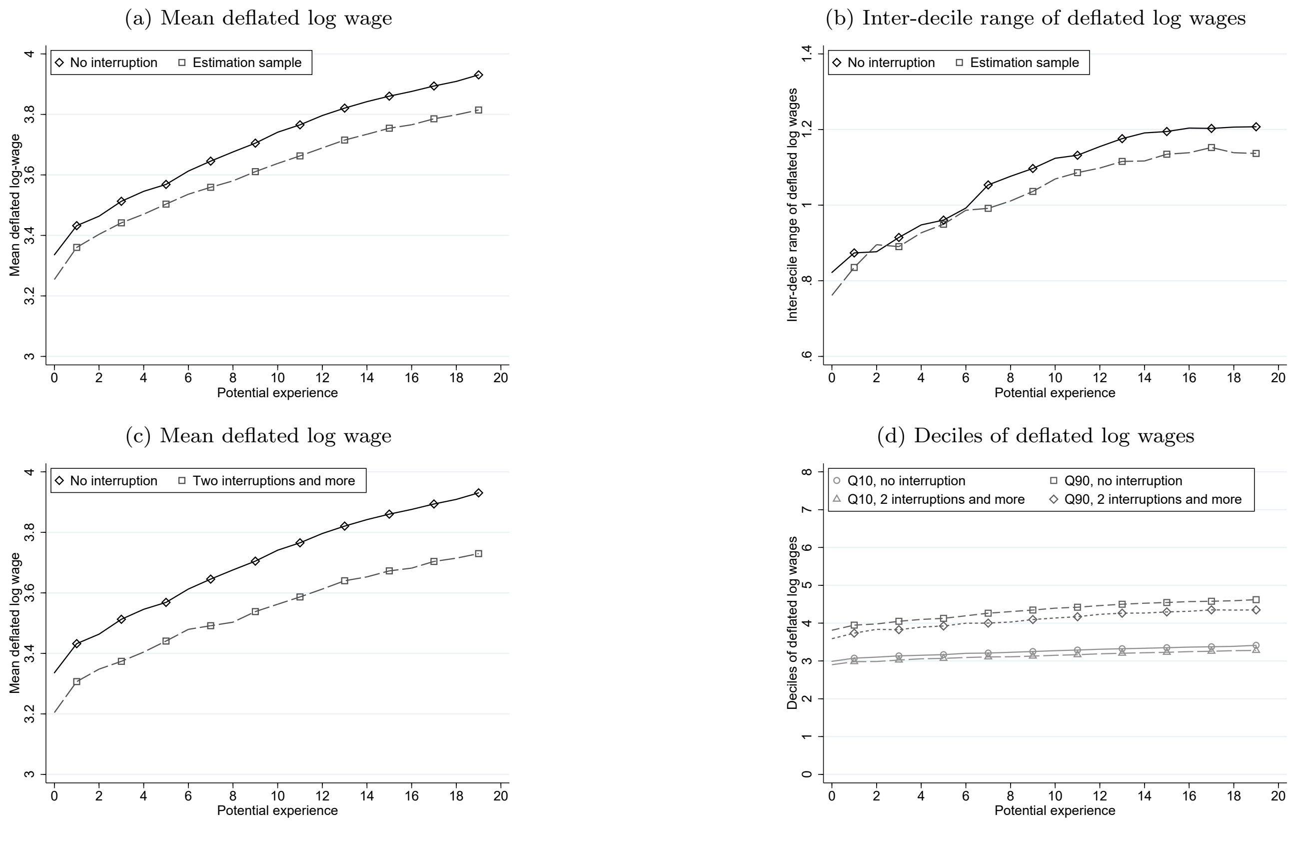
We focus on males who enter the market over the 1985-1992 period and who are $16-30$ years old at the entry date. We recode person-year observations as missing when the daily wage is lower than 80% of the minimum wage and when the number of days of work is lower than 180. A non-missing observed daily wage defines "employment in the private sector", a sector which is denoted as e, while the alternative is denoted n. The year of entry into the panel is defined as the first year an individual works in sector e. We finally select individuals whose wages in the private sector are observed for at least 15 years, and we end up with a working sample of $145,166$ yearly observations involving $7,339$ males. Further details on the sample construction are given in Supplementary Appendix S.1.1.

[1] e.g. the OECD series, at https://data.oecd.org/fr/price/inflation-ipc.htm.

3.2 Descriptive statistics

In Figure 1, we report the profile of statistical summaries of CPI-deflated log wages as a function of potential experience for different samples, where potential experience is defined as the number of years since first entry in the private sector. Our goal is to illustrate selection effects when one does not restrict the analysis to a balanced panel, but also considers individuals with non-employment spells. Selection effects arise from both static selection defined as selection into participation in the private sector at the current date, and dynamic selection defined as the effect of past interruptions on current wages. We display profiles up to 20 years of potential experience since the youngest cohort enters in 1992 and the panel ends in 2011.

Figure 1: Mean and inter-decile of CPI-deflated wages as a function of potential experience



Note: Individuals in our sample are individuals entering the labour market between 1985 and 1992 who are employed at least 15 years in our panel data. We distinguish those with no interruption and those with two interruptions or more. "deflated" means that log-wages are deflated by the CPI.

We first compare results obtained from our working sample and the balanced sample obtained when restricting it to workers with no interruption. Figure 1(a) shows that the mean log-wage is larger and its increase is steeper in the balanced sample than in our working sample. The selection of workers with no interruption thus leaves aside workers occupying less attractive jobs with less favorable career prospects. As displayed in Figure 1(b), the inter-decile range, defined as the difference between the 90% quantile and the 10% quantile, is also larger for the balanced sample and the gap with our working sample is more important at larger values of potential experience.

To further assess selection effects, we also compare two disjoint samples: the balanced sample and the subsample of workers with two interruptions and more. As expected, the mean log-wage for workers with two and more interruptions is lower and with a flatter profile (see Figure 1 (c)). Indeed, these workers have unstable employment trajectories as they go in and out of the private sector (static selection), and past non-employment spells negatively affect their wages (dynamic selection). In our empirical application, we disentangle the roles of static selection and dynamic selection by contrasting the log-wage distributions in the two samples.

We also investigate differences in the first and last deciles between the two samples (Figure 1(d)). Interestingly, whereas both deciles for workers with no interruption are above those of workers with two interruptions and more, the inter-decile range is larger for workers with no interruption. This suggests that workers with no interruption do better, even at the bottom of the wage distribution, but career profiles differ to a larger extent among them.

Interruptions in participation in the private sector make real and potential individual experience different and play an important role in our empirical results. In Table 1, we describe these interruptions excluding attrition periods. For instance, for an individual exiting the private sector in 2007 and absent until the end of the panel in 2011, we do not consider years between 2007 and 2011. This Table shows that the cumulative duration outside the private sector is 2.6 years and the average proportion of years spent in the alternative sector is 11.1%. Interestingly, the frequency of individuals with a single complete spell in the private sector is only 1579/7339 = 21.5%.

We do not have information in our data on the date of exit from the education system, but we can document the existence of short employment spells in the private sector before what we define as the entry year in that sector. Most of these spells are part-time job spells (61%), whereas full-time job spells of fewer than 180 days are far less important (24%). Remaining spells include those in the public sector (7%), in apprenticeship (6%) and those during which wages are unusually low, i.e. below 80% of the minimum wage (1%). In robustness checks below, we evaluate the impact of changing the definition of the entry year.

Table 1: Descriptive statistics on interruptions

Number of interruptions	Number of individuals	Number of observations	Interruption frequency	Cumulated duration in interruptions
All	7339	145166	0.111	2.6
0	1579	34218	0.000	0.0
1	2558	51515	0.086	1.9
2	1925	36636	0.156	3.5
3	870	15778	0.213	4.9
4	316	5501	0.266	6.3
5+	91	1518	0.310	7.5

Note: For a given individual, observations after the last year employed in the private sector are ignored. "Interruption frequency" is the number of years during which an individual is in the alternative sector divided by the number of years between entry and last observation in the private sector (averaged over our estimation sample). "Cumulated duration in interruptions" is the number of years spent in the alternative sector. For instance, consider an individual entering the private sector in 1991 and being last observed in that sector in 2012. Assume that he experiences interruptions in 1993 and 1995. In that case, his proportion in interruption is equal to $2/(2012 - 1991 + 1) = .09$ and his cumulative duration in interruptions is 2.

Regarding years spent outside the private sector between entry and exit (i.e. the last year individuals are observed in the private sector), information in the data shows that individuals very unfrequently work in the public sector (only 0.12% of the time spent outside the private sector). This means that we are not missing much by not distinguishing public sector spells from other spells in the alternative sector. Nonetheless, after the exit year, the time spent in the public sector is on average much more important (29.7% of the years remaining until the end of the panel). Still, this does not affect our analysis since, in our estimations, we use available information until exit only.

4 The economic model

In this section, we set up the model, analyze its structural predictions and derive the reduced form to be brought to the data. Notation is summarized in Appendix A. We then discuss the robustness of our predicted reduced-form wage equation when loosening our assumptions.

4.1 Set up

We start with the description of human capital accumulation that extends the framework of Magnac *et al.* (2018) to two distinct sectors. We then present the timing of decisions, and define value functions. We end up with the description of terminal conditions.

4.1.1 Human capital accumulation in two sectors

An individual, indexed by i, chooses to participate, in each year t, in one of two labour market sectors that are either the private sector (n) or the alternative sector (n). This choice is denoted $s_{i,t} \in \{e, n\}$, and the year of entry in the private sector is denoted, $t_{i,0}$.[2]

Individual wages in the private sector ($s = e$), or a wage-equivalent notion in the alternative sector ($s = n$), are written as:

$$w_{i,t}^{s} = \exp(\delta_{i,t}^{s}) H_{i,t} \exp(-\tau_{i,t}^{s}), \tag{1}$$

in which $H_{i,t}$ is a single dimensional stock of human capital at the beginning of year t. We discuss below the robustness of our results when human capital is multi-dimensional. The process $\delta_{i,t}^{s}$ is the rental rate or (log) "price" of human capital in sector s at year t. The term $\exp(-\tau_{i,t}^{s})$ is a function of the decision variable, $\tau_{i,t}^{s}$, and can be interpreted as the fraction of time spent working, whereas $1-\exp(-\tau_{i,t}^{s})$ is the fraction of time devoted to investing in human capital when in sector s. The latter function is increasing in $\tau_{i,t}^{s}$, equal to zero when $\tau_{i,t}^{s} = 0$ and equal to one when $\tau_{i,t}^{s} = +\infty$. We call, $\tau_{i,t}^{s} \geq 0$, the individual specific investment in human capital in sector s at year t.

At date $t_{i,0}$, the individual is endowed with an initial stock of human capital $H_{i,t_{i,0}}$ that captures the endowment from education as well as experience in the alternative sector before entry in the private one. The technology of production of human capital in sector s is described by

$$H_{i,t+1} = H_{i,t} \exp\left(\rho_i^{s} \tau_{i,t}^{s} - \lambda_{i,t}^{s}\right), \tag{2}$$

in which ρ_i^{s} is the individual- and sector-specific rate of return of human capital investments and $\lambda_{i,t}^{s}$ is the depreciation of human capital. Depreciation $\lambda_{i,t}^{s}$ embeds individual specific and aggregate shocks that depreciate previous vintages of human capital. Depreciation shocks are sector-specific if human capital depreciation is different in the two sectors. The individual prices of human capital, $\delta_{i,t}^{s}$, and depreciations, $\lambda_{i,t}^{s}$, are treated as stochastic processes whose properties

[2]In the empirical model, participation in the private sector means a full-time job in that sector while any other status, e.g. part-time, self-employment, public sector employment, and non employment, is assigned to the alternative sector.

are detailed below.

We assume that investing in human capital is the only way of smoothing consumption over time, and we investigate below the generalizability of our results if this assumption does not hold. Absent consumption smoothing, year-t utility in sector s is a function of income, effort and participation:

$$\ln w_{i,t}^s - c_i \frac{(\tau_{i,t}^s)^2}{2} + \psi_{i,t}\mathbf{1}\{s = e\}, \tag{3}$$

in which the variable $\psi_{i,t}$ is the utility difference between sectors e and n. Furthermore, the cost of investment is quadratic and indexed by an individual specific parameter, c_i, that is assumed to be independent of sector s as it is a parameter of the utility function. Moreover, we omit the linear component of the cost in terms of $\tau_{i,t}^s$ because it cannot be identified.[3] Increasing marginal costs fits well with the interpretation of $\tau_{i,t}^s$ as an exerted effort which decreases current earnings and provides future returns. Convex costs make unique the solution of the dynamic programming decision problem.

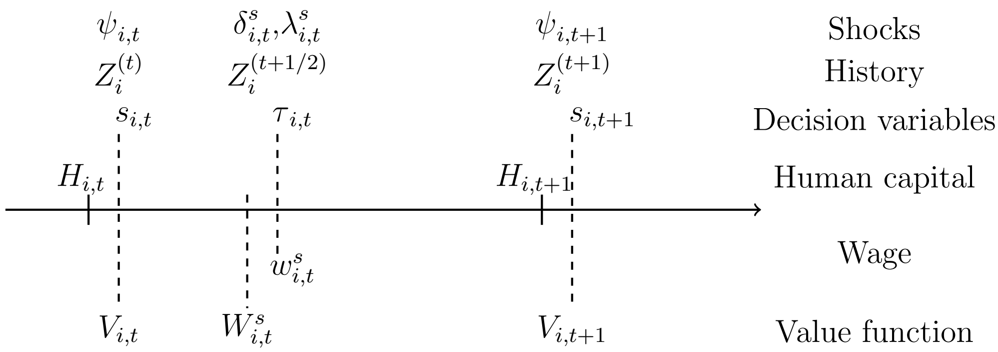
4.1.2 Timing and value functions

At each period, the individual maximizes his inter-temporal utility by computing the optimal human capital investments in the two sectors, and choosing the sector yielding the highest utility. The timing of revelation of shocks, state variables and decisions about sectors and human capital investments is plotted in Figure 2. A key assumption is that the revelation of sector preference shocks, $\psi_{i,t}$, and the choice of sector, $s_{i,t}$, are made before shocks on prices and depreciations of human capital are revealed, and before decisions about human capital investments are made as in Blevins (2014). An economic justification for this timing is that participation decisions are yearly commitments made before substantial uncertainty about price shocks ($\delta_{i,t}^s$ and $\lambda_{i,t}^s$) is lifted, while human capital investments can be made more adaptively all along the year. This is a specific version of the Roy model which is known, under conditions developed below, to lead to the absence of selection effects on wages. In the current paper, this absence of selection results from the conditioning on factors and factor loadings (see Section 5.2 below).

The first row of Figure 2 reports the timing of the revelation of shocks on sector preferences,

[3]Adding hypothetically a linear term $b_i\tau_{i,t}^s$ to the cost and combining equations (1) and (3) would yield year-t utility: $\delta_{i,t}^s + \ln H_{i,t} - (1 + b_i)\tau_{i,t}^s - c_i(\tau_{i,t}^s)^2/2 + \psi_{i,t}\mathbf{1}\{s_{i,t} = e\}$. As the unit in which $\tau_{i,t}^s$ is expressed is not identified, we can use the renormalization, $\tilde{\tau}_{i,t}^s = (1 + b_i)\tau_{i,t}^s$, and $\tilde{c}_i = c_i/(1 + b_i)^2$ so that the linear part of the cost is absorbed into the term $\tilde{\tau}_{i,t}^s$.

Figure 2: Timing of the model



$\psi_{i,t}$, and on price and depreciation of human capital, $\delta^s_{i,t}$ and $\lambda^s_{i,t}$. The second row reports the history – denoted $Z_i^{(.)}$ – of the year processes, $\delta^s_{i,t}$, $\lambda^s_{i,t}$ and $\psi_{i,t}$ up to the years described by the first row. In particular $Z_i^{(t)}$ contains the history of $\psi_{i,t}$ up to year t and the history of $\delta^s_{i,t}$, $\lambda^s_{i,t}$ up to year $t-1$. History $Z_i^{(t+1/2)}$ completes $Z_i^{(t)}$ with year-t information on $\delta^s_{i,t}$ and $\lambda^s_{i,t}$. The third row reports the timing of decisions: the choice of sector, $s_{i,t}$, is made after sector preference shocks are revealed, and human capital investments, $\tau_{i,t}$, are made after the revelation of shocks on prices and depreciation. The state variable $H_{i,t}$ is inherited from the past according to equation (2) at the very beginning of year t. Below the timeline, the wage $w_{i,t}$ is a function of shocks on prices and depreciation.

Value functions at each stage of this timeline can now be constructed. We denote $V_{i,t+1}$ the value function at the beginning of year $t+1$, whose arguments are the state variables, $H_{i,t+1}$ and $Z_i^{(t+1)}$. At the previous interim stage $t+1/2$, these state variables are $H_{i,t}$, $Z_i^{(t+1/2)}$. Because of equations (1) and (3), human capital investments are derived for each sector $s \in \{n, e\}$ from the following decision program:

$$W^s_{i,t}(H_{i,t}, Z_i^{(t+1/2)}) = \max_{\tau} \Bigg\{ \delta^s_{i,t} + \ln H_{i,t} - \left(\tau + c_i \frac{(\tau)^2}{2} \right) \quad (4)$$

$$+\beta \mathbb{E}_{t+1/2} \left[V_{i,t+1}(H_{i,t+1}, Z_i^{(t+1)}) \right] \Bigg\} \quad (5)$$

subject to the human capital accumulation equation (2) in sector s.

In this expression, $\mathbb{E}_{t+1/2}(.) = \mathbb{E}(. \mid H_{i,t}, Z_i^{(t+1/2)})$ and the discount rate β is restricted to be homogeneous among individuals, as is commonly assumed.[4] This means in particular that the delay between t and $t+1/2$ is smaller than the delay between $t+1/2$ and $t+1$ despite our abusive but clear notation, 1/2.

At the beginning of year t, we model sector choice as resulting from:

$$s_{i,t} = e \iff \mathbb{E}_t W^e_{i,t}(H_{i,t}, Z_i^{(t+1/2)}) + \psi_{i,t} > \mathbb{E}_t W^n_{i,t}(H_{i,t}, Z_i^{(t+1/2)}), \tag{6}$$

where $\mathbb{E}_t(.) = \mathbb{E}(. \mid H_{i,t}, Z_i^{(t)})$, and we complete the definition of the recursive equation in $V_{i,t}$ as:

$$V_{i,t}(H_{i,t}, Z_i^{(t)}) = \max(\mathbb{E}_t W^e_{i,t}(H_{i,t}, Z_i^{(t+1/2)}) + \psi_{i,t}, \mathbb{E}_t W^n_{i,t}(H_{i,t}, Z_i^{(t+1/2)})). \tag{7}$$

As sector choice, denoted by $s_{i,t}$, affects the accumulation of human capital, the optimal level of investment is denoted, $\tau_{i,t}^{s_{i,t}}$.

4.1.3 Terminal conditions and information

Individual i is assumed to invest in human capital at least until calendar year, T_i.[5] The terminal condition of his decision program is written for year $T_i + 1$. Specifically, the value function or discounted value of the utility stream from $T_i + 1$ onwards is assumed to be given by:

$$V_{i,T_i+1}(H_{i,T_i+1}, Z_i^{(T_i+1)}) = a_{i,T_i+1}(Z_i^{(T_i+1)}) + \kappa_i \ln H_{i,T_i+1}, \tag{8}$$

in which the level a_{i,T_i+1} generically depends on $Z_i^{(T_i+1)}$, and in which parameter κ_i is the individual specific marginal valuation of log human capital in the terminal year. This latter parameter commands the horizon effects for wages as we show below that the condition $\kappa_i < \frac{1}{1-\beta}$ makes sure that wage profiles are concave. Parameter κ_i is not indexed by $T_i + 1$ for notational simplicity, and is assumed to be independent of $Z_i^{(T_i+1)}$. To complete the description of the economic model, we further assume that agents have full information about their structural parameters, $c_i, \rho_i^s, H_{i,0}$ and κ_i. These parameters are included in the information set at the date of entry in the private

[4] In the theoretical model, β could be made individual specific at no cost. This would lead however to a more involved non linear factor model in the empirics which we leave for future work.

[5] In our empirical application, year T_i is the same for all individuals, (e.g., $T_i = T$), since it is the last observed year in our panel data. Note that the choice of a date for T_i is innocuous as long as individuals invest in human capital until that date. This is because, as shown below, provided that the terminal condition (8) is verified, it is also verified at all previous dates according to Proposition 1.

sector, $Z_i^{(t_{i,0})}$.

We also assume that the distribution of shocks $(\psi_{i,t}, \delta_{i,t}^s, \lambda_{i,t}^s)$ conditional on the appropriate information set does not depend on human capital so that workers do not have superior information:

Assumption:

$$\psi_{i,t} \amalg H_{i,t} \mid Z_i^{(t-1/2)}; \quad (\delta_{i,t}^s, \lambda_{i,t}^s) \amalg H_{i,t} \mid Z_i^{(t)}. \tag{9}$$

in which $\amalg$ denotes full independence.

This implies in particular that the expectation operators verify:

$$\mathbb{E}_{t+1/2}(.) = \mathbb{E}(. \mid Z_i^{(t+1/2)}) \text{ and } \mathbb{E}_t(.) = \mathbb{E}(. \mid Z_i^{(t)}) \tag{10}$$

Further properties of the stochastic processes will be set in the econometric Section 5.

4.2 Analysis

We now construct the steps that fully characterize the decisions of sectoral choice and human capital investments, through a sequence of Propositions whose proofs are relegated to Supplementary Appendix S.2.

4.2.1 Value functions and life-cycle profile of investments

We can analytically solve the dynamic model backwards because of linear assumptions, and we now show that the value functions are log-linear in human capital stocks at any period.

Proposition 1 *The sequence of value functions writes:*

$$W_{i,t}^s(H_{i,t}, Z_i^{(t+1/2)}) = a_{i,t}^s(Z_i^{(t+1/2)}) + \kappa_{i,t} \log H_{i,t} \text{ for } s = e, n \tag{11}$$

and:

$$V_{i,t}(H_{i,t}, Z_i^{(t)}) = a_{i,t}(Z_i^{(t)}) + \kappa_{i,t} \log H_{i,t} \tag{12}$$

in which for $t \leq T_i$,

$$\kappa_{i,t} = \frac{1}{1-\beta} + \beta^{T_i - t}(\kappa_i - \frac{1}{1-\beta}). \tag{13}$$

and functions $a^s_{i,t}(Z_i^{(t+1/2)})$ and $a_{i,t}(Z_i^{(t)})$ are defined in Proposition 3 and its proof.

Note that, as a consequence of condition $\kappa_i < 1/(1-\beta)$, $\kappa_{i,t}$ is decreasing in t. From this Proposition, we derive a closed form for human capital investments that depends on individual specific parameters.

Proposition 2 *The sequence of potential investments between $t = t_{i,0}$ and $t = T_i$ in each sector s is:*

$$\tau^s_{i,t} = \max\{0, \frac{1}{c_i}(\rho^s_i \beta \kappa_{i,t+1} - 1)\} \tag{14}$$

Note that the requirement that human capital investments remain positive until date T_i imposes that $\rho^s_i \beta \kappa_{i,t+1} > 1$ for $t \in \{t_{i,0}, ., T_i\}$. As human capital investments are decreasing over time when $\kappa_i < 1/(1-\beta)$, this condition becomes $\rho^s_i \beta \kappa_i > 1$ and is assumed to hold true below.
The previous Proposition also determines the dynamic equation for the additive terms in the value functions:

Proposition 3 *The sector specific additive terms in Proposition 1 are:*

$$a^s_{i,t}(Z_i^{(t+1/2)}) = \delta^s_{i,t} - \beta \kappa_{i,t+1} \lambda^s_{i,t} + c_i \frac{(\tau^s_{i,t})^2}{2} + \beta \mathbb{E}_{t+1/2}\left[a_{i,t+1}(Z_i^{(t+1)})\right].$$

in which $\tau^s_{i,t}$ is the optimal value of human capital investment in sector s as defined in equation (14).

The determination of the value functions in each sector finally leads to the determination of sectoral choice.

Proposition 4 *The sectoral choice is determined by:*

$$\begin{aligned} s_{i,t} &= e \text{ iff} \\ \psi_{i,t} + \mathbb{E}_t\left(\delta^e_{i,t} - \beta\kappa_{i,t+1}\lambda^e_{i,t} + c_i \frac{(\tau^e_{i,t})^2}{2}\right) &\geq \mathbb{E}_t\left(\delta^n_{i,t} - \beta\kappa_{i,t+1}\lambda^n_{i,t} + c_i \frac{(\tau^n_{i,t})^2}{2}\right). \end{aligned} \tag{15}$$

This is the structural equation that determines static selection. In particular, we will show below in Section 5 that equation (15) has an interactive effect structure if shocks $\delta_{i,t}$, $\lambda_{i,t}$ and $\psi_{i,t}$ have themselves a factor structure. We now turn to our main object of interest, the profile of log wages in the private sector.

4.3 The wage equation

By definition, a worker enters into sector e at year $t_{i,0}$, e.g. $s_{i,t_{i,0}} = e$. We adopt the following construction for the subsequent timing of interruptions of participation in the private sector. Denote $t_{i,1}$ the first year in which individual i stops being in the private sector and is in the alternative one. If it never happens, $t_{i,1}$ is set to $+\infty$. By construction, we have $t_{i,1} > t_{i,0}$. Similarly, denote $t_{i,2}$ the first year after $t_{i,1}$ in which individuals i stops being in the alternative sector, and works again in the private one. If it never happens, $t_{i,2}$ is set to $+\infty$ and by construction $t_{i,2} > t_{i,1}$. This construction is repeated for any transition between sectors after $t_{i,2}$ over the whole period. In sum, sub-sequence $(t_{i,0}, t_{i,2}, .)$ stands for years of (re-)entry into sector e (it gathers even values of the index), whereas sub-sequence $(t_{i,1}, t_{i,3}, .)$ denotes transition years into sector n (it gathers odd values).

Given the combined sequence $(t_{i,0}, t_{i,1}, ., t_{i,K_i+1} = +\infty)$ in which K_i is the number of transitions between sectors over the period, the mapping between year t and sectoral choice is described for appropriate k by:

$$\begin{aligned} s_{i,t} &= e \text{ for } t_{i,2k} \leq t \leq t_{i,2k+1} - 1, \\ &= n \text{ for } t_{i,2k+1} \leq t \leq t_{i,2k+2} - 1. \end{aligned} \tag{16}$$

Given this structure of spells in the private and alternative sectors, we can now derive the equation for log wages in the private sector. It is shown in the following Proposition to be a random coefficient model where explanatory variables are functions of potential experience, $t - t_{i,0}$, and functions of interruptions in participation in the private sector.

Proposition 5 *Consider a worker in sector e at date $t \in \{t_{i,0}, ., T_i\}$ and assume that $\tau_{i,l}^{s_{i,l}} > 0$ for any $t_{i,0} \leq l < T_i + 1$. Log wages in the private sector are given by:*

$$\ln w_{i,t} = \eta_{i0} + \eta_{i1}(t - t_{i,0}) + \eta_{i2}\beta^{-(t-t_{i,0})} + \eta_{i3}x_{i,t}^{(3)} + \eta_{i4}x_{i,t}^{(4)} + \underbrace{\delta_{i,t}^{e} - \sum_{l=t_{i,0}}^{t-1} \lambda_{i,l}^{s_{i,l}}}_{v_{i,t}}, \tag{17}$$

in which $\eta_i = (\eta_{i0}, \eta_{i1}, \eta_{i2}, \eta_{i3}, \eta_{i4})$ are functions of parameters $(\rho_i^s, c_i, \beta, \kappa_i)$ as well as of the initial value of human capital stock $\ln H_{i,t_{i,0}}$. The composite shocks, i.e. log-prices of human capital net of depreciations, are $v_{i,t} = \delta_{i,t}^{e} - \sum_{l=t_{i,0}}^{t-1} \lambda_{i,l}^{s_{i,l}}$, and variables $x_{i,t}$ are defined as $(1, t -$

$t_{i,0}, \beta^{-(t-t_{i,0})}, x^{(3)}_{i,t}, x^{(4)}_{i,t})$ *in which variables* $x^{(3)}_{i,t}$ *and* $x^{(4)}_{i,t}$ *are:*

$$x^{(3)}_{i,t} = \sum_{k=0}^{K_{i,t}-1} (t_{i,2k+2} - t_{i,2k+1})\,; \qquad x^{(4)}_{i,t} = \sum_{k=0}^{K_{i,t}-1} \left(\beta^{-(t_{i,2k+2}-t_{i,0})} - \beta^{-(t_{i,2k+1}-t_{i,0})}\right). \tag{18}$$

Reduced form (17) is our wage equation of interest in the empirical application.[6] This equation is a random coefficient model where the first three terms involve observed factors $(1, t, \beta^{-t})$, and other terms involve two variables $(x^{(3)}_{i,t}, x^{(4)}_{i,t})$ derived from the interruptions in participation in the private sector. These interruption variables generate dynamic selection effects in wages. We will deal with selection effects by imposing restrictions on the correlations between the wage shocks $v_{i,t}$ and participation shocks $\psi_{i,t}$ that are described in Section 5.2.

4.4 Generalizing the reduced form

We now investigate whether this reduced form is robust to relaxing two important assumptions of our set-up: the absence of consumption smoothing and the single dimensionality of human capital.

4.4.1 Exogenous savings

Magnac *et al.* (2018) derive predictions when allowing for endogenous consumption smoothing although their approach requires consumption information that is not available in our data. In the absence of consumption data, we can also relax the absence of consumption smoothing in the model by considering an exogenous savings rate, denoted $\sigma_{i,t}$. This savings rate is not a decision variable although it may differ from zero. Consumption is then equal to $w^s_{i,t}(1-\sigma_{i,t})$ and the logarithmic utility given by equation (3) becomes:

$$\delta^s_{i,t} + \ln H_{i,t} + \ln(1-\sigma_{i,t}) - (\tau^s_{i,t} + c_i \frac{(\tau^s_{i,t})^2}{2}) + \psi_{i,t}\mathbf{1}\{s_{i,t} = e\}, \tag{19}$$

in which we have used equation (1). Admittedly, savings feed the accumulation of financial assets at each period (not detailed here) and the value of these assets adds up to the value function. Assets are kept for consumption during the retirement period and for bequests. Human capital investments are however not affected because of additive separability. In other words, the exogenous term $\ln(1-\sigma_{i,t})$ can be confounded with the stochastic process $\delta^s_{i,t}$ and can even be

[6] We dropped the sectoral exponent e on wages since it is assumed that only private sector wages are observed.

sector-dependent. The only endogenous process is the human capital one.

In sum, our work adds to the literature by considering an alternative polar case compared to the standard consumption and income models in which the income process is exogenous while savings are endogenous (e.g. Blundell *et al.*, 2008). In this paper, we have to shut down one of the two channels of intertemporal smoothing in order to get tractable predictions as we do not observe both consumption and income. Our assumptions seem reasonable in light of Carroll (1997) who shows that consumption closely follows income, which would mean that variations of saving rates have a second order importance. The covariation of consumption with income is also in line with Browning and Crossley (2001) which discusses the different timings of consumption smoothing, and Kaplan *et al.* (2014) about the existence of hand-to-mouth wealthy consumers because of liquidity issues. Finally, it is consistent with Kolsrun *et al.* (2018) who show that consumption decreases when workers experience an unemployment spell.

4.4.2 Multidimensional human capital

We can generalize the model by allowing human capital to depend on an observable variable such as the skill group, as in Heckman *et al.* (1998). We will use this set up in our empirical application in which we will allow for skill-specific human capital prices and will apply the so called flat spot approach.

Most interestingly, we can also generalize equation (1) to the case in which human capital is multidimensional. Sector-specific wages are then written as:

$$\log(w_{i,t}^s) = \delta_{i,t}^s + \sum_{r=1}^{R} \delta_{i,r} \ln H_{i,r,t} - \sum_{r=1}^{R} \tau_{i,r,t}^s, \tag{20}$$

in which $H_{i,t} = (H_{i,r,t})_{k=1,.,R}$, with $R \geq 1$, are multiple human capital components. Denote investments in the different components as $\tau_{i,t} = (\tau_{i,r,t})_{r=1,.,R}$. Utility is logarithmic and verifies the following generalized version of equation (3):

$$\log(w_{i,t}^s) - \frac{(\tau_{i,t}^s)' C_i \tau_{i,t}^s}{2} + \psi_{i,t} \mathbf{1}\{s_{i,t} = e\}, \tag{21}$$

in which C_i is a definite positive matrix of dimension R that interrelates costs of investments across human capital types. Finally, the accumulation of human capital in equation (2) is rewritten as:

$$H_{i,r,t+1} = H_{i,r,t} \exp(\rho^s_{i,r} \tau^s_{i,r,t} - \lambda^s_{i,r,t}), \tag{22}$$

in which $\lambda^s_{i,r,t}$ are depreciation terms.

In this new environment, we obtain the same reduced form (17) as previously (see Supplementary Appendix S.2.6), although the interpretation of reduced-form coefficients changes. The important assumption in the specification above is that coefficients $\{\delta_{i,r}\}_{r=1,.,R}$ affecting the mix of human capital types do not depend on the sector chosen by the individual while the rates of return $\{\rho^s_{i,r}\}_{r=1,.,R}$ and the depreciations $\{\lambda^s_{i,r,t}\}_{r=1,.,R}$ can freely depend on the sector (see Supplementary Appendix S.2.6). Identifying separately individual-specific parameters related to multidimensional human capital would require additional external information as shown by Roys and Taber (2019) and Lise and Postel-Vinay (2020).

5 Econometric model

In our empirical analysis, we use panel data on male wages in France for different cohorts indexed by their years of entry into the private sector, $t_{i,0}$, that varies between 1985 and 1992. The end of the observation period is however given by a common value $T_i = T$ marking the end of the panel for all cohorts (the year 2011).[7] We rely on information on wages observed during spells of employment in the private sector to estimate equation (17) since no information is available when individuals are not employed in the private sector. This reduced-form equation allows for individual-specific parameters, η_i, which are functions of structural parameters (returns in both sectors, ρ^e_i, ρ^n_i, the cost of effort, c_i, the terminal value of human capital, κ_i, and the initial value of (log) human capital stock, $\log(H_{i,t_{i,0}})$), although we do not impose structural restrictions.[8] As seen above, this reduced form can also be derived from more general specifications at the cost of underidentification for some parameters.

In this Section, we first handle the distinction in the log wage equation between human capital

[7]Accordingly, our definition of T_i implies that human capital investments remain positive for all cohorts during the period we observe them.

[8]Note that the number of structural parameters and the number of reduced form parameters are both equal to 5 for every individual. The derivation of structural parameters from the reduced form, as well as sufficient conditions for this derivation, is available upon request.

stocks and prices by using a flat spot approach. Next, we introduce a restriction (called MARCOF) that is based on a factor decomposition of the exogenous stochastic processes. Under this restriction, static selection related to participation in the private sector and dynamic selection related to past interruptions are exogenous conditionally on factors and factor loadings. This means that selection and explanatory variables in equation (17) are conditionally exogenous. Finally, we discuss identification and present an Expectation Maximization algorithm to estimate parameters.

5.1 Flat spots

The flat spot approach was proposed by Heckman *et al.* (1998) and further investigated by Bowlus and Robinson (2012). This technique is meant to identify human capital prices and to recover human capital stocks, our main object of interest, by deflating log wages by human capital prices. We now show how the flat spot approach can be formalized in our theoretical framework. We briefly present the moment restrictions that this set-up generates while the full analytical development is relegated to Appendix S.1.2.

The two assumptions that underpin the technique of flat spots are that (1) human capital investments have stopped for individuals aged over 50 (this is consistent with Proposition 2 at these ages), and (2) human capital stocks are perfectly substitutable within skill groups. Human capital prices in each skill group can then be estimated yearly as the group-specific average wage for individuals aged 50-55. The remaining idiosyncratic individual variations in human capital prices have mean zero conditional on the information set.

More formally, the first flat spot restriction implies that $\tau_{i,t} = 0$ for individuals i with age $a_{i,t} \in [50, 55]$. Using this restriction in equation (1) yields the accounting identity that values are the product of volumes and prices:

$$\log w_{i,t} = \log H_{i,t} + \delta_{i,t}. \tag{23}$$

Equation (2) further yields:

$$\Delta \log H_{i,t} = -\lambda_{i,t-1}. \tag{24}$$

Using these two equations, we thus have for individuals aged 50-55:

$$\Delta \log w_{i,t} = -\lambda_{i,t-1} + \Delta\delta_{i,t} = \Delta v_{i,t},$$

using the definition of $v_{i,t}$ given in Equation (17). We compute averages in each skill group, $\varpi_{g,t} = E(\log w_{i,t} \mid i \in g, t, a_{i,t} \in [50, 55])$, so that:

$$\mathbb{E}(\Delta v_{i,t} \mid i \in g, t, a_{i,t} \in [50, 55]) = \Delta \varpi_{g,t}. \tag{25}$$

Variable $\varpi_{g,t}$ can then be interpreted as a human capital price net of depreciations. Supplementary Appendix S.1.2 reports the estimates of $\varpi_{g,t}$ using external information on male wages for workers aged 50-55 in different skill groups. Subtracting $\varpi_{g,t}$ from equation (17), we now define the deflated log wage as

$$\ln y_{i,t} = \ln w_{i,t} - \varpi_{g,t}, \tag{26}$$

in which $\ln y_{i,t}$ is the individual stock of human capital (in logs). According to the second restriction underlying flat spots, the idiosyncratic shock $v_{i,t}$ deflated by human capital prices $\varpi_{g,t}$ has mean zero at any period t and for any history of shocks $Z_i^{(t-1/2)}$:

$$\mathbb{E}(v_{i,t} - \varpi_{g,t} \mid Z_i^{(t-1/2)}, t) = 0. \tag{27}$$

Note that, at this stage, the expectation is conditional on $Z_i^{(t-1/2)}$ and not $Z_i^{(t)}$, because we do not condition on preference shock $\psi_{i,t}$. The selection process is thus not yet specified, and we impose restrictions on selection in the next section.

5.2 Exogeneity of Selection and Covariates

Our main specification of stochastic shocks and our main identifying restriction – labelled missing at random conditionally on factors (MARCOF) – are set out in this section. Using linear factor structures, we develop conditions under which selection is exogenous conditional on factors and factor loadings, and we show that covariates affecting the deflated wage equation (26) are conditionally exogenous. We proceed in two steps. By assuming a factor structure for preference, price and depreciation shocks, we first show that the deflated wage also has a factor structure. Next, by assuming conditional independence between preference and wage shocks, we show that the participation equation has a factor structure and that shocks in deflated wage and participation equations are conditionally independent. All proofs of lemmas and corollaries are relegated to Supplementary Appendix S.3.1.

5.2.1 Factor structure

We assume that shocks on preferences, prices and depreciations have a linear factor structure:

$$\begin{aligned}\psi_{i,t} &= \varphi_{\psi,t}\theta_{\psi,i} + \widetilde{\psi}_{i,t},\\ \delta^s_{i,t} &= \varphi_{\delta^s,t}\theta_{\delta^s,i} + \widetilde{\delta}^s_{i,t},\\ \lambda^s_{i,t} &= \varphi_{\lambda^s,t}\theta_{\lambda^s,i} + \widetilde{\lambda}^s_{i,t}.\end{aligned} \tag{28}$$

in which residual shocks $(\widetilde{\psi}_{i,t}, \widetilde{\delta}^s{}_{i,t}, \widetilde{\lambda}^s_{i,t})$ have conditional-on-factor mean zero and verify:

Assumption IND. Independence of residual shocks over time, with factors and individual effects:

$$(\widetilde{\psi}_{i,t}, \widetilde{\delta}^s{}_{i,t}, \widetilde{\lambda}^s_{i,t}) \amalg ((\psi_i^{(t-1)}, \delta_i^{s,(t-1)}, \lambda_{i,t}^{s,(t-1)}), (\varphi_{\psi,t}, \varphi_{\delta^s_t}, \varphi_{\lambda^s,t}), (\theta_{\psi,i}, \theta_{\delta^s,i}, \theta_{\lambda^s,i}, c_i, \rho_i^s, \kappa_i)), \tag{29}$$

where the history of shocks $(\psi_i^{(t-1)}, \delta_i^{s,(t-1)}, \lambda_{i,t}^{s,(t-1)})$ is defined such that, say, $\psi_i^{(t)} = \left(\psi_{i,t_{i,0}}, ..., \psi_{i,t}\right)$, and other terms are defined similarly.

We complete the specification of factor structure (28) with the additional assumption that:

Assumption INV. Invariance of interactive effects in depreciation across sectors:

$$\varphi_{\lambda^e,t} = \varphi_{\lambda^n,t} = \varphi_{\lambda,t} \text{ and } \theta_{\lambda^e,i} = \theta_{\lambda^n,i} = \theta_{\lambda,i} \tag{30}$$

This assumption allows the deflated wage equation to be rewritten as an interactive effect model as shown below. Allowing for differences between factors and factor loadings in depreciations across sectors would lead to a non linear factor model for wages that is significantly more involved to estimate.

Under assumptions (28) and (30), shocks in the deflated wage equation (26) derived from equation (17) verify:

$$\ln y_{i,t} - x_{i,t}\eta_i = \underbrace{\delta^e_{i,t} - \sum_{l=t_{i,0}}^{t-1} \lambda^{s_{i,l}}_{i,l}}_{v_{i,t}} - \varpi_{g,t} = \varphi_{\delta^e,t}\theta_{\delta^e,i} - \sum_{l=t_{i,0}}^{t-1} \varphi_{\lambda,l}\theta_{\lambda,i} + \varepsilon_{i,t}, \tag{31}$$

in which the residual shock is defined as:

$$\varepsilon_{i,t} = \widetilde{\delta}^{e}_{i,t} - \sum_{l=t_{i,0}}^{t-1} \tilde{\lambda}^{e}_{i,l} 1\{s_{i,l} = e\} - \sum_{l=t_{i,0}}^{t-1} \tilde{\lambda}^{n}_{i,l} 1\{s_{i,l} = n\} - \varpi_{g,t}. \tag{32}$$

Equation (31) gives the interactive structure of the deflated wage:

$$\ln y_{i,t} = x_{i,t}\eta_i + \varphi_{\delta^e,t}\theta_{\delta^e,i} - \sum_{l=t_{i,0}}^{1985} \varphi_{\lambda,l}\theta_{\lambda,i} + \sum_{l=t-1}^{1985} \varphi_{\lambda,l}\theta_{\lambda,i} + \varepsilon_{i,t}, \tag{33}$$

in which year 1985 is the entry date of the first cohort, and the second right-hand side term $\sum_{l=t_{i,0}}^{1985} \varphi_{\lambda,l}\theta_{\lambda,i}$ is an individual specific term independent of t. The right-hand side has a factor structure because it sums three interactive effects and a shock $\varepsilon_{i,t}$ orthogonal to factors and factor loadings. Furthermore, all right-hand side terms except the first one have conditional mean 0 because of the flat spot condition (27).

5.2.2 Conditional independence

We further restrict preference residual shocks $\tilde{\psi}_{i,t}$ and human capital price and depreciation residual shocks $\left(\tilde{\delta}^{s}_{i,t}, \tilde{\lambda}^{s}_{i,t}\right)$ to be independent conditionally on the history of factors and on factor loadings and permanent parameters, η_i. We write:

Assumption MARCOF. Missing At Random Conditionally On Factors:

$$\text{For all } t,\ (\tilde{\delta}^{s}_{i,t}, \tilde{\lambda}^{s}_{i,t}) \amalg \tilde{\psi}_{i,t} \mid (\varphi^{(t)}_{\psi}, \varphi^{(t)}_{\delta^s}, \varphi^{(t)}_{\lambda}, \theta_{\psi,i}, \theta_{\delta^s,i}, \theta_{\lambda,i}, \eta_i) \tag{34}$$

This assumption means that the correlations between shocks in the wage equation and the participation condition are governed by the factor and factor loadings structure and not by the idiosyncratic shocks $(\widetilde{\delta}^{s}_{i,t}, \widetilde{\lambda}^{s}_{i,t})$ and $\widetilde{\psi}_{i,t}$ which are conditionally independent. The factor structure of our model captures the individual-specific effects of business cycles, i.e. the fact that bad years and good years for the economy affect individuals differently. The main limitation of our MARCOF assumption is that it restricts the dynamic impact of individual-specific shocks in wages and participation, and in particular Ashenfelter's dips (Heckman and Smith, 1999) are assumed away. For instance, an individual health shock or job loss affecting wages in year $t-1$ cannot have lasting effects, and cannot affect wages and participation in year t.

It is remarkable that under the MARCOF assumption, the participation equation in the theoretical model also admits a linear factor structure:

Lemma 6 *The participation condition (15) verifies:*

$$s_{i,t} = e \Longleftrightarrow \widetilde{\psi}_{i,t} > \varphi_{s,t}\theta_{s,i}, \tag{35}$$

in which $\varphi_{s,t}$ are linear factors that are functions of $(\varphi_{\psi,t}, \varphi_{\delta^s,t}, \varphi_{\lambda,t})$ and year t, and $\theta_{s,i}$ is a function of factor loadings $\theta_{\psi,i}, \theta_{\delta^s,i}, \theta_{\lambda,i}$ and parameters η_i.

As this Lemma shows that factors (resp. factor loadings) affecting participation are linear functions of $\varphi_{\psi}^{(t)}, \varphi_{\delta^s}^{(t)}, \varphi_{\lambda}^{(t)}$ (resp. of factor loadings $\theta_{\psi,i}, \theta_{\delta^s,i}, \theta_{\lambda,i}$ and parameters η_i), the conditioning in the MARCOF assumption with respect to $(\varphi_{\psi}^{(t)}, \varphi_{\delta^s}^{(t)}, \varphi_{\lambda}^{(t)}, \theta_{\psi,i}, \theta_{\delta^s,i}, \theta_{\lambda,i}, \eta_i)$ is also true when conditioning additionally on participation factors and factor loadings, that is, on:

$$\varphi^{(t)*} = (\varphi_{\psi}^{(t)}, \varphi_{\delta^s}^{(t)}, \varphi_{\lambda}^{(t)}, \varphi_{s}^{(t)}); \quad \theta_i^* = (\theta_{\psi,i}, \theta_{\delta^s,i}, \theta_{\lambda,i}, \theta_{s,i}, \eta_i), \tag{36}$$

We are now in a position to state how selection effects in deflated wages can be conditioned out when we use factors and factor loadings and more generally all the information available just before the revelation of residual shocks:

Lemma 7 *Under assumptions (28) and (30):*

$$\varepsilon_{i,t} \perp\!\!\!\perp \widetilde{\psi}_{i,t} \mid (\varphi^{(t)*}, \theta_i^*) \tag{37}$$

$$\varepsilon_{i,t} \perp\!\!\!\perp (\varphi^{(t)*}, \theta_i^*). \tag{38}$$

5.2.3 Exogeneity

The MARCOF equation (34) provides a general conditioning argument to obtain independence between deflated wage residuals and residuals entering the selection equation conditional on factors and individual specific effects. As a consequence, the deflated wage equation (33) can be rewritten as:

$$\ln y_{i,t} = x_{i,t}\eta_i + \varphi_t\theta_i + \varepsilon_{i,t}. \tag{39}$$

The conditional mean of the residual shock writes:

$$E\left(\varepsilon_{i,t} \mid s_{i,t}=1, \varphi_t, \theta_i, Z_i^{(t-1/2)}\right) = E\left(\varepsilon_{i,t} \mid \widetilde{\psi}_{i,t} > \varphi_{s,t}\theta_{s,i}, \varphi_t, \theta_i, Z_i^{(t-1/2)}\right) \tag{40}$$
$$= E(\varepsilon_{i,t} \mid \varphi_t, \theta_i, Z_i^{(t-1/2)}) = 0.$$

The first equality derives from Lemma 6, the second one from Lemma 7, and the third one from the factor structure (28) and the flat spot condition (27). This proves that selection is exogeneous conditional on factors. Moreover, the right-hand side variables in the deflated wage equation (39) are conditionally exogenous as shown by the next Corollary.

Corollary 8 *Under assumptions (28), (30) and (34), selection is exogenous conditional on factors and factor loadings, and experience variables* $x_{i,t}^{(3)}$ *and* $x_{i,t}^{(4)}$ *are conditionally exogenous.*

As shown below, this allows a simple procedure that can be used when pooling all cohorts together and estimating parameters.

5.3 Identification: Movers and stayers

We now discuss the identification of parameters, η_i. Given Corollary 8, selection in the private sector is exogenous as well as $x_{i,t}^{(3)}$ and $x_{i,t}^{(4)}$. This implies that the individual-specific coefficients in the estimated equation (26) can be identified in the absence of multicolinearity for each individual. This last condition is not satisfied for some individuals because it requires enough mobility across sectors. This issue is akin to the one identified by Chernozhukov *et al.* (2013) in a treatment set-up. Indeed, consider an individual i who is employed during the whole observation period in sector e, or who moves only once out of sector e to sector n. In this case, $x_{i,t}^{(3)} = x_{i,t}^{(4)} = 0$ for all dates t during which this individual is working in sector e. In consequence, parameters η_{i3} and η_{i4} are not identified. Turn now to an individual making two transitions, one from e to n first, and then a return from n to e later. In this case, $x_{i,t}^{(3)} = (t_{i,2} - t_{i,1})\, 1_{\{t \geqslant t_{i,2}\}}$ and $x_{i,t}^{(4)} = \left(\beta^{-t_{i,2}} - \beta^{-t_{i,1}}\right) 1_{\{t \geqslant t_{i,2}\}}$, and the two variables $x_{i,t}^{(3)}$ and $x_{i,t}^{(4)}$ are proportional to $1_{\{t \geqslant t_{i,2}\}}$ where $1_{\{A\}}$ is the indicator function of the event A. Parameters η_{i3} and η_{i4} are not separately identified but the linear combination $\eta_{i3}\,(t_{i,2} - t_{i,1}) + \eta_{i4}\left(\beta^{-(t_{i,2}-t_{i,0})} - \beta^{-(t_{i,1}-t_{i,0})}\right)$ is. An additional final exit from employment would not have any additional identifying power. It is only if an individual makes four transitions (two from e to n and two from n to e such that $K_i \geq 4$) that parameters η_{i3} and η_{i4} are identified separately. Note that the underidentification of parameters η_{i3} and η_{i4} does not affect other parameters η_{i0}, η_{i1} and η_{i2} which are identified by the moment restrictions generated by the flat spot condition (27).

5.4 The Expectation-Maximization algorithm

In our interactive effect equation (39), our approach consists in minimizing the sum of squares of residuals for observations of workers in the private sector, and this is equivalent to maximizing the pseudo-likelihood function when disturbances are normal. As the model involves interactive effects and the panel is not balanced, we use an Expectation-Maximization (EM) algorithm as suggested by Bai (2009). In the expectation step, we replace missing observations with their linear predictions in years during which workers are not employed (sector n). In two sequential maximization steps, we maximize the pseudo-likelihood function for observations corresponding to all individuals and dates as in Bai (2009). More details on our EM algorithm can be found in Supplementary Appendix S.3.3.

6 Counterfactual Analysis: Structural functions

We now explain how we recover several interesting structural functions by using the set up of Blundell and Powell (2003). Specifically, we assess the relative importance of selection effects in the present and past of life-cycle histories as well as the impact of early versus late interruptions in those histories. We start with the general definition of structural objects and then explain how to compute them by manipulating the history of interruptions along the life-cycle while keeping constant the individual structural parameters $\eta_i = (\eta_{i0}, \eta_{i1}, \eta_{i2}, \eta_{i3}, \eta_{i4})$. To focus on life-cycle issues, we consider in the following, the counterfactual simulations of wages over the life cycle of all cohorts entering between 1985 and 1992, as a function of potential experience and not as a function of calendar time.

The interactive effect component of equation (39) can be written as a function of potential experience $d = t - t_{i,0}$:

$$\begin{aligned} x_{i,t}\eta_i &= \eta_{i0} + \eta_{i1}(t - t_{i,0}) + \eta_{i2}\beta^{-(t-t_{i,0})} + \eta_{i3}x_{i,t}^{(3)} + \eta_{i4}x_{i,t}^{(4)} \\ &= \eta_{i0} + \eta_{i1}d + \eta_{i2}\beta^{-d} + \eta_{i3}x_{i,t_{i,0}+d}^{(3)} + \eta_{i4}x_{i,t_{i,0}+d}^{(4)} = x_{i,t_{i,0}+d}\eta_i. \end{aligned} \tag{41}$$

Note that there are other terms in wage equation (39) such as idiosyncratic shocks, $\varepsilon_{i,t}$, and interactive effects, $\varphi_t \theta_i$. Those terms, which control for selection, have conditional mean zero because of flat spot equation (27). We leave them out since they are not affected by counterfactuals. For each skill group g, we then impute human capital prices that are constant over the whole period. These prices, denoted ϖ_g, are computed as the average value over time of human capital prices in skill group g, $\varpi_{g,t}$. The resulting prediction of log wages, $x_{i,t_{i,0}+d}\eta_i + \varpi_g$ is labelled the *adjusted log wage*. For notational simplicity, we neglect ϖ_g in the following development.

Denote $S_{i,t}$ any counterfactual individual choice of a sector at year t and $S_i^{(d)} = (S_{i,t_{i,0}}, ., S_{i,t_{i,0}+d})$ the counterfactual history as a function of potential experience, d, starting from $t_{i,0}$.[9] Furthermore, denote $x^{(3)}_{i,t_{i,0}+d}(S_i^{(d-1)})$ and $x^{(4)}_{i,t_{i,0}+d}(S_i^{(d-1)})$ the covariates defined by equation (18) as functions of counterfactual history $S_i^{(d-1)}$ e.g. for $k = 3,4$,

$$x^{(k)}_{i,t_{i,0}+d}(S_i^{(d-1)}) = x^{(k)}_{i,t}(S_{i,t_{i,0}}, ., S_{i,t_{i,0}+d-1}) \tag{42}$$

By extension, the observed values are $x^{(3)}_{i,t} = x^{(3)}_{i,t}(s_i^{(d-1)})$ and $x^{(4)}_{i,t} = x^{(4)}_{i,t}(s_i^{(d-1)})$ when $S_i^{(d-1)} = s_i^{(d-1)}$.

We define the counterfactual outcome as the term expressed in equation (41):

$$x_{i,t_{i,0}+d}(S_i^{(d-1)})\eta_i = \eta_{i0} + \eta_{i1} d + \eta_{i2}\beta^{-d} + \eta_{i3} x^{(3)}_{i,t_{i,0}+d}(S_i^{(d-1)}) + \eta_{i4} x^{(4)}_{i,t_{i,0}+d}(S_i^{(d-1)}), \tag{43}$$

which is the expectation of log wages of an individual given by a specific realization of η_i and $S_i^{(d-1)}$.

Given the distribution of η_i, and the counterfactual distribution of $S_i^{(d-1)}$, average structural functions of potential experience d are then defined as:

$$\mathbb{E}\left[Q\left(x_{i,t_{i,0}+d}(S_i^{(d-1)})\eta_i\right)\right]. \tag{44}$$

in which the expectation is taken over parameters η_i and variables $S_i^{(d-1)}$. Function $Q(.)$ can be of various types: the identity function to recover means, a quadratic function to recover variances, or indicator functions to recover deciles that can be used to compute inter-decile ranges. For instance,

[9] Recall that $s_{i,t}$ is the observed sector at year t and by extension $s_i^{(t)}$ is the realized history of sectoral choices between $t_{i,0}$ and t. By extension, $S_i^{(d)}$ is any realized or counterfactual history until reaching potential experience d for individual i. For simplicity, $S_i^{(d)}$ is called a counterfactual history.

if $S_i^{(d-1)}$ stands for the history of continuous participation in sector e, $S_i^{(d-1)} = (e, .., e)$, this expression defines the average counterfactual log wages at potential experience d, had participation been continuous for everyone.

Structural functions can also be defined conditionally on observed participation in the private sector, e.g:

$$\mathbb{E}\left[Q\left(x_{i,t_{i,0}+d}(S_i^{(d-1)})\eta_i\right) \mid s_{i,t_{i,0}+d} = e\right], \tag{45}$$

is a summary statistic of counterfactual log-wages for those who are working in sector e in year $t_{i,0} + d$ and setting the potential history to $S_i^{(d-1)}$.

We now review in more detail specific structural functions. We have to keep in mind that under-identification of (η_{i3}, η_{i4}) in the subpopulation with fewer than two interruptions propagates to any counterfactual in which the number of interruptions is set larger than the observed one. Hence, any estimation of counterfactuals affecting the timing of interruptions will be restricted to individuals with two and more interruptions.

6.1 Selection and interruption effects

We compute structural functions (44) contrasting the observed and counterfactual situations where selection effects on wages are neutralized. Recall that we distinguish static selection effects due to individuals being out of the private sector at a given date, and dynamic selection effects due to past spells out of the private sector. We specify the object of equation (44) in the following four cases:

- The benchmark case in which the potential history $S_i^{(d-1)}$ is equal to its observed value, $s_i^{(d-1)}$, for those who currently participate:

$$Q_d^{(0)} = \mathbb{E}\left(Q\left(x_{i,t_{i,0}+d}(s_i^{(d-1)})\eta_i\right) \mid s_{i,t_{i,0}+d} = e\right). \tag{46}$$

- The counterfactual in which interruption effects are neutralized, i.e. $S_i^{(d-1)} = (e, ., e)$ is labelled "No interruption" below. In that case, there is no career interruption and $x_t^{(3)}(S_i^{(d-1)}) = x_t^{(4)}(S_i^{(d-1)}) = 0$:

$$Q_d^{(1)} = \mathbb{E}\left(Q\left(x_{i,t_{i,0}+d}(e, ., e)\eta_i\right) \mid s_{i,t_{i,0}+d} = e\right). \tag{47}$$

- The counterfactual in which current selection effects are neutralized, labelled "No selection" below:

$$Q_d^{(2)} = \mathbb{E}\left(Q\left(x_{i,t_{i,0}+d}(s_i^{(d-1)})\eta_i\right)\right). \tag{48}$$

- The counterfactual in which both selection and interruption effects are neutralized, labelled "No selection, No interruption" below:

$$Q_d^{(3)} = \mathbb{E}\left(Q\left(x_{i,t_{i,0}+d}(e,.,e)\eta_i\right)\right). \tag{49}$$

We can contrast the life cycle profiles $\{Q_d^{(j)}\}_{j=1,.,3}$ with the benchmark profile $Q_d^{(0)}$ when potential experience d varies. The first contrast, $Q_d^{(1)} - Q_d^{(0)}$ identifies dynamic selection effects, while $Q_d^{(2)} - Q_d^{(0)}$ identifies static selection effects, and $Q_d^{(3)} - Q_d^{(0)}$ captures both static and dynamic selection effects.

6.2 The impact of interruptions: Random, early or late interruptions

The effect of the timing of interruptions on wages can also be estimated using this framework. This timing influences current wages since they are partly determined by all past spells out of employment. We restrict our attention to individuals experiencing at least two such spells followed by employment spells ($K_i \geq 4$) whose parameters η_{i3} and η_{i4} related to spells out of private sector are identified.

We then compute our statistics in three different counterfactual situations. In the first one, years out of the private sector are randomly assigned over time for every individual. We hold the total number of years of interruption constant and set the last year to which an interruption can be randomly assigned to the last year of observation. Contrasting this counterfactual to the benchmark allows the correlation of the timing of interruptions with individual specific parameters, η_i, to be assessed. Note that this correlation is entirely due to factors and factor loadings because of the MARCOF assumption.

The other counterfactual exercises consist in reassigning interruptions either at the beginning of the observed life cycle (imposing at least one year of presence in the private sector) or at the end. By comparing these two counterfactuals to the benchmark, we can measure the wage changes due to career interruptions at the beginning and at the end of the life-cycle, for those individuals who have intermittent careers, in the same spirit as Light and Ureta (1995).

7 Specification choice

In this section, we first present estimation results for different specifications of the model, and justify our preference for the specification given by equation (39) when considering two unobserved factors.

7.1 Model selection and comparisons

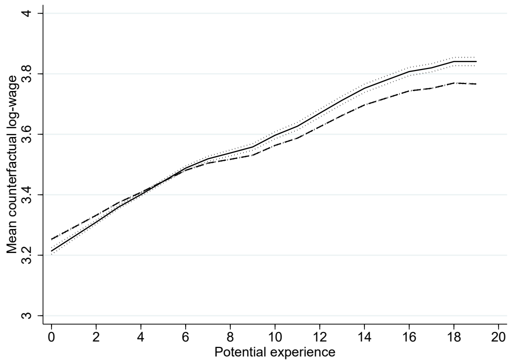
We estimate five models: a *basic* model that includes neither interruption variables nor factors while the others include interruption variables and an increasing number of factors $(0, 1, 2$ and $3)$.[10] Our preferred specification, called *main* below, includes interruption variables and two factors, and this preference rests on three arguments: (1) A significance test for estimated coefficients of interruption variables and factor loadings in no-, one- and two-factor specifications rejects that those are equal to zero; (2) Three out of six model selection criteria proposed by Bai and Ng (2002) point to the two-factor model as the best one among the one- to three-factor alternatives (see Supplementary Table S.2); (3) Estimates for the three-factor specification are quite unstable signaling possible identification issues and overfitting.

We report summaries of predicted wages over the life-cycle, and contrast results for basic and main specifications. Specifically, Figure 3 displays profiles of mean, median, variance and inter-decile range of potential wages that are defined as adjusted log wages obtained when there is no transition to the alternative sector ($x_{i,t}^{(3)} = x_{i,t}^{(4)} = 0$). Note that these potential wages only depend on the estimates of the first three individual specific coefficients, i.e. $\eta_{i0}, \eta_{i1}, \eta_{i2}$, for any model specification.

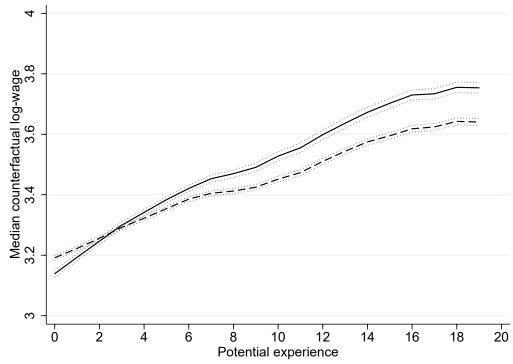
Figures 3(a) and 3(b) display mean and bias-corrected median profiles. There is a marked contrast between the basic and main specifications. Mean or median profiles are steeper when using the main specification. This indicates that interruptions and/or selection into the private sector have significant effects on wages, and that ignoring them biases downward returns to potential experience.

[10] Supplementary Appendix S.1.3 discusses results of the "Mincer" equation derived from (39) in which coefficients are homogenous and factors are absent.

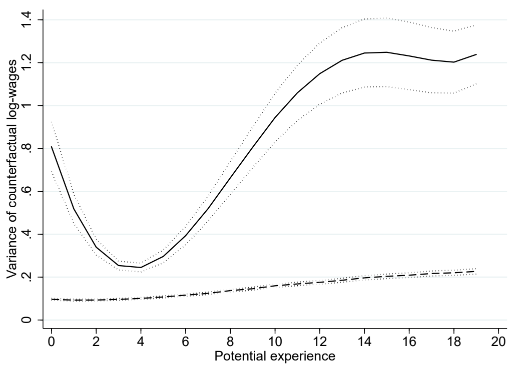
Figure 3: Mean, median, variance and inter-decile range of adjusted log-wages as a function of potential experience, main and basic specifications



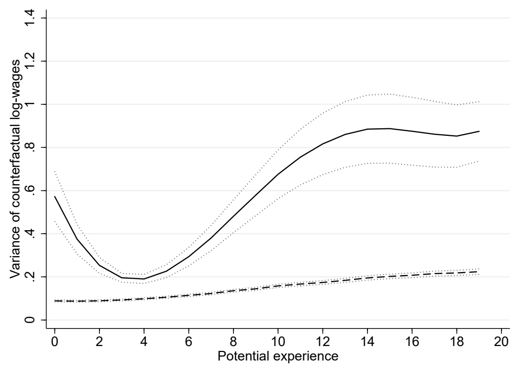
(a) Mean: Main (solid), Basic (dashed)



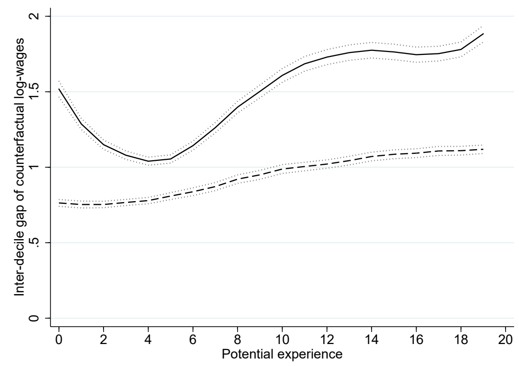
(b) Corrected median: Main (solid), Basic (dashed)



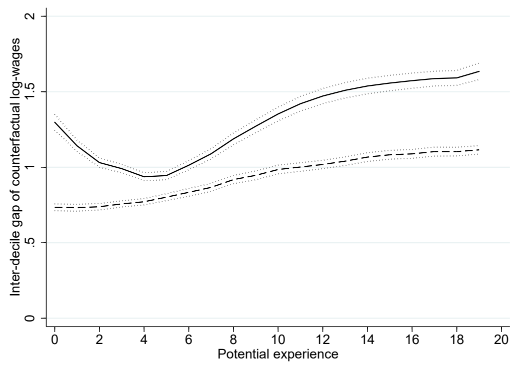
(c) Variance: Main (solid), Basic (dashed)



(d) Corrected variance: Main (solid), Basic (dashed)



(e) Inter-decile: Main (solid), Basic (dashed)



(f) Corrected inter-decile: Main (solid), Basic (dashed)

Note: Log-wages are computed as $\eta_{i0} + \eta_{i1} d + \eta_{i2} \beta^{-d}$ where d is potential experience, and are ajusted by (i) changing the timing of cohorts such that all cohorts enter the labour market in 1984; (ii) deflating wages with skill-specific prices of human capital; (iii) adding time-constant skill-specific prices of human capital. Wage statistics are computed on the whole set of individuals at each value of potential experience (whether they are employed or not). "Corrected" statistics are obtained after bias correction as described in Supplementary Appendix S.3.6. We consider the following estimated specifications: "Main": it includes variables $x^{(1)}_{i,t_{i,0}+d}$, $x^{(2)}_{i,t_{i,0}+d}$, $x^{(3)}_{i,t_{i,0}+d}$ and $x^{(4)}_{i,t_{i,0}+d}$ as well as the additive individual effect and two interactive factors; "Basic": it includes only variables $x^{(1)}_{i,t_{i,0}+d}$ and $x^{(2)}_{i,t_{i,0}+d}$, and the additive individual effect.

Figure 3(c) displays the profiles of the uncorrected variance of potential log wages and the bias-corrected estimates (see Supplementary Appendix S.3.6 for their computation) are displayed in Figure 3(d). The comparison between them shows the extent of the bias in variances. Furthermore, these graphs show that variance estimates are larger for the main specification than for the basic one. In particular, results for the main specification display a Mincer dip in line with Mincer (1974) since the profile of variances is U-shaped – at least before ten years of experience. The profile of high-return workers, who invest more in human capital at the beginning of their life-cycle, crosses the profile of low-return workers after a few years. The crossing point is estimated at about 4 years.

Monte Carlo experiments show, however, that biases in corrected variances might remain sizable (Gobillon *et al.*, 2022). This is why we turn to the profile of the inter-decile range of potential log wages (Figure 3(e)). Correcting the bias for the inter-decile range mildly affects these profiles by at most 15% (Figure 3(f)). The inter-decile range for the main specification is hovering between 90% and 160%, and here also, profiles are slightly higher than for the basic specification. In contrast with variances though, the Mincer dip is slightly dampened although the trough is still estimated at about 4 years of potential experience.

7.2 Estimated components of wages

For our preferred specification, we decompose adjusted log wages into the contributions of their different components: potential experience, interruptions and factors. A widespread approach to quantify the importance of those components is to rely on a variance decomposition. As already explained, we instead focus on the more robust inter-decile ranges and report rank correlations that are corrected for biases. Results on inter-decile ranges in Table 2 show that the contribution of the effect of potential experience is the largest and that the effect of interruptions is sizable while factors play a lesser role. Remarkably, the effects of potential experience and interruptions are highly negatively rank-correlated. This is mainly due to the strong negative correlation between linear coefficients, η_{i1} and η_{i3}. Their Spearman rank correlation is equal to -0.299.

Table 2: Corrected inter-decile ranges and rank correlations of the effects, main specification

		Rank correlation			
	Inter-decile	Adjusted log-wage	Potential experience effect	Interruption effect	Factor effect
Adjusted log-wage	0.819	1.000	0.517	0.011	0.056
Potential experience effect	1.250	0.517	1.000	-0.564	-0.306
Interruption effect	0.657	0.011	-0.564	1.000	-0.028
Factor effect	0.414	0.056	-0.306	-0.028	1.000

Note: Adjusted log-wages are computed from raw wages by (i) changing the timing of cohorts such that all cohorts enter the labour market in 1984; (ii) deflating wages with skill-specific prices of human capital; (iii) adding time-constant skill-specific prices of human capital. "Potential experience effect": sum of all effects related to potential experience d and the individual additive effect: $\eta_{i0}+\eta_{i1}d+\eta_{i2}\beta^{-d}$; "Interruption effect": sum of all effects related to being absent from the panel: $\eta_{i3}x^{(3)}_{i,t_{i,0}+d}+\eta_{i4}x^{(4)}_{i,t_{i,0}+d}$; " Factor effect": sum of all effects related to factors and factor loadings $\theta^{(1)}_i\varphi^{(1)}_t+\theta^{(2)}_i\varphi^{(2)}_t$. Statistics are computed on the whole sample of individuals (whether they are employed or not). They are corrected for bias as described in Supplementary Appendix S.3.6.

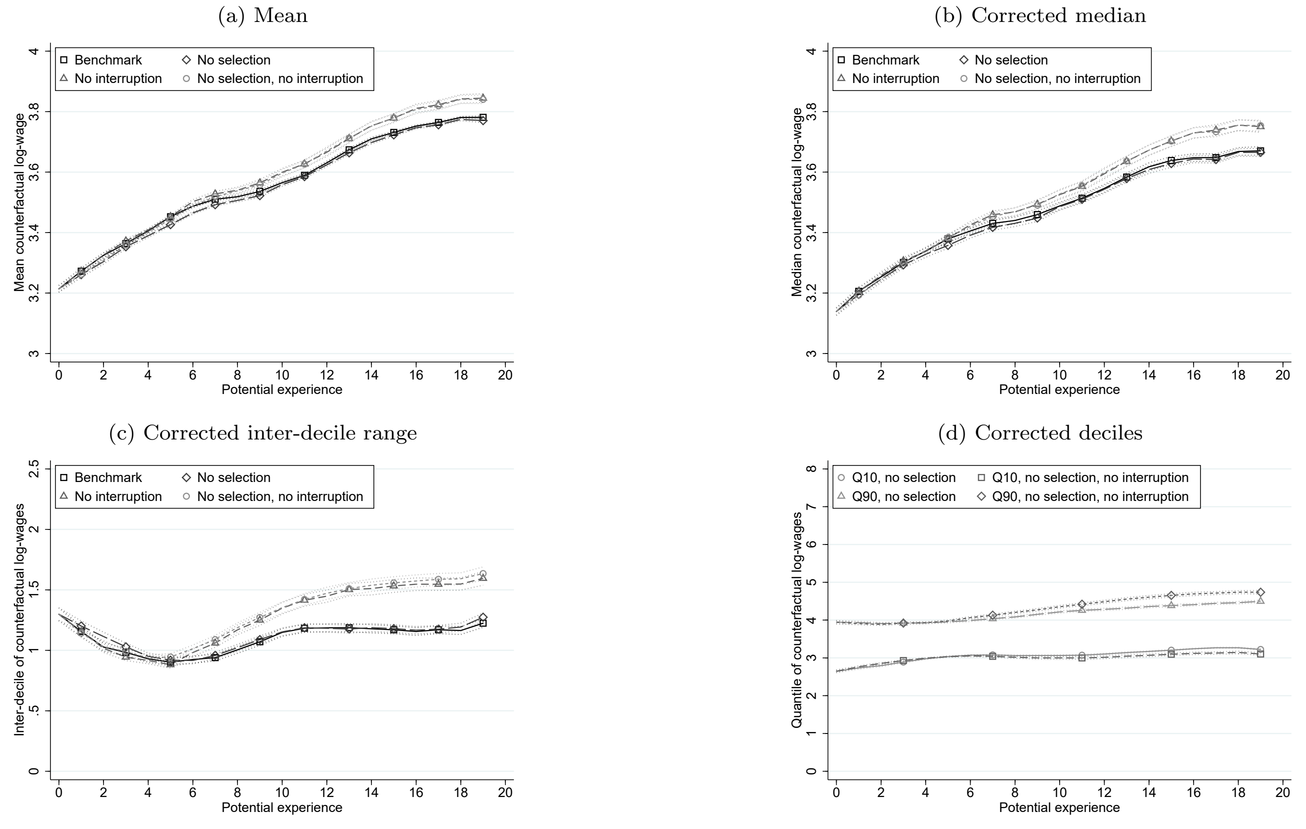
8 Counterfactual empirical analysis

We characterize counterfactual wage profiles in the four cases introduced in Section 6.1 to investigate the economic importance of static and dynamic selection in our working sample, and their impact on the mean and dispersion of wages. We then turn to the analysis of counterfactuals for two disjoint subsamples of individuals: those with no interruption and those with two interruptions or more, to assess how individual unobserved heterogeneity affects returns to experience. Next, we study the effects of the timing of interruptions by resorting to counterfactual wage profiles when interruptions are drawn randomly, or reallocated over the life-cycle, as introduced in Section 6.2. We also investigate the selection in our main sample of individuals working more than 15 years in the private sector. Finally, we conduct robustness checks when changing entry dates and the discount factor. For all our counterfactual exercises, we implement bias corrections.

8.1 Counterfactuals: Static and dynamic selection

We contrast different structural objects defined in Section 6 to evaluate the economic impact of interruptions and participation on the profile of log wages for private sector employees when log wages are predicted using potential experience and interruptions, but excluding factors as in equation (41). Figure 4 presents summary statistics of wage profiles in three counterfactual situations defined in Section 6.1: the absence of static selection (labelled "No selection"); the absence of dynamic selection due to interruptions ("No interruption") or both ("No selection, No interruption"). We compare them to the benchmark in which static and dynamic selections are present.

Figure 4: Mean, median, inter-decile range, and first and last deciles of counterfactual adjusted log-wage as a function of potential experience, counterfactual scenarii when neutralizing static and/or dynamic selections



Note: "No interruption" and "No selection, no interruption": Log-wages are computed as $\eta_{i0}+\eta_{i1}d+\eta_{i2}\beta^{-d}$ where d is the potential experience. "Benchmark" and "No selection": Log-wages are computed as $\eta_{i0} + \eta_{i1}d + \eta_{i2}\beta^{-d} + \eta_{i3}x^{(3)}_{i,t_{i,0}+d} + \eta_{i4}x^{(4)}_{i,t_{i,0}+d}$. In both cases, log-wages are adjusted by (i) changing the timing of cohorts such that all cohorts enter the labour market in 1984; (ii) deflating wages with skill-specific prices of human capital; (iii) adding time-constant skill-specific prices of human capital. "Benchmark" and "No interruption": Wage statistics are computed using the subsample of individuals employed at the value of potential experience that is considered."No selection" and "No selection, no interruption": Wage statistics are computed on the whole sample of individuals at each value of potential experience (whether they are employed or not). "Corrected" statistics are obtained after bias correction as described in Supplementary Appendix S.3.6.

First, static selection has no significant effect on medians, means, inter-decile ranges and first and last deciles as shown by Figure 4. By contrast, interruptions have a strong and significant dynamic selection effect. In other words, both potential and real experience matter (e.g. Light and Ureta, 1995; Das and Polachek, 2019) and their effects differ. Figures 4(a) and 4(b) show that potential experience increases mean log wages by around 80% (= [exp(0.6)−1] ∗ 100) in 20 years. This result squares well with other studies which cover many countries and use homogenous Mincer equations (e.g. Lagakos *et al.*, 2018). The average wage loss due to interruptions is about

7% after 20 years.

The impact of interruptions on the dispersion of wages is shown in Figure 4(c) through the lens of inter-decile ranges (see also Figure S.2 for variances). After 20 years, interruptions decrease dispersion by -0.36 (-30%).[11] This effect plays on both tails at the 90% quantile and the 10% quantile. Its magnitude is stronger at the 90% quantile (-0.24 after 20 years) than at the 10% quantile (+0.12) as shown by Figure 4(d). These results on the impact of interruptions on dispersion is related to the negative rank correlation between the effects of potential experience and interruptions that we mentioned when commenting Table 2.

We now consider other counterfactuals to better understand this impact of interruptions on wage dispersion which is a new stylized fact to our knowledge.

8.2 Counterfactuals: contrasting wage profiles between stayers and movers

We distinguish two disjoint subsamples, one composed of individuals with no interruption (e.g. "stayers") and the other consisting of individuals with two interruptions or more (e.g. "movers"). We show that movers have a much more dispersed distribution of wages than stayers.

We first focus on the "No selection, No interruption" case in which static and dynamic selection are neutralized. The left panel of Table 3 reports descriptive statistics on counterfactual adjusted log-wage distributions for four values of potential experience (1, 5, 10 and 20), and contrasts stayers in the top left with movers in the bottom left. It shows that there are sizable differences with mean, median and first decile of log wages after ten years being larger for stayers by 0.127, 0.089 and 0.413 points, respectively. Conversely, the last decile of wages is greater by 0.094 points for movers. Interestingly, this yields a wage dispersion measured by the inter-decile range that is greater by as much as 0.508 points for movers.

We next turn to the bottom right panel of Table 3 which provides summaries of adjusted log-wages for the subsample of movers in the "No selection" counterfactual.[12] Dynamic selection effects for this subsample can be inferred by comparing bottom-right and bottom-left panels. Interestingly, the sign and extent of interruption effects vary along the wage distribution. Interruptions increase wages after ten years at the first decile by 0.175 points but decrease them at

[11] Over 20 years, the average duration of interruptions is 2.3 years.

[12] Counterfactual adjusted-log wages remain the same for stayers since they have no interruptions.

Table 3: Descriptive statistics on distributions of counterfactual adjusted log-wages as a function of potential experience, subsamples of individuals with no interruption and two interruptions and more

	No selection, no interruption				No selection			
Potential experience	1	5	10	20	1	5	10	20
Subsample: Individuals with no interruption								
Mean wages	3.270	3.475	3.644	3.872				
Variance	0.318	0.116	0.187	0.349				
Inter-decile	1.126	0.890	1.033	1.295				
Q5	2.521	3.047	3.136	3.204				
Q10	2.725	3.111	3.206	3.322				
Q25	2.988	3.229	3.346	3.499	Same as left panel			
Median	3.228	3.396	3.546	3.767				
Q75	3.552	3.647	3.852	4.161				
Q90	3.851	4.001	4.239	4.616				
Q95	4.142	4.149	4.437	4.891				
N	1579	1579	1579	1579				
Subsample: Individuals with two interruptions and more								
Mean wages	3.167	3.356	3.517	3.820	3.167	3.332	3.449	3.692
Variance	0.700	0.236	0.851	1.267	0.700	0.269	0.257	0.660
Inter-decile	1.436	0.987	1.541	2.030	1.436	0.970	1.082	1.209
Q5	2.079	2.724	2.316	2.296	2.079	2.698	2.762	3.073
Q10	2.558	2.901	2.793	2.821	2.558	2.891	2.968	3.177
Q25	2.871	3.103	3.140	3.318	2.871	3.086	3.168	3.353
Median	3.098	3.306	3.457	3.743	3.098	3.283	3.378	3.592
Q75	3.407	3.580	3.874	4.297	3.407	3.523	3.668	3.937
Q90	3.994	3.888	4.333	4.851	3.994	3.861	4.051	4.386
Q95	4.396	4.217	4.768	5.541	4.396	4.173	4.294	4.695
N	3202	3202	3202	3202	3202	3202	3202	3202

Note: Log-wages when "No selection, no interruption" are computed as $\eta_{i0} + \eta_{i1}d + \eta_{i2}\beta^{-d}$, where d is the potential experience, and wages when "no selection" are computed as $\eta_{i0} + \eta_{i1}d + \eta_{i2}\beta^{-d} + \eta_{i3}x^{(3)}_{i,t_{i,0}+d} + \eta_{i4}x^{(4)}_{i,t_{i,0}+d}$. In both cases, log-wages are adjusted by (i) changing the timing of cohorts such that all cohorts enter the labour market in 1984; (ii) deflating wages with skill-specific prices of human capital; (iii) adding time-constant skill-specific prices of human capital. Wage statistics are computed for five values of potential experience: 1, 5, 10, 15 and 20 over the whole set of individuals in subsamples (whether they are employed or not). We report "Corrected" statistics obtained after bias correction as described in Supplementary Appendix S.3.6. N: number of observations in the considered subsample.

the last decile by as much as 0.282 points. As a consequence, wage dispersion measured by the inter-decile range ends up being much lower by 0.459 points than it would have been in the absence of interruptions. This confirms the effect of interruptions on dispersion found in the previous subsection to which stayers do not contribute.

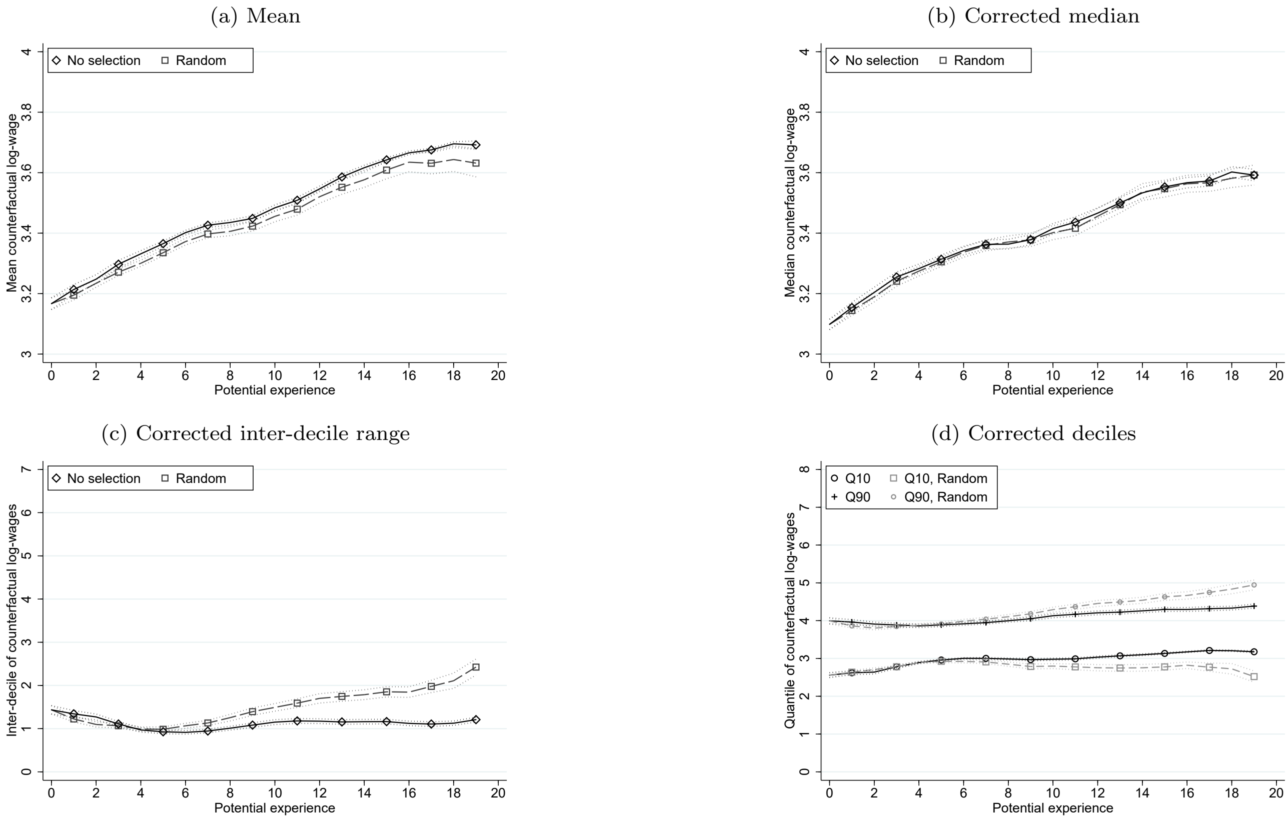
Our favourite story explaining the negative effect of interruptions on wage dispersion distinguishes workers at the top and those at the bottom of the wage distribution. First, it could be that some of the workers at the top returning to the private sector after an interruption (say as self-employed or abroad) are those who discovered that the alternative sector offered them lower returns on their human capital investments compared to the private sector. In consequence, their accumulation of human capital had been less intense than what it would have been, had they stayed in the private sector, and this shows up in their wages when they return to the private sector. In contrast, at the bottom of the distribution, near the minimum wage, the reverse may happen. Among workers exiting to the alternative sector, mostly those who get larger returns to human capital investments in that sector (including when unemployed) and invest more in human capital, re-enter the private sector. They then earn larger wages than they would have earned, had they stayed in the private sector.

8.3 Counterfactual timing of interruptions

We now assess the importance of the timing of interruptions exploring additional counterfactuals. We neutralize static selection effects and focus on effects related to dynamic selection. We restrict our attention to the subsample of movers with two interruptions or more since parameters related to interruptions, η_{i3} and η_{i4}, are identified only for those individuals.

In the first counterfactual, we randomly assign years of interruptions over time for each worker. This neutralises the effects of individual heterogeneity on the timing of interruption spells. We hold the number of years of interruption constant and set the last year to which an interruption can be randomly assigned, to the last year of observation before final attrition. We compare the resulting couterfactual profiles of adjusted log wages (the “Random” case) to those when years of interruptions are the observed ones, when participation selection is absent (the “No selection” case). Results reported in Figures 5(a) and 5(b) show that mean or median wage profiles are very close in the random and no selection cases. In consequence, mean or median returns to potential experience are not much affected by endogenous choices of interruptions. By contrast, inter-decile ranges start diverging after 5 years (Figure 5(c)), and wage dispersion increases more quickly in the random case than in the no selection one.

Figure 5: Mean, median, deciles and inter-decile range of counterfactual adjusted log-wages as a function of potential experience, counterfactual when years of interruption are random, sample of individuals with two interruptions or more



Note: Adjusted log-wages are computed from raw wages by (i) changing the timing of cohorts such that all cohorts enter the labour market in 1984; (ii) deflating wages with skill-specific prices of human capital; (iii) adding time-constant skill-specific prices of human capital. "Corrected" statistics are obtained after bias correction as described in Supplementary Appendix S.3.6. For this counterfactual exercise, we focus on dynamic selection effects while static selection is neutralized. This is why "No selection" is our benchmark. In the "Random" case, interruptions years are drawn randomly in the period up to and including the individual's final year of employment.

This result helps understand why wage dispersion is larger when interruptions are suppressed as shown in previous Subsection 8.1. When interruptions are set to zero, it neutralizes both the timing and the number of years of interruptions. What Figure 5(c) shows, is that making the timing of interruptions random explains part of the larger wage dispersion only. In consequence, the larger wage dispersion when interruptions are suppressed is not only due to the endogeneity of interruptions but also to the heterogeneity of parameters commanding the effect on wages of interruptions, $\eta_{i,3}$ and $\eta_{i,4}$. This is consistent with the interpretation developed at the end of the previous section. Furthermore, this dispersion comes from changes in both the first and last deciles (Figure 5(d)) as in Section 8.2.

8.4 Early and late interruptions

We finally estimate counterfactuals related to the structure of interruptions by reassigning interruptions either at the beginning or at the end of the observed life-cycle (between the initial year in the private sector and the last year before final attrition) as was studied by Light and Ureta (1995). Again, we contrast those counterfactuals with the counterfactual in which static selection is absent ("No selection"). Results are reported in Figure 6 in which "Interruptions first" (respectively "Interruptions last") refers to moving interruption spells at the beginning (resp. at the end) of the observed life cycle.

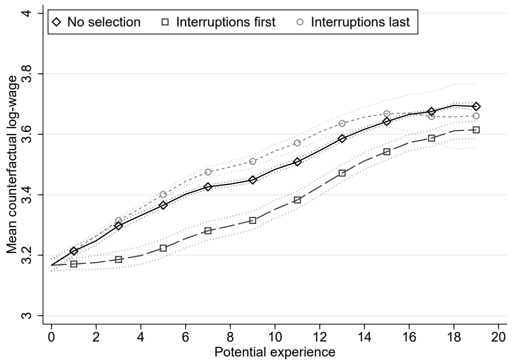
Reassigning interruptions at the beginning has an important negative effect on mean and median log wages over the whole period (Figures 6(a) and 6(b)). Mean log wages never recover what they have initially lost while median log wages do. By contrast, when reassigning interruptions at the end of observed life-cycle, mean log wages increase above what is observed for any number of years of experience. Effects are smaller and insignificant for median log wage profiles.

Interestingly, reassigning interruptions at the beginning of the observed life-cycle largely increases the inter-decile range when compared to the "No selection" counterfactual (Figure 6(c)) over the whole life-cycle. This increase is larger at the beginning of the life-cycle as expected, and it slowly decreases after 6 years, presumably because the time already spent in interruptions is getting closer for individuals in the two counterfactuals. This widening of the inter-decile range is due to both a larger last decile and a lower first decile (Figure 6(d)).

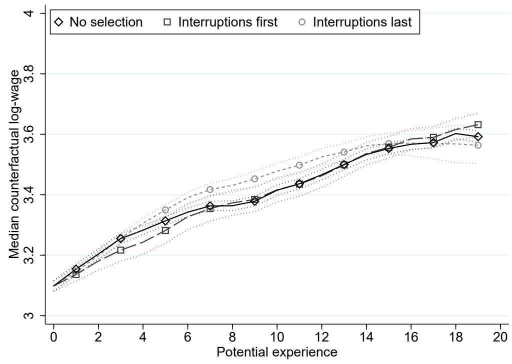
Since we focus on the subsample of movers, the interpretation of the results we propose is based on the heterogeneity of workers within this group. Since individuals who experience early interruptions in their careers are hardly affected, the results are mostly due to individuals who experience late interruptions. Among these, consider low wage individuals first and assume that those who experience late interruptions accumulate less human capital in the alternative sector than in the private sector. Shifting these detrimental interruptions to the beginning of the career results in lower wages for these individuals in the early years of potential experience than in the benchmark situation. In contrast, high wage individuals who experience late interruptions may choose the alternative sector because of the job opportunities associated with higher human

Figure 6: Mean, median, deciles and inter-decile range of counterfactual adjusted log-wages as a function of potential experience, counterfactual scenario when interruptions years occur in the first or last years in the private sector, sample of individuals with two interruptions or more

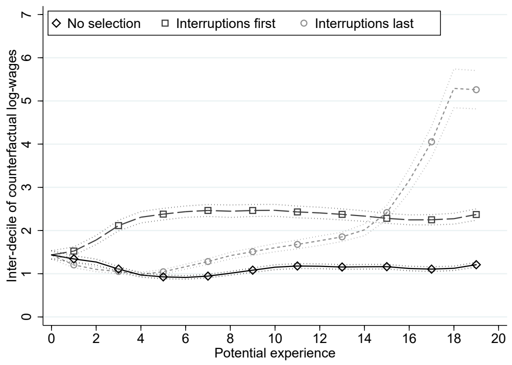
(a) Mean



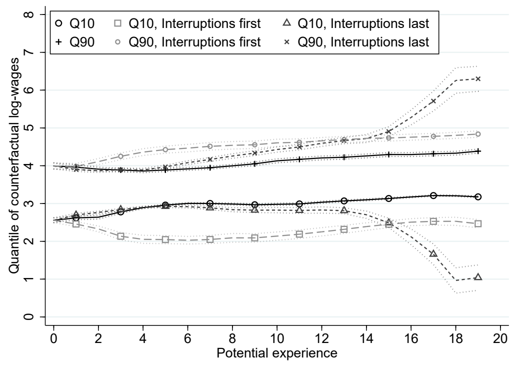
(b) Corrected median



(c) Corrected inter-decile range



(d) Corrected deciles



Note: Adjusted log-wages are computed from raw wages by (i) changing the timing of cohorts such that all cohorts enter the labour market in 1984; (ii) deflating wages with skill-specific prices of human capital; (iii) adding time-constant skill-specific prices of human capital. “Corrected” statistics are obtained after bias correction as described in Supplementary Appendix S.3.6. For this counterfactual exercise, we focus on dynamic selection effects while static selection is neutralized. This is why “No selection” is our benchmark. “Interruptions first” corresponds to the case where all the years of interruption are assigned to the first years of observations (except the very first year). “Interruptions last” corresponds to the case where all the years of interruption are assigned to the last years of the period during which the individual is observed.

capital accumulation. Shifting these advantageous interruptions at the beginning of the career yields higher wages.

We finally consider the counterfactual when artificially moving years of interruptions to the period just preceding the last year of observation before final attrition. The rise in the inter-decile range first parallels the trend observed in Figure 6(c) before taking off more steeply after 15 years. This rise is probably due to a combination of two effects. First, suppressing interruptions at the beginning of the life cycle mildly increases the wage dispersion during the first part of the life cycle as it was found before in Section 8.1. Second, reallocating interruptions at the very end magnifies their impacts on wage dispersion in a dramatic way because there is more divergence of wage trajectories with potential experience before the last years.

8.5 Evaluating sample selection effects

Our working sample excludes workers who have very incomplete histories, since we selected out workers for whom we have fewer than fifteen years of observation.

One can wonder whether such a selection has an effect on our results on wage dispersion. For instance, there is evidence within our working sample of 15+ observations that this dispersion increases with the number of years of interruption as shown in Table 3. There are two explanations: The first one is an asymptotic bias in $1/T^2$ that remains when using our estimation method; The second one is the substantive fact that the more incomplete the employment histories, the more dispersed the wage profiles.

We now summarize an experiment detailed in Gobillon *et al.* (2022) in which we assess the relative importance of these two explanations. We consider three samples: The first one that we denote (10/14) comprises individuals who are observed working in the private sector between 10 and 14 years between their entry year and 2011. By construction, they are excluded from our working sample. The second sample, denoted (20+), is a subsample of our working sample which comprises individuals who are observed more than 20 years. Using sample (20+), we construct a reduced sample $(20+, Censored)$ by randomly drawing the number of years in the private sector for every individual in such a way that the marginal distribution of this variable in sample $(20+, Censored)$ is the same as its marginal distribution in the first sample (10/14).

We then estimate individual specific coefficients for the 2 samples (10/14) and $(20+, Censored)$ while using factors estimated using our working sample. Estimates in both samples are compared to the original estimates for individuals in sample (20+). In sum, we show that results remain unchanged when we (artificially) decrease the number of observations from sample (20+) to sample $(20+, Censored)$. This suggests that bias correction works accurately, and that biases in $1/T^2$ are negligible as these negligible differences between sample $(20+, Censored)$ and sample (20+) are only due to the exogenous reduction of the sample size. We also find large differences between estimates in sample $(10-14)$ and sample $(20+, Censored)$ although those could be attributed to structural differences between the two sub-populations or to our specific way of reducing sample size.

8.6 Robustness checks

We implemented robustness checks on counterfactuals when varying the year of entry, $t_{i,0}$, and the discount factor, β. We now discuss the corresponding results which are reported in Supplementary Tables S.3 and S.4.

Considering first the entry year $t_{i,0}$, note again that there is a degree of arbitrariness in its definition. The first year in the model could be any year after entry provided that we condition on the stock of human capital in that year as an unobservable. We chose for $t_{i,0}$ the first year in which a full-time job duration in the private sector is above a threshold of 180 days is observed. Consequently, this entry year $t_{i,0}$ is individual specific and exogenous conditionally on factors and factor loadings in line with our MARCOF assumption. We also experimented with other thresholds, 90, 210, 240, 270 and 360 days, which entail some variations in the working samples. Results on model selection and counterfactual experiments are very similar across thresholds.

Second, we conducted a sensitivity analysis when varying the value of the discount factor, β. This parameter was set to different values in the range from 0.93 to 0.99, a usual range in the literature of microeconometric dynamic models, and the whole analysis was repeated for each value. We found that the likelihood criteria is quite insensitive to the discount factor value, and this confirms that this parameter is hard to identify and estimate. Moreover all results, included counterfactuals, are barely affected when changing the value of β across the $0.93 - 0.99$ range.

9 Conclusion

In this paper, we estimated a model of human capital acccumulation with lots of individual heterogeneity to assess how wage inequalities build up over the life cycle. We simultaneously deal with wage processes and missing data within the same structural economic model. Furthermore, our empirical strategy extends the common yet unconvincing restrictive MAR assumptions. We propose an assumption of missing at random conditionally on factors and factor loadings (MARCOF), which is weaker than usual MAR assumptions.

In our empirical application, we use French administrative data for young cohorts of males entering the private sector between 1985 and 1993, and who are followed until 2011. Life cycle inequalities within cohorts can be accurately measured in our working sample since wage disparities in the working population remain stable during the 1985-2011 period in France.

We show that dynamic selection effects are important using location and dispersion summaries of wage profiles, whereas static selection effects are much weaker. Past interruptions in participation in the private sector decrease mean and median wages as expected. Interestingly, wage dispersion becomes larger when we set the number of years of interruptions to zero. This increase in wage dispersion can be partly attributed to the endogeneity of past participation choices.

To save on space, we chose to display our results in terms of wage profiles. We could also have produced other statistics of interest such as the discounted sums of log wages, e.g. the integral of wage profiles. One can also wonder about the external validity of our results since we restricted our working sample to individuals who participated in the private sector during at least 15 years between 1985 and 2011. Future research should explore the empirical strategy that consists in using restrictions on individual heterogeneity that would weaken further sample selection issues. Trade-offs exist however because these restrictions are affecting precisely what we want to measure, i.e. wage inequalities.

Acknowledgments

We are grateful to Stephane Bonhomme, Xavier d'Haultfeuille, Koen Jochmans, Francesca Molinari and three referees as well as participants of seminars and conferences at Sciences Po, French Econometrics'18, Luxembourg, IAAE'19, EALE'19, Western Ontario, Marseilles, Venice, Tilburg, Carnegie Mellon, KU Leuven, Oxford, IPDC'23, Paris-Dauphine and TSE for their helpful comments. Access to the data was carried through the Secure Data Access Center (CASD) dedicated to researchers authorised by the French *Comité du secret statistique*.

Funding

We acknowledge funding from the Institut Universitaire de France and ANR under grants ANR-17-EURE-0001, ANR-17-EURE-0010 (Investissements d'Avenir programs) and ANR-20-CE26-0011.

References

Aakvik, Arild, James J. Heckman, and Edward Vytlacil. 2005. Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to norwegian vocational rehabilitation programs. *Journal of Econometrics* 125:15–51.

Abrevaya, Jason and Stephen G. Donald. 2017. A GMM approach for dealing with missing data on regressors. *Review of Economics and Statistics* 99:657–662.

Ahn, Seung, Young H. Lee, and Peter Schmidt. 2001. GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics* 101(2):219–255.

Alan, Sule, Martin Browning, and Mette Ejrnaes. 2018. Income and consumption: A micro semistructural analysis with pervasive heterogeneity. *Journal of Political Economy* 126:1827–1864.

Altuğ, Sumru and Robert A. Miller. 1998. The effect of work experience on female wages and labour supply. *The Review of Economic Studies* 65(1):45–85.

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme. 2017. Earnings and consumption dynamics: a nonlinear panel data framework. *Econometrica* 85(3):693–734.

Arellano, Manuel and Stéphane Bonhomme. 2012. Identifying distributional characteristics in random coefficient panel data models. *The Review of Economic Studies* 79(3):987–1020.

Bai, Jushan. 2009. Panel data models with interactive fixed effects. *Econometrica* 77(4):1229–1279.

Bai, Jushan and Serena Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70(1):191–221.

Balakrishnan, Sivaraman, Martin J. Wainwright, and Bin Yu. 2017. Statistical guarantees for the em algorithm: From population to sample-based analysis. *The Annals of Statistics* 45(1):77–120.

Ben-Porath, Yoram. 1967. The production of human capital and the life-cycle of earnings. *Journal of Political Economy* 75(4, Part.1):352–365.

Beyhum, Jad and Eric Gautier. 2022. Square-root nuclear norm penalized estimator for panel data models with approximately low-rank unobserved heterogeneity. *Journal of Business & Economic Statistics* 41(1):270–281.

Biewen, Martin, Bernd Fitzenberger, and Jakob de Lazzer. 2018. The role of employment interruptions and part-time work for the rise in wage inequality. *IZA Journal of Labor Economics* 7(1):1–34.

Blevins, Jason R. 2014. Nonparametric identification of dynamic decision processes with discrete and continuous choices. *Quantitative Economics* 5:531–554.

Blundell, Richard, Monica Costa-Dias, Costas Meghir, and Johnathan M. Shaw. 2016. Female labor supply, human capital, and welfare reform. *Econometrica* 84(5):1705–1753.

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. Consumption inequality and partial insurance. *American Economic Review* 98:1887–1921.

Blundell, Richard and James L. Powell. 2003. Endogeneity in nonparametric and semiparametric regression models. In M. Dewatripont, L.P Hansen, and S.J. Turnovsky (eds.) *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. II.* Cambridge University Press, 312–357.

Bowlus, Audra J. and Chris Robinson. 2012. Human capital prices, productivity, and growth. *American Economic Review* 102(7):3483–3515.

Browning, Martin and Thomas F. Crossley. 2001. The life-cycle model of consumption and saving. *Journal of Economic Perspectives* 15:3–22.

Browning, Martin, Mette Ejrnaes, and Javier Alvarez. 2012. Modelling income processes with lots of heterogeneity. *Review of Economic Studies* 77(4):1353–1381.

Browning, Martin, Lars P. Hansen, and James J. Heckman. 1999. Micro data and general equilibrium models. In *Handbook of Macroeconomics*. Elsevier Science, Amsterdam, 543–633.

Bunting, Jackson, Paul Diegert, and Arnaud Maurel. 2024. Heterogeneity, uncertainty and learning: Semiparametric identification and estimation. NBER Working Paper, n°32164.

Carneiro, Pedro, Hansen Karsten T., and Heckman James J. 2003. 2001 lawrence r. klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* 44:316–422.

Carroll, Christopher. 1997. Buffer-stock saving and the life-cycle/permanent income hypothesis. *The Quarterly journal of economics* 112:1–55.

Chernozhukov, Victor, Ivan Fernández-Val, Jinyong Hahn, and Whitney Newey. 2013. Average and quantile effects in nonseparable panel models. *Econometrica* 81(2):535–580.

Das, Tirthatanmoy and Solomon W. Polachek. 2019. Microfoundations of earnings differences. In T. Raa and Greene W. (eds.) *The Palgrave Handbook of Economic Performance Analysis*. Palgrave Macmillan, Cham, 9–76.

Deming, David J. 2023. Multidimensional human capital and the wage structure. Working Paper 31001, National Bureau of Economic Research.

Dominitz, Jeff and Robert P. Sherman. 2005. Some convergence theory for iterative estimation procedures with an application to semiparametric estimation. *Econometric Theory* 21(4):838–863.

Eckstein, Zvi and Kenneth I. Wolpin. 1989. Dynamic labour force participation of married women and endogenous work experience. *The Review of Economic Studies* 56(3):375–390.

Eika, Lasse, Magne Mogstad, and Ola Vestad. 2020. What can we learn about household consumption expenditure from data on income and assets? *Journal of Public Economics* 189.

Eisenhauer, Philipp, James J. Heckman, and Edward Vitlacyl. 2015. The generalized roy model and the cost-benefit analysis of social programs. *Journal of Political Economy* 123:413–443.

Fernández-Val, Ivan, Hugo Freeman, and Martin Weidner. 2021. Low-rank approximations of nonseparable panel models. *The Econometrics Journal* 24:C40–C77.

Fernández-Val, Ivan and Martin Weidner. 2018. Fixed effects estimation of large-T panel data models. *Annual Review of Economics* 10:109–138.

Gobillon, Laurent and Thierry Magnac. 2016. Regional policy evaluation: Interactive fixed effects and synthetic controls". *Review of Economics and Statistics* 98(3):535–551.

Gobillon, Laurent, Thierry Magnac, and Sébastien Roux. 2022. Lifecycle wages and human capital investments: Selection and missing data. TSE Working Papers 22-1299, Toulouse School of Economics.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2021. What do data on millions of US workers reveal about life-cycle earnings risk? *Econometrica* 89(5):2303–2339.

Heckman, James J. and V. Joseph Hotz. 1989. Alternative methods for evaluating the impact of training programs. *Journal of the American Statistical Association* 84:862–874.

Heckman, James J., Lance Lochner, and Christopher Taber. 1998. Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics* 1:1 – 58.

Heckman, James J., Lance Lochner, and Petra E. Todd. 2006. Earnings functions, rates of return and treatment effects: the mincer equation and beyond. In E. A. Hanushek and F. Welch (eds.) *Handbook of Economics of Education*. Elsevier Science, Amsterdam, 307–458.

Heckman, J.J and Jeffrey Smith. 1999. The pre-programme earnings dip and the determinants of participation in a social programme. *The Economic Journal* 109:313–34.

Heyde, Chris C. and Richard Morton. 1996. Quasi-likelihood and generalizing the em algorithm. *Journal of the Royal Statistical Association B* 59(2):317–327.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen. 1988. Estimating vector autoregressions with panel data. *Econometrica* 56(6):1371–1395.

Hsiao, Cheng. 2018. Panel models with interactive effects. *Journal of Econometrics* 206(2):645–673.

Jappelli, Tullio and Luigi Pistaferri. 2017. *The economics of consumption: theory and evidence*. Oxford: Oxford University Press.

Jiang, Bin, Yanrong Yang, Jiti Gao, and Cheng Hsiao. 2021. Recursive estimation in large panel data models: Theory and practice. *Journal of Econometrics* 224:439–465.

Jochmans, Koen and Martin Weidner. 2024. Inference on a distribution from noisy draws. *Journal of Econometrics* 40(1):60–97.

Kaplan, Greg, Giovanni Violante, and Justin Weidner. 2014. The wealthy hand-to-mouth. *Brookings Papers on Economic Activity* 1:77–153.

Keane, Michael P. and Nada Wasi. 2016. Labour supply: the roles of human capital and the extensive margin. *The Economic Journal* 126(592):578–617.

Kline, Patrick and Andres Santos. 2013. Sensitivity to missing data assumptions: Theory and an evaluation of the us wage structure. *Quantitative Economics* 4(2):231–267.

Kolsrun, Jonas, Camille Landais, Peter Nilsson, and Johannes Spinnewijn. 2018. The optimal timing of unemployment benefits: Theory and evidence from sweden. *American Economic Review* 108:985–1033.

Lagakos, David, Benjamin Moll, Tomaso Porzio, Nancy Qian, and Todd Schoellman. 2018. Life cycle wage growth across countries. *Journal of Political Economy* 126(2):797–849.

Light, Audrey and Manuelita Ureta. 1995. Early-career work experience and gender wage differentials. *Journal of Labor Economics* 13(1):121–154.

Lise, Jeremy and Fabien Postel-Vinay. 2020. Multidimensional skills, sorting, and human capital accumulation. *American Economic Review* 110(8):2328–76.

Little, Roderick J. and Donald B. Rubin. 2019. *Statistical analysis with missing data*. 3 edition. Wiley.

Magnac, Thierry, Nicolas Pistolesi, and Sébastien Roux. 2018. Post schooling human capital investments and the life-cycle variance of earnings. *Journal of Political Economy* 126(3):1219–1249.

Magnac, Thierry and Sébastien Roux. 2021. Heterogeneity and wage inequalities over the life-cycle. *European Economic Review* 134:1–23.

Meghir, Costas and Luigi Pistaferri. 2010. Earnings, consumption and life-cycle choices. In O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics, Vol. 4b*. Elsevier Science, Amsterdam, 773–854.

Mincer, Jacob A. 1974. *Schooling experience and earnings*. National Bureau of Economic Research, New York.

Moon, Hyungsik R. and Martin Weidner. 2019. Nuclear norm regularized estimation of panel regression models. Working Paper 14/19, National Bureau of Economic Research.

Pesaran, M. Hashem. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74(4):967–1012.

Polachek, Solomon, Tirthatanmoy Das, and Rewat Thamma-Apiroam. 2015. Heterogeneity in the production of human capital. *Journal of Political Economy* 123(6):1410–1455.

Roys, Nicolas A. and Christopher R. Taber. 2019. Skill prices, occupations, and changes in the wage structure for low skilled men. NBER Working Paper 26453.

Rubinstein, Yona and Yoram Weiss. 2006. Post schooling wage growth: investment, search and learning. In E. Hanushek and F. Welch (eds.) *Handbook of the Economics of Education*. Elsevier, Amsterdam, 1–67.

Sasaki, Yuya. 2015. Heterogeneity and selection in dynamic panel data. *Journal of Econometrics* 188(1):236–249.

Song, Minkee. 2013. Asymptotic theory for dynamic heterogenous panels with cross section dependence and its applications. Unpublished manuscript.

Su, Liangjun and Gaosheng Ju. 2018. Identifying latent grouped patterns in panel data models with interactive fixed effects. *Journal of Econometrics* 206:554–573.

A Glossary of Variables and Parameters

Name	Object	Values	Definition
$s_{i,t}$	Sector of activity	e, n	Page 11
$t_{i,0}$	Year of first entry	$\mathbb{N}$	11
$w_{i,t}^s$	Wages in year t and sector s	$\mathbb{R}^+$	11
$\delta_{i,t}^s$	Log price of human capital	$\mathbb{R}$	11
$H_{i,t}$	Human capital stock at time t	$\mathbb{R}^+$	11
$\tau_{i,t}^s$	Optimal human capital investment	$\mathbb{R}^+$	11
ρ_i^s	Rate of return of investment in sector s	$\mathbb{R}^+$	11
$\lambda_{i,t}^s$	Human capital depreciation shock	$\mathbb{R}^+$	11
c_i	Marginal cost of investment	$\mathbb{R}^+$	12
$\psi_{i,t}$	Private sector specific utility	$\mathbb{R}$	12
$Z_i^{(t)}$	Shocks revealed before time t	$\{\psi_{i,u}\}_{u=t_{i,0},.,t}$	13
$Z_i^{(t+1/2)}$	Shocks revealed before time $t+1/2$	$\{\delta_{i,u}^s, \lambda_{i,u}^s\}_{u=t_{i,0},.,t}$	13
β	Discount rate	0.95	13
$V_{i,t}$	Value function at time t	$\mathbb{R}$	13
$W_{i,t}^s$	Sector specific value function	$\mathbb{R}$	13
T_i	Horizon of investment	$\mathbb{R}$	14
$a_{i,t}(.)$	Levels of value functions	$\mathbb{R}$	14
$\kappa_{i,t}$	Marginal value of human capital stock	$\mathbb{R}^+$	14
K_i	Number of private sector interruptions	$\mathbb{N}$	17
$t_{i,2k}(t_{i,2k+1})$	Entry (resp.exit) years into private sector	$k \in \{0,.,K_i-1\}$	17

Name	Object	Values	Definition
$x_{i,t}^{(3)}$	Time spent outside sector e	$\mathbb{R}^+$	Page 18
$x_{i,t}^{(4)}$	Discounted years outside sector e	$\mathbb{R}^+$	18
$x_{i,t}$	$x_{i,t} = (1, t, \beta^{-t}, x_{i,t}^{(3)}, x_{i,t}^{(4)})$	$\mathbb{R}^5$	17
$\eta_i = (\eta_{ik})_{k=0,1,2,3,4}$	Random coefficients, wage equation	$\mathbb{R}$	17
$v_{i,t}$	Residual shock in wage equation	$\mathbb{R}$	17
$\sigma_{i,t}$	Exogenous savings rate	$[0,1)$	18
C_i	Cost matrix	$\mathbb{R}^R$	19
$\varpi_{g,t}$	Skill-group price index	$\mathbb{R}$	22
$\varphi_t = (\varphi_{\psi,t}, \varphi_{\delta,t}^s, \varphi_{\lambda,t}^s)$	Linear factors in $\psi_{i,t}, \delta_{i,t}^s, \lambda_{i,t}^s$	$\mathbb{R}$	23
$\theta_i = (\theta_{\psi,i}, \theta_{\delta,i}^s, \theta_{\lambda,i}^s)$	Linear factor loadings in $\psi_{i,t}, \delta_{i,t}^s, \lambda_{i,t}^s$	$\mathbb{R}$	23
$\tilde{\psi}_{i,t}, \tilde{\delta}_{i,t}, \tilde{\lambda}_{i,t}^s$	Residuals of factorisation of $\psi_{i,t}, \delta_{i,t}, \lambda_{i,t}^s$	$\mathbb{R}$	23
$y_{i,t}$	Deflated wage	$\mathbb{R}$	22
$\varepsilon_{i,t}$	Residual wage shock	$\mathbb{R}$	24
$\varphi_{s,t}, \theta_{s,i}$	Factors and loadings in participation	$\mathbb{R}$	25
φ_t^*, θ_i^*	Full list of factors and loadings	$\mathbb{R}$	25
φ_t, θ_i	Final list of interactive effects	$\mathbb{R}$	25
d	Potential experience $(t - t_{i,0})$	$\mathbb{R}$	27
$S_i^{(d)}$	Counterfactual selection history	$\mathbb{R}$	28
$x_{i,t_{i,0}+d}^{(k)}(S_i^{(d)})$	Counterfactual experience terms	$\mathbb{R}$	28
$Q(.)$	Generator: mean, var. and quantiles		28
$Q_d^{(k)}, k = 1,.,4$	Counterfactual statistics	$\mathbb{R}$	30

S.1 Data appendix

S.1.1 Data construction

In the raw data, there are $4,720,011$ person-job-year observations in the public and private sectors over the 1976-2011 period corresponding to individuals born in the first four days of October of even years. When restricting the sample to males, we are left with $2,551,964$ observations. When considering only jobs in the private sector, we are left with $1,889,371$ observations, and when considering only full-time positions, the sample size decreases to $1,616,598$ observations. We also delete jobs for workers on a training period and apprentices, and this leaves us with $1,581,304$ observations. Once jobs are aggregated per individual-year, we end up with $1,445,603$ observations. We ignore overlaps of job spells because they are exceptional for full-time jobs.

We then exclude jobs in which the wage is lower than 80% of the minimum wage. To compute the minimum wage, we use a national time series of gross hourly values. Over the 1976-1998 period, we transform them into monthly values by multiplying them with the number of working hours fixed legally to 169 (i.e. 39 hours per week). After 1998, some firms change their number of working hours to 151.67 (i.e. 35 hours per week) and this becomes the legal number in 2001. Therefore, from 1999 onwards, we compute two monthly values depending on whether the number of working hours is 169 or 151.67, and we consider that there is a transition over the 1999-2006 period between the two values consistently with the evolution of the proportion of individuals working 35 hours per week.[13] From 2007 onwards, we consider that the number of working hours is 151.67. We then decrease monthly values by 20% to remove payroll taxes and obtain net monthly values. The deletion of observations such that the wage is lower than 80% of the minimum wage makes the sample decrease to $1,431,109$ observations.

We keep only individual-year observations when the total amount of working days is larger than 6 months, and the sample then includes $1,253,730$ observations corresponding to $110,523$ males. We keep only observations for individuals entering the labor market over the 1985-1992 period (i.e. individuals observed for the first time in the panel during that period), and we are left with $210,810$ observations corresponding to $15,661$ males. After restricting the sample to individuals aged $16-30$, our sample includes $186,351$ corresponding to $12,707$ males. We delete observations for which individuals are older than 50 because we assume that human capital

[13]We use as proportions for every year over the 1999-2006 period: 10%, 20%, 30%, 40%, 60%, 70%, 80% and 90%.

investments become negligible after this age (see Section S.1.2 below) and this leaves us with 184,684 observations. Finally, we keep individuals who were present at least 15 years, which leaves us with 7,339 individuals with 145,166 observations.

The education level is defined as the highest diploma obtained by individuals. Using French diploma names, high-school drop-outs include no diploma, CAP, BEPC and CEP; high-school graduates include baccalauréat and low-level technical diplomas; short-track college graduates gather BTS, DUT and DEUG diploma holders; college graduates include 3-year and more college diplomas and Grandes Ecoles.

When constructing potential experience since entry in the private sector, we have to deal with the issue that no information is available in 1990. We use an imputation rule to fill the hole that year for employment in the private sector. We consider that a worker is employed (resp. non-employed) in 1990, if she was already employed (resp. non-employed) in 1989.

S.1.2 Human capital prices

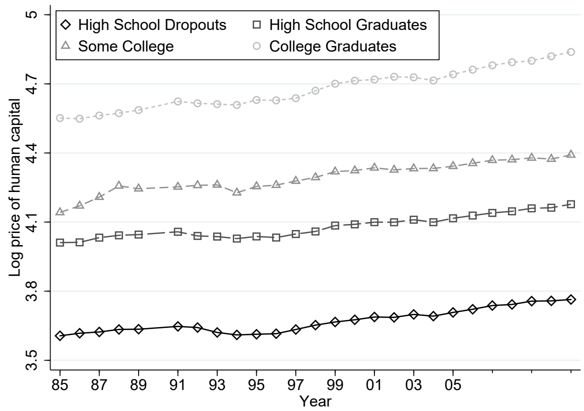
In Section 5.1, we deflated log wages by human capital price indexes whose growth is given by:

$$\Delta\varpi_{g,t} = \mathbb{E}(\Delta \log w_{i,t} \mid i \in g, t, a_{i,t} \in [50, 55]) \tag{50}$$

In practice, we compute these quantities as the log-wage growth of individuals aged 50-55 working in the private sector at both dates $t-1$ and t for each skill group g. To get human capital price indexes in level, we add to these growth terms the average log-wage computed on the same skill-group subsample over the whole study period. The resulting index $\varpi_{g,t}$ can be interpreted as a net human capital price index that is used to deflate wages and obtain the profiles of human capital stocks in equation (26). Note that we use levels $\varpi_{g,t}$ and not first differences $\Delta\varpi_{g,t}$ which will distort the estimation of the level effects or parameters η_{i0}, but in an homogeneous way across individuals within each skill group.

In our application, we consider that skill groups are diploma groups. Figure S.1 displays the time profiles of $\varpi_{g,t}$ resulting from the estimation method of flat spots by diploma group. These prices are deflated by the INSEE Consumer Price Index. In contrast with the US (Bowlus and Robinson, 2012), human capital prices are slightly increasing for all diploma groups. The increase over the $1985-2012$ period is higher for high skill groups (33% for college graduates and 28% for some college) than for low skill groups (18% for high-school graduates and 17% for high-school dropouts).

Figure S.1: Log-price of human capital by diploma group



Note: The price of human capital is measured with the mean log-wage growth for individuals aged 50-55 in the whole population of individuals in the private sector (that includes all the cohorts and not only the ones selected for our study). This price is deflated by the INSEE Consumer Price Index. Computations are stratified by diploma group.

S.1.3 Mincer regressions

As a descriptive device, we also ran homogenous Mincer regressions for deflated wages without and with correction for static selection into the private sector (using Mill's ratio with marriage and children variables as exclusion restrictions in the selection equation). Estimates are reported in Table S.1. Coefficient estimates of the interruption variables ($x^{(3)}$ and $x^{(4)}$) are respectively significantly negative and positive even when the selection correction term is introduced. In fact, the effect of this selection term is not significant and its inclusion does not affect much the estimates for other coefficients (except to some extent that of $x^{(4)}$). We can draw three partial conclusions before the full analysis with heterogeneous parameters: (i) years of interruptions in the participation to the private sector negatively affect potential wages and this indicates that returns to human capital investments are lower outside the private sector; (ii) interruptions move the Mincer dip to a lower value of potential experience; (iii) the effect of static selection is weak.

Table S.1: Mincer regression in line with the theoretical model

	(1)	(2) 2^{nd} stage	(3) 1^{st} stage (probit)
$x_{it}^1 = t$	0.073***	0.072***	0.038***
	(0.001)	(0.001)	(0.005)
$x_{it}^2 = \beta^{-t}$	-0.523***	-0.491***	-1.542***
	(0.012)	(0.028)	(0.054)
x_{it}^3	-0.078***	-0.071***	-0.393***
	(0.003)	(0.006)	(0.009)
x_{it}^4	0.377***	0.215	8.796***
	(0.037)	(0.134)	(0.111)
Married			-0.082***
			(0.017)
Marriage tenure			0.007***
			(0.002)
Having a child			0.003
			(0.019)
Number of children 3+			-0.055***
			(0.013)
Number of children 18+			0.017
			(0.010)
$\lambda(\hat{p}_{it}^*)$		-0.015	
		(0.012)	
Cohort fixed effects	X	X	X
N	145166	145166	165808
R^2	0.084	0.084	

Note: Column (1) reports OLS estimates. Column (2) reports OLS estimates when including a Mill's ratio in the specification to take into account selection. Results of the probit model used to compute the Mill's ratio are presented in Column (3).

S.2 Theoretical appendix

S.2.1 Proof of Proposition 1

Consider an individual who evaluates the consequences of working in sector s and choosing human capital investments, $\tau^s_{i,t}$, whether it is positive or equal to zero.

The marginal value of human capital can be expressed as the derivative of the interim value function $W^s_{i,t}$ with respect to the level of human capital. Using the envelope theorem if $\tau^s_{i,t}$ is an interior solution, or replacing with the corner solution, $\tau^s_{i,t} = 0$, we have that for any $H_{i,t}$:

$$\frac{\partial W^s_{i,t}}{\partial H_{i,t}} = \frac{1}{H_{i,t}} + \beta \left\{ \exp\left(\rho^s_i \tau^s_{i,t} - \lambda^s_{i,t}\right) \mathbb{E}_{t+1/2} \left[\frac{\partial V_{i,t+1}}{\partial H_{i,t+1}} \right] \right\} \tag{51}$$

$$= \frac{1}{H_{i,t}} + \frac{H_{i,t+1}}{H_{i,t}} \beta \mathbb{E}_{t+1/2} \left[\frac{\partial V_{i,t+1}}{\partial H_{i,t+1}} \right] \tag{52}$$

since we have $\frac{H_{i,t+1}}{H_{i,t}} = \exp\left(\rho^s_i \tau^s_{i,t} - \lambda^s_{i,t}\right)$ whenever s is chosen by the individual. This expression is equivalent to:

$$H_{i,t} \frac{\partial W^s_{i,t}}{\partial H_{i,t}} = 1 + \beta \mathbb{E}_{t+1/2} \left[H_{i,t+1} \frac{\partial V_{i,t+1}}{\partial H_{i,t+1}} \right],$$

and implies that:

$$H_{i,t} \mathbb{E}_t \frac{\partial W^s_{i,t}}{\partial H_{i,t}} = 1 + \beta \mathbb{E}_t \left[H_{i,t+1} \frac{\partial V_{i,t+1}}{\partial H_{i,t+1}} \right]. \tag{53}$$

We proceed by backward induction to show that $H_{i,t} \frac{\partial V_{i,t}}{\partial H_{i,t}}$ is a constant independent of $H_{i,t}$ and $Z^{(t)}_i$. Note first that for $t = T_i + 1$, specification (8) implies:

$$\frac{\partial V_{i,T_i+1}}{\partial H_{i,T_i+1}} = \frac{\kappa_i}{H_{i,T_i+1}} \Longrightarrow H_{i,T_i+1} \frac{\partial V_{T_i+1}}{\partial H_{T_i+1}} = \kappa_i, \tag{54}$$

in which κ_i does not depend on H_{i,T_i+1} or on $Z^{(T_i+1)}_i$. We assume that the property $H_{i,t+1} \frac{\partial V_{t+1}}{\partial H_{t+1}} = \kappa_{i,t+1}$, in which $\kappa_{i,t+1}$ is a constant that does not depend on $H_{i,t+1}$ or on $Z^{(t+1)}_i$, is true at year $t+1$ and we should prove that it is also true at year t to prove it for all years. Indeed, because of equation (53), derivatives on the left-hand side do not depend on s, i.e. $\mathbb{E}_t \frac{\partial W^e_{i,t}}{\partial H_{i,t}} = \mathbb{E}_t \frac{\partial W^n_{i,t}}{\partial H_{i,t}}$. This property is used to prove that:

$$H_{i,t} \frac{\partial V_{i,t}}{\partial H_{i,t}} = H_{i,t} \frac{\partial}{\partial H_{i,t}} (\max(\mathbb{E}_t W^e_{i,t} + \psi_{i,t}, \mathbb{E}_t W^n_{i,t})) = 1 + \beta \mathbb{E}_t \left[H_{i,t+1} \frac{\partial V_{i,t+1}}{\partial H_{i,t+1}} \right]. \tag{55}$$

We then have:

$$H_{i,t} \frac{\partial V_{i,t}}{\partial H_{i,t}} = \kappa_{i,t}, \tag{56}$$

where $\kappa_{i,t} = 1 + \beta E_t \left[H_{i,t+1} \frac{\partial V_{i,t+1}}{\partial H_{i,t+1}} \right] = 1 + \beta \kappa_{i,t+1}$ is a constant term independent of $H_{i,t}$ and $Z_i^{(t)}$. Using the initial condition (54) and backward induction, all values $\kappa_{i,t}$ are deterministic and we obtain:

$$\kappa_{i,t} = 1 + \beta \kappa_{i,t+1} \Longrightarrow \kappa_{i,t} - \frac{1}{1-\beta} = \beta(\kappa_{i,t+1} - \frac{1}{1-\beta}) \tag{57}$$

so that:

$$\kappa_{i,t} = \frac{1}{1-\beta} + \beta^{T_i+1-t}(\kappa_i - \frac{1}{1-\beta}). \tag{58}$$

By integration of equations (53) and (56), we obtain the value functions of Proposition 1 in which the arbitrary constants of integration, $a_{i,t}^s(Z_i^{(t+1/2)})$ and $a_{i,t}(Z_i^{(t)})$ are further defined below. ■

S.2.2 Proof of Proposition 2

The first order condition of the maximization problem for $t \in \{t_{i,0}, ., T_i\}$ with respect to the level of investment $\tau_{i,t}^s$ is

$$-\left(1 + c_i \tau_{i,t}^s\right) + \beta \rho_i^s \mathbb{E}_{t+1/2} \left[H_{i,t+1} \frac{\partial V_{i,t+1}}{\partial H_{i,t+1}} \right] = 0, \tag{59}$$

in which $H_{i,t+1}$ is determined by equation (2). This first order condition delivers a positive optimal human capital investment, $\tau_{i,t}^s > 0$, if the following condition holds:

$$\beta \rho_i^s \mathbb{E}_{t+1/2} \left[H_{i,t+1} \frac{\partial V_{i,t+1}}{\partial H_{i,t+1}} \right] > 1. \tag{60}$$

Using equation (56), this condition is equivalent to $\beta \rho_i^s \kappa_{i,t+1} > 1$ and equation (59) yields the optimal investment which verifies:

$$\left(1 + c_i \tau_{i,t}^s\right) = \beta \rho_i^s \kappa_{i,t+1}, \tag{61}$$

and the second term in equation (14) follows. When $\beta \rho_i^s \kappa_{i,t+1} \leq 1$, we obtain that $\tau_{i,t}^s = 0$. Furthermore, as the second left hand side term in (59) is constant, the second order condition is satisfied if and only if $c_i > 0$.

S.2.3 Proof of Proposition 3

By induction, define:

$$a_{i,t+1}(Z_i^{(t+1)}) = \mathbb{E}_{t+1}(a_{i,t+1}^{s_{i,t+1}}(Z_i^{(t+1)})) \tag{62}$$

Using Proposition 1, we have:

$$
\begin{aligned}
W^s_{i,t}(H_{i,t}, Z_i^{(t+1/2)}) &= \delta^s_{i,t} + \ln H_{i,t} - \left(\tau^s_{i,t} + c_i \frac{(\tau^s_{i,t})^2}{2}\right) + \beta \mathbb{E}_{t+1/2}\left[V_{i,t+1}\right] && (63)\\
&= \delta^s_{i,t} + \ln H_{i,t} - \left(\tau^s_{i,t} + c_i \frac{(\tau^s_{i,t})^2}{2}\right) && (64)\\
&\quad + \beta \mathbb{E}_{t+1/2}\left[a_{i,t+1}(Z_i^{(t+1)}) + \kappa_{i,t+1} \log H_{i,t+1}\right] && (65)\\
&= \delta^s_{i,t} + \ln H_{i,t} - \left(\tau^s_{i,t} + c_i \frac{(\tau^s_{i,t})^2}{2}\right) && (66)\\
&\quad + \beta \mathbb{E}_{t+1/2}\left[a_{i,t+1}(Z_i^{(t+1)}) + \kappa_{i,t+1}(\ln H_{i,t} + \rho^s_i \tau^s_{i,t} - \lambda^s_{i,t})\right].
\end{aligned}
$$

By identifying constant terms and using equation (61) and Proposition 2, we get:

$$
\begin{aligned}
a^s_{i,t}(Z_i^{(t+1/2)}) &= \delta^s_{i,t} + \left(\beta \kappa_{i,t+1} \rho^s_i \tau^s_{i,t} - \tau^s_{i,t} - c_i \frac{(\tau^s_{i,t})^2}{2}\right) - \beta \kappa_{i,t+1} \lambda^s_{i,t} && (67)\\
&\quad + \beta \mathbb{E}_{t+1/2}\left[a_{i,t+1}(Z_i^{(t+1)})\right], && (68)\\
&= \delta^s_{i,t} + c_i \frac{(\tau^s_{i,t})^2}{2} - \beta \kappa_{i,t+1} \lambda^s_{i,t} + \beta \mathbb{E}_{t+1/2}\left[a_{i,t+1}(Z_i^{(t+1)})\right]. && (69)
\end{aligned}
$$

S.2.4 Proof of Proposition 4

By equation (6), we have:

$$
\begin{aligned}
&\psi_{i,t} + \mathbb{E}_t\left[\delta^e_{i,t} + c\frac{(\tau^e_{i,t})^2}{2} - \beta \kappa_{i,t+1} \lambda^e_{i,t}\right] + \beta \mathbb{E}_t\left[a_{i,t+1}(Z_i^{(t+1)})\right] + \kappa_{i,t} \log(H_{i,t}) && (70)\\
\geq\ & \mathbb{E}_t\left[\delta^n_{i,t} + c_i \frac{(\tau^n_{i,t})^2}{2} - \beta \kappa_{i,t+1} \lambda^n_{i,t}\right] + \beta \mathbb{E}_t\left[a_{i,t+1}(Z_i^{(t+1)})\right] + \kappa_{i,t} \log(H_{i,t}).
\end{aligned}
$$

and noting that neither $H_{i,t}$ nor $E_t\left[a_{i,t+1}(Z_i^{(t+1)})\right]$ depend on current sector choice (absent any transition costs), we obtain condition (15). It also yields:

$$
\begin{aligned}
&a_{i,t}(Z_i^{(t)}) && (71)\\
=\ & \max(\psi_{i,t} + \mathbb{E}_t(\delta^s_{i,t} - \beta \kappa_{i,t+1} \lambda^s_{i,t} + c_i \frac{(\tau^s_{i,t})^2}{2}), \mathbb{E}_t(\delta^n_{i,t} - \beta \kappa_{i,t+1} \lambda^n_{i,t} + c_i \frac{(\tau^n_{i,t})^2}{2}))\\
&+ \beta \mathbb{E}_t\left[a_{i,t+1}(Z_i^{(t+1)})\right].
\end{aligned}
$$

S.2.5 Proof of Proposition 5

First, the stock of human capital in period t depends on previous investment choices and past depreciation, that is:

$$\begin{aligned}
H_{i,t} &= H_{i,t_{2K_{i,t}}} \exp\left[\sum_{l=t_{i,2K_{i,t}}}^{t-1} \rho_i^e \tau_{i,l}^e - \sum_{l=t_{i,2K_{i,t}}}^{t-1} \lambda_{i,l}^e\right] && (72)\\
&= H_{i,t_{2K_{i,t}-1}} \exp\left[\sum_{l=t_{i,2K_{i,t}}}^{t-1} \rho_i^e \tau_{i,l}^e - \sum_{l=t_{i,2K_{i,t}}}^{t-1} \lambda_{i,l}^e + \sum_{l=t_{i,2K_{i,t}-1}}^{t_{i,2K_{i,t}}-1} \rho_i^n \tau_{i,l}^n \right. && (73)\\
&\quad \left. - \sum_{l=t_{i,2K_{i,t}-1}}^{t_{i,2K_{i,t}}-1} \lambda_{i,l}^n \right]\\
&\quad \ldots\\
&= H_{i,t_{i,0}} \exp\left[\sum_{l=t_{i,0}}^{t-1} \rho_i^{s_{i,l}} \tau_{i,l}^{s_{i,l}} - \sum_{l=t_{i,0}}^{t-1} \lambda_{i,l}^{s_{i,l}}\right]. && (74)
\end{aligned}$$

At each date, we have that:

$$\tau_i^{s_{i,l}} = \max\left\{0, \frac{1}{c_i}\left(\rho_i^{s_{i,l}}\beta\kappa_{i,l+1} - 1\right)\right\} \tag{75}$$

As long as investments remain strictly positive in both sectors we have that:

$$\begin{aligned}
\ln H_{i,t} &= \ln H_{i,t_{i,0}} - \sum_{l=t_{i,0}}^{t-1} \lambda_{i,l}^{s_{i,l}} + \sum_{l=t_{i,0}}^{t-1} \rho_i^{s_{i,l}} \tau_{i,l}^{s_{i,l}} && (76)\\
&= \ln H_{i,t_{i,0}} - \sum_{l=t_{i,0}}^{t-1} \lambda_{i,l}^{s_{i,l}} + \sum_{l=t_{i,0}}^{t-1} \frac{\rho_i^{s_{i,l}}}{c_i}\left(\rho_i^{s_{i,l}}\beta\kappa_{i,l+1} - 1\right) && (77)
\end{aligned}$$

Using the sequence of periods in every sector and replacing $\kappa_{i,l+1}$ by its expression $\kappa_{i,l+1} = \frac{1}{1-\beta} + \beta^{T_i - l}(\kappa_i - \frac{1}{1-\beta})$ (see Proposition 1), the term $\sum_{l=t_{i,0}}^{t-1} \rho_i^{s_{i,l}} \frac{1}{c_i}\left(\rho_i^{s_{i,l}}\beta\kappa_{i,l+1} - 1\right)$ can be decomposed into:

$$\sum_{l=t_{i,0}}^{t-1} \frac{\rho_i^{s_{i,l}}}{c_i} \left(\rho_i^{s_{i,l}} \beta \kappa_{i,l+1} - 1\right) \tag{78}$$

$$= \sum_{k=0}^{K_{i,t}-1} \sum_{l=t_{i,2k}}^{t_{i,2k+1}-1} \frac{\rho_i^e}{c_i} \left(\rho_i^e \frac{\beta}{1-\beta} - 1 + \rho_i^e \beta^{T_i+1-l} \left(\kappa_i - \frac{1}{1-\beta}\right)\right)$$
$$+ \sum_{l=t_{i,2K_{i,t}}}^{t-1} \frac{\rho_i^e}{c_i} \left(\rho_i^e \frac{\beta}{1-\beta} - 1 + \rho_i^e \beta^{T_i+1-l} \left(\kappa_i - \frac{1}{1-\beta}\right)\right)$$
$$+ \sum_{k=0}^{K_{i,t}-1} \sum_{l=t_{i,2k+1}}^{t_{i,2k+2}-1} \frac{\rho_i^n}{c_i} \left(\rho_i^n \frac{\beta}{1-\beta} - 1 + \rho_i^n \beta^{T_i+1-l} \left(\kappa_i - \frac{1}{1-\beta}\right)\right)$$

in which $K_{i,t}$ is equal to the number of spells in the alternative sector before year t. The first two right-hand-side terms correspond to the accumulation of human capital when the worker is in sector e and the last one when he is in sector n.

Since

$$\sum_{l=l_0}^{l_1-1} \frac{\rho_i^e}{c_i} \left(\rho_i^e \frac{\beta}{1-\beta} - 1 + \rho_i^e \beta^{T_i+1-l} \left(\kappa_i - \frac{1}{1-\beta}\right)\right)$$
$$= \frac{\rho_i^e}{c_i} \left(\rho_i^e \frac{\beta}{1-\beta} - 1\right)(l_1 - l_0) + \frac{(\rho_i^e)^2}{c_i} \beta^{T_i+1-l_0} \left(\kappa_i - \frac{1}{1-\beta}\right) \sum_{l=0}^{l_1-l_0-1} \beta^{-l} \tag{79}$$
$$= \frac{\rho_i^e}{c_i} \left(\rho_i^e \frac{\beta}{1-\beta} - 1\right)(l_1 - l_0) + \frac{(\rho_i^e)^2}{c_i} \frac{\beta^{T_i+1}}{1-\beta} \left(\kappa_i - \frac{1}{1-\beta}\right) \left(\beta^{-l_1} - \beta^{-l_0}\right), \tag{80}$$

the term $\sum_{l=t_{i,0}}^{t-1} \frac{\rho_i^{s_{i,l}}}{c_i} \left(\rho_i^{s_{i,l}} \beta \kappa_{i,l+1} - 1\right)$ simplifies into:

$$\frac{\rho_i^e}{c_i} \left(\rho_i^e \frac{\beta}{1-\beta} - 1\right) \sum_{k=0}^{K_{i,t}-1} (t_{i,2k+1} - t_{i,2k})$$
$$+ \frac{(\rho_i^e)^2}{c_i} \left(\kappa_i - \frac{1}{1-\beta}\right) \frac{\beta^{T_i+1}}{1-\beta} \sum_{k=0}^{K_{i,t}-1} \left(\beta^{-t_{i,2k+1}} - \beta^{-t_{i,2k}}\right) \tag{81}$$
$$+ \frac{\rho_i^e}{c_i} \left(\rho_i^e \frac{\beta}{1-\beta} - 1\right)\left(t - t_{i,2K_{i,t}}\right) + \frac{(\rho_i^e)^2}{c_i} \left(\kappa_i - \frac{1}{1-\beta}\right) \frac{\beta^{T_i+1}}{1-\beta} \left(\beta^{-t} - \beta^{-t_{i,2K_{it}}}\right)$$
$$+ \frac{\rho_i^n}{c_i} \left(\rho_i^n \frac{\beta}{1-\beta} - 1\right) \sum_{k=0}^{K_{i,t}-1} (t_{i,2k+2} - t_{i,2k+1})$$
$$+ \frac{(\rho_i^n)^2}{c_i} \left(\kappa_i - \frac{1}{1-\beta}\right) \frac{\beta^{T_i+1}}{1-\beta} \sum_{k=0}^{K_{i,t}-1} \left(\beta^{-t_{i,2k+2}} - \beta^{-t_{i,2k+1}}\right)$$

This term can be rearranged considering the differential accumulation of human capital between sectors e and n when the worker is in sector n. This leads to introducing the accumulation of human capital if the individual had been employed in sector e during the whole period:

$$\begin{aligned}
&\sum_{l=t_{i,0}}^{t-1} \frac{\rho_i^{s_{i,l}}}{c_i}\left(\rho_i^{s_{i,l}}\beta\kappa_{i,l+1}-1\right) \\
=\ & \frac{\rho_i^e}{c_i}\left(\rho_i^e\frac{\beta}{1-\beta}-1\right)\left\{\sum_{k=0}^{K_{i,t}-1}\left[(t_{i,2k+1}-t_{i,2k})+(t_{i,2k+2}-t_{i,2k+1})\right]+(t-t_{i,2K_{it}})\right\} \quad (82)\\
&+\left[\frac{\rho_i^n}{c_i}\left(\rho_i^n\frac{\beta}{1-\beta}-1\right)-\frac{\rho_i^e}{c_i}\left(\rho_i^e\frac{\beta}{1-\beta}-1\right)\right]\sum_{k=0}^{K_{i,t}-1}\left(t_{i,2k+2}-t_{i,2k+1}\right)\\
&+\frac{(\rho_i^e)^2}{c_i}\frac{\beta^{T_i+1}}{1-\beta}\left(\kappa_i-\frac{1}{1-\beta}\right)\left\{\sum_{k=0}^{K_{i,t}-1}\left[\beta^{-t_{i,2k+1}}-\beta^{-t_{i,2k}}+\beta^{-t_{i,2k+2}}-\beta^{-t_{i,2k+1}}\right]\right.\\
&\qquad\left.+\beta^{-t}-\beta^{-t_{i,2K_{i,t}}}\right\}\\
&+\frac{(\rho_i^n)^2-(\rho_i^e)^2}{c_i}\frac{\beta^{T_i+1}}{1-\beta}\left(\kappa_i-\frac{1}{1-\beta}\right)\sum_{k=0}^{K_{i,t}-1}\left(\beta^{-t_{i,2k+2}}-\beta^{-t_{i,2k+1}}\right)\\
=\ & \frac{\rho_i^e}{c_i}\left(\rho_i^e\frac{\beta}{1-\beta}-1\right)(t-t_{i,0})+\frac{(\rho_i^e)^2}{c_i}\frac{\beta^{T_i+1}}{1-\beta}\left(\kappa_i-\frac{1}{1-\beta}\right)\left(\beta^{-t}-\beta^{-t_{i0}}\right) \quad (83)\\
&+\left[\frac{\rho_i^n}{c_i}\left(\rho_i^n\frac{\beta}{1-\beta}-1\right)-\frac{\rho_i^e}{c_i}\left(\rho_i^e\frac{\beta}{1-\beta}-1\right)\right]\sum_{k=0}^{K_{i,t}-1}\left(t_{i,2k+2}-t_{i,2k+1}\right)\\
&+\frac{(\rho_i^n)^2-(\rho_i^e)^2}{c_i}\frac{\beta^{T_i+1}}{1-\beta}\left(\kappa_i-\frac{1}{1-\beta}\right)\sum_{k=0}^{K_{i,t}-1}\left(\beta^{-t_{i,2k+2}}-\beta^{-t_{i,2k+1}}\right)
\end{aligned}$$

Defining

$$x_{i,t}^{(3)} = \sum_{k=0}^{K_{i,t}-1}\left(t_{i,2k+2}-t_{i,2k+1}\right) \quad (84)$$

$$x_{i,t}^{(4)} = \sum_{k=0}^{K_{i,t}-1}\left(\beta^{-t_{i,2k+2}}-\beta^{-t_{i,2k+1}}\right) \quad (85)$$

and

$$\eta_{i3} = \frac{\rho_i^n}{c_i}\left(\rho_i^n\frac{\beta}{1-\beta}-1\right)-\frac{\rho_i^e}{c_i}\left(\rho_i^e\frac{\beta}{1-\beta}-1\right) \quad (86)$$

$$\eta_{i4} = \frac{1}{c_i}\frac{\beta^{T_i+1}}{1-\beta}\left(\kappa_i-\frac{1}{1-\beta}\right)\left((\rho_i^n)^2-(\rho_i^e)^2\right) \quad (87)$$

Human capital at date t has the following expression:

$$\begin{aligned}\ln H_{i,t} &= \ln H_{i,t_{i,0}} - \sum_{l=t_{i,0}}^{t-1} \lambda_{i,l}^{s_{i,l}} + \eta_{i3} x_{i,t}^{(3)} + \eta_{i4} x_{i,t}^{(4)} \\ &+ \frac{\rho_i^e}{c_i}\left(\rho_i^e \frac{\beta}{1-\beta} - 1\right)(t - t_{i,0}) \\ &+ \frac{(\rho_i^e)^2}{c_i(1-\beta)} \beta^{T_i+1}\left(\kappa_i - \frac{1}{1-\beta}\right)\left(\beta^{-t} - \beta^{-t_{i,0}}\right)\end{aligned} \tag{88}$$

This expression can then be plugged into the earnings equation which becomes:

$$\begin{aligned}\ln w_{i,t} &= \delta_{i,t} + \ln H_{i,t} - \tau_{i,t} && (89)\\ &= \delta_{i,t} + \ln H_{i,t_{i,0}} - \sum_{l=t_{i,0}}^{t-1} \lambda_{i,l}^{s_{i,l}} + \eta_{i3} x_{i,t}^{(3)} + \eta_{i4} x_{i,t}^{(4)} && (90)\\ &+ \frac{\rho_i^e}{c_i}\left(\rho_i^e \frac{\beta}{1-\beta} - 1\right)(t - t_{i,0}) + \frac{(\rho_i^e)^2}{c_i(1-\beta)} \beta^{T_i+1}\left(\kappa_i - \frac{1}{1-\beta}\right)\left(\beta^{-t} - \beta^{-t_{i,0}}\right) \\ &- \frac{1}{c_i}\left(\frac{\rho_i^e \beta}{1-\beta} + \rho_i^e \beta^{T_i+1-t}(\kappa_i - \frac{1}{1-\beta}) - 1\right) \\ &= \ln H_{i,t_{i,0}} - \frac{\rho_i^e t_{i,0} + 1}{c_i}\left(\frac{\rho_i^e \beta}{1-\beta} - 1\right) - \frac{(\rho_i^e)^2 \beta}{c_i} \frac{\beta^{T_i - t_{i,0}+1}}{1-\beta}\left(\kappa_i - \frac{1}{1-\beta}\right) && (91)\\ &+ \frac{\rho_i^e}{c_i}\left(\rho_i^e \frac{\beta}{1-\beta} - 1\right) t \\ &+ \frac{\rho_i^e}{c_i}\left(\frac{\rho_i^e \beta}{1-\beta} - 1\right) \beta^{T_i+1}\left(\kappa_i - \frac{1}{1-\beta}\right)\beta^{-t} \\ &+ \delta_{i,t} - \sum_{l=t_{i,0}}^{t-1} \lambda_{i,l}^{s_{i,l}} + \eta_{i3} x_{i,t}^{(3)} + \eta_{i4} x_{i,t}^{(4)}.\end{aligned}$$

We can then set

$$\eta_{i0} = \ln H_{i,t_{i,0}} - \frac{\rho_i^e t_{i,0} + 1}{c_i}\left(\frac{\rho_i^e \beta}{1-\beta} - 1\right) - \frac{(\rho_i^e)^2 \beta}{c_i} \frac{\beta^{T_i - t_{i,0}+1}}{1-\beta}\left(\kappa_i - \frac{1}{1-\beta}\right) \tag{92}$$

$$\eta_{i1} = \frac{\rho_i^e}{c_i}\left(\rho_i^e \frac{\beta}{1-\beta} - 1\right) \tag{93}$$

$$\eta_{i2} = \beta^{T_i+1} \frac{\rho_i^e}{c_i}\left(\kappa_i - \frac{1}{1-\beta}\right)\left(\frac{\rho_i^e \beta}{1-\beta} - 1\right) \tag{94}$$

$$\eta_{i3} = \left(\frac{\rho_i^n}{c_i}\left(\frac{\rho_i^n \beta}{1-\beta} - 1\right) - \frac{\rho_i^e}{c_i}\left(\frac{\rho_i^e \beta}{1-\beta} - 1\right)\right) \tag{95}$$

$$\eta_{i4} = \frac{1}{c_i}\left((\rho_i^n)^2 - (\rho_i^e)^2\right)\left(\kappa_i - \frac{1}{1-\beta}\right)\frac{\beta^{T_i+1}}{1-\beta} \tag{96}$$

and we obtain the reduced form given by equation (17).

S.2.6 Multidimensional human capital

We first develop the arguments of Section 4.4.2 in the case of a single sector. We then develop the restrictions under which the same arguments hold in the multisector case. The value function is a function of $H_{i,t}$ and $\tau_{i,t}$:

$$\begin{aligned} V_{i,t}(H_{i,t},\tau_{i,t}) &= \delta_{i,t}+\sum_{r=1}^{R}\delta_{i,r}\ln H_{i,r,t}-\sum_{r=1}^{R}\tau_{i,r,t}-\frac{\tau_{i,t}{}'C_i\tau_{i,t}}{2} \\ &\quad +\beta\mathbb{E}_t W_{i,t+1}(H_{i,t+1}) \end{aligned} \tag{97}$$

using the same notation as in the main text although we dropped the dependence on exogenous processes Z_i. We also write that:

$$W_{i,T_i+1}(H_{i,T_i+1})=a_{i,T_i+1}+\sum_{r=1}^{R}\kappa_{i,r}\log(H_{i,r,T_i+1}). \tag{98}$$

The first order condition is given by:

$$-1-e_r'C_i\tau_{i,t}+\beta\rho_{i,r}H_{i,r,t+1}\mathbb{E}_t\frac{\partial W_{i,t+1}}{\partial H_{i,r,t+1}}=0, \tag{99}$$

if e_r is a $(R,1)-$vector whose only non-null element is the r^{th} one which is equal to 1. Furthermore, the derivatives of the value function are given by:

$$\frac{\partial W_{i,t}}{\partial H_{i,r,t}}=\frac{\delta_{i,r}}{H_{i,r,t}}+\beta\frac{H_{i,r,t+1}}{H_{i,r,t}}E_t\frac{\partial W_{i,t+1}}{\partial H_{i,r,t+1}}. \tag{100}$$

When $t=T_i$, we get:

$$\frac{\partial W_{i,T_i}}{\partial H_{i,r,T_i}}=\frac{\delta_{i,r}}{H_{i,r,T_i}}+\beta\frac{\kappa_{i,r}}{H_{i,r,T_i}}, \tag{101}$$

so that:

$$H_{i,r,T_i}\frac{\partial W_{i,T}}{\partial H_{i,r,T_i}}=\delta_{i,r}+\beta\kappa_{i,r}=\kappa_{i,r,T_i}. \tag{102}$$

This implies that the constants $\kappa_{i,r,t}$ are obtained by backward induction as in the case of a single dimensional human capital:

$$\kappa_{i,r,t}=\delta_{i,r}+\beta\kappa_{i,r,t+1} \tag{103}$$

which leads to:

$$\kappa_{i,r,t+1}=\frac{1-\beta^{T_i-t}}{1-\beta}\delta_{i,r}+\beta^{T_i-t}\kappa_{i,r,T_i+1}=\frac{\delta_{i,r}}{1-\beta}+\beta^{T_i-t}(\kappa_{i,r,T_i+1}-\frac{\delta_{i,r}}{1-\beta}). \tag{104}$$

The first order conditions (99) in year t keep their simple structure:

$$-1 - e_r' C_i \tau_{i,t} + \beta \rho_{i,r} \kappa_{i,r,t+1} = 0. \tag{105}$$

Stacking over r yields:

$$-j_R - C_i \tau_{i,t} + \beta \rho_i \odot \kappa_{i,t+1} = 0, \tag{106}$$

in which we defined j_R the $(R, 1)$ vector of ones and $\rho_i \odot \kappa_{i,t+1} = (\rho_{i,r} \kappa_{i,r,t+1})_{r=1,.,R}$ the element-by-element vector multiplication. We thus obtain:

$$\tau_{i,t} = C_i^{-1} (\beta \rho_i \odot \kappa_{i,t+1} - j_R). \tag{107}$$

By using equation (104) and by defining $\delta_i \odot \rho_i = (\delta_{i,r} \rho_{i,r})_{r=1,.,R}$,

$$\rho_i \odot \kappa_{i,t+1} = \frac{\delta_i \odot \rho_i}{1 - \beta} + \beta^{T_i - t} (\kappa_{i,T_i+1} \odot \rho_i - \frac{\delta_i \odot \rho_i}{1 - \beta}), \tag{108}$$

and the first order condition (107) can be written as :

$$\tau_{i,t} = C_i^{-1} \left(\beta \frac{\delta_i \odot \rho_i}{1 - \beta} + \beta^{T_i - t + 1} (\kappa_{i,T_i+1} \odot \rho_i - \frac{\delta_i \odot \rho_i}{1 - \beta}) - j_R \right). \tag{109}$$

Assume that all elements of $\tau_{i,t}$ are non-negative for all t in such a way that investments continue in each type of human capital until at least T_i.

Replacing human capital stocks by their expressions in equation (22) yields:

$$\log(w_{i,t}) = \delta_{i,t} + \sum_{r=1}^{R} \delta_{i,r} \left(\ln H_{i,r,t_{i,0}} + \sum_{l=1}^{t-1} (\rho_{i,r} \tau_{i,r,l} - \lambda_{i,r,l}) \right) - \sum_{r=1}^{R} \tau_{i,r,t} \tag{110}$$

$$= \overline{\ln H_{i,t_{i,0}}} + \sum_{r=1}^{R} \delta_{i,r} \rho_{i,r} \left(\sum_{l=1}^{t-1} \tau_{i,r,l} \right) - \sum_{r=1}^{R} \tau_{i,r,t} + \zeta_{i,t}, \tag{111}$$

in which $\overline{\ln H_{i,t_{i,0}}} = \sum_{r=1}^{R} \delta_{i,r} \ln H_{i,r,t_{i,0}}$ and $\zeta_{i,t} = \delta_{i,t} - \sum_{r=1}^{R} \delta_{i,r} \left(\sum_{l=1}^{t-1} \lambda_{i,r,l} \right)$. Rewrite:

$$\log(w_{i,t}) = \overline{\ln H_{i,t_{i,0}}} + \sum_{l=1}^{t-1} (\delta_i \odot \rho_i)' \tau_{i,l} - j_R' \tau_{i,t} + \zeta_{i,t}. \tag{112}$$

Replace $\tau_{i,l}$ by its expression (109) to obtain:

$$\begin{aligned}
& \log(w_{i,t}) & (113)\\
= \; & \overline{\ln H_{i,t_{i,0}}} + \sum_{l=1}^{t-1}(\delta_i \odot \rho_i)' C_i^{-1} \left(\beta \frac{\delta_i \odot \rho_i}{1-\beta} - j_R + \beta^{T_i - l+1}(\kappa_i \odot \rho_i - \frac{\delta_i \odot \rho_i}{1-\beta}) \right) \\
& - j_R' C_i^{-1} \left(\beta \frac{\delta_i \odot \rho_i}{1-\beta} - j_R + \beta^{T_i - t+1}(\kappa_i \odot \rho_i - \frac{\delta_i \odot \rho_i}{1-\beta}) \right) + \zeta_{i,t} \\
= \; & \overline{\ln H_{i,t_{i,0}}} + (t-1)(\delta_i \odot \rho_i)' C_i^{-1} \left(\beta \frac{\delta_i \odot \rho_i}{1-\beta} - j_R \right) & (114)\\
& + (\delta_i \odot \rho_i)' C_i^{-1} \left(\kappa_i \odot \rho_i - \frac{\delta_i \odot \rho_i}{1-\beta} \right) \sum_{l=1}^{t-1} \beta^{T_i - l+1} \\
& - j_R' C_i^{-1} \left(\beta \frac{\delta_i \odot \rho_i}{1-\beta} - j_R \right) - \beta^{T_i - t+1} j_R' C_i^{-1} \left(\kappa_i \odot \rho_i - \frac{\delta_i \odot \rho_i}{1-\beta} \right) + \zeta_{i,t} \\
= \; & \overline{\ln H_{i,t_{i,0}}} + (t-1)(\delta_i \odot \rho_i)' C_i^{-1} \left(\beta \frac{\delta_i \odot \rho_i}{1-\beta} - j_R \right) & (115)\\
& + (\delta_i \odot \rho_i)' C_i^{-1} \left(\kappa_i \odot \rho_i - \frac{\delta_i \odot \rho_i}{1-\beta} \right) \left(\frac{\beta^{T_i - t+2} - \beta^{T_i+1}}{1-\beta} \right) \\
& - j_R' C_i^{-1} \left(\beta \frac{\delta_i \odot \rho_i}{1-\beta} - j_R \right) - \beta^{T_i - t+1} j_R' C_i^{-1} \left(\kappa_i \odot \rho_i - \frac{\delta_i \odot \rho_i}{1-\beta} \right) + \zeta_{i,t}.
\end{aligned}$$

We can then rewrite this equation as:

$$\log(w_{i,t}) = \eta_{i0} + \eta_{i1} t + \eta_{i2} \beta^{-t} + \zeta_{i,t} \tag{116}$$

in which:

$$\begin{aligned}
\eta_{i0} \; &= \; \overline{\ln H_{i,t_{i,0}}} - (\delta_i \odot \rho_i)' C_i^{-1} \left(\beta \frac{\delta_i \odot \rho_i}{1-\beta} - j_R \right) & (117)\\
& \quad - (\delta_i \odot \rho_i)' C_i^{-1} \left(\kappa_i \odot \rho_i - \frac{\delta_i \odot \rho_i}{1-\beta} \right) \frac{\beta^{T_i+1}}{1-\beta} - j_R' C_i^{-1} \left(\beta \frac{\delta_i \odot \rho_i}{1-\beta} - j_R \right), \\
\eta_{i1} \; &= \; (\delta_i \odot \rho_i)' C_i^{-1} \left(\beta \frac{\delta_i \odot \rho_i}{1-\beta} - j_R \right), & (118)\\
\eta_{i2} \; &= \; (\delta_i \odot \rho_i)' C_i^{-1} \left(\kappa_i \odot \rho_i - \frac{\delta_i \odot \rho_i}{1-\beta} \right) \frac{\beta^{T_i+2}}{1-\beta} & (119)\\
& \quad - \beta^{T_i+1} j_R' C_i^{-1} \left(\kappa_i \odot \rho_i - \frac{\delta_i \odot \rho_i}{1-\beta} \right).
\end{aligned}$$

The same line of arguments can be extended to two sectors under the assumption that the coefficients $\{\delta_{i,r}\}_{r=1,.,R}$ do not depend on the sector chosen by the individual. Nevertheless, coefficients $\{\rho_{i,r}\}_{r=1,.,R}$ and $\{\lambda_{i,r,t}\}_{r=1,.,R}$ can depend on the sector. Indeed, the thorny issue lies with the induction relationship (108):

$$\kappa_{i,r,t} = \delta^s_{i,r} + \beta\kappa_{i,r,t+1} \tag{120}$$

which breaks the separability between investment profile and sectoral choices when we allow for dependence of the composition of the portfolio of human capital types on sectoral choice, which leads to:

$$\kappa_{i,r,t+1} = \frac{1-\beta^{T_i-t}}{1-\beta}\delta_{i,r} + \beta^{T_i-t}\kappa_{i,r,T_i+1} = \frac{\delta_{i,r}}{1-\beta} + \beta^{T_i-t}(\kappa_{i,r,T_i+1} - \frac{\delta_{i,r}}{1-\beta}). \tag{121}$$

S.3 Econometric appendix

S.3.1 Proofs of lemmas and corollaries of Econometric Section 5

S.3.1.1 Proof of Lemma 6

Using equations (28), the participation condition (15) can be rewritten as:

$$\widetilde{\psi}_{i,t} + \mathbb{E}_t\left(\widetilde{\delta}^e_{i,t} - \beta\kappa^e_{i,t+1}\widetilde{\lambda}^e_{i,t}\right) - \mathbb{E}_t\left(\widetilde{\delta}^n_{i,t} - \beta\kappa^n_{i,t+1}\widetilde{\lambda}^n_{i,t}\right) \geq \varphi_{s,t}\theta_{s,i}, \tag{122}$$

in which the notation, $\mathbb{E}_t$, is defined in Section 4.1.2, and conditions on available information at the beginning of period t. The interactive effect, $\varphi_{s,t}\theta_{s,i}$, is a function of all factors and factor loadings in equations (28), and implicitly includes investment terms like $c_i(\tau^s_{i,t})^2/2$. Indeed, the terms τ^s_{it} in equation (15) depend themselves on ρ^s_i (and hence of η_i) through equation (14), and thus have a factor structure that enters into $\varphi_{s,t}\theta_{s,i}$. Furthermore, conditionally on factors, residual shocks $\widetilde{\delta}^s_{i,t}, \widetilde{\lambda}^s_{i,t}$ for $s \in \{s,n\}$ are independent of preference residual shock $\widetilde{\psi}_{i,t}$ because of the MARCOF assumption (34), and are independent of the history of any other shocks that enter the information set at time t, because of the assumption of independence over time (IND). As a consequence, we have $\mathbb{E}_t\left(\widetilde{\delta}^s_{i,t} - \beta\kappa_{i,t+1}\widetilde{\lambda}^s_{i,t}\right) = 0$ for $s = e, n$. The selection equation thus rewrites as:

$$\widetilde{\psi}_{i,t} \geq \varphi_{s,t}\theta_{s,i}. \tag{123}$$

S.3.1.2 Proof of Lemma 7

The term $\varepsilon_{i,t}$ was defined in equation (32) as:

$$\varepsilon_{i,t} = \widetilde{\delta}^e_{i,t} - \sum_{l=t_{i,0}}^{t-1} \widetilde{\lambda}^e_{i,l} 1\{s_{i,l} = e\} - \sum_{l=t_{i,0}}^{t-1} \widetilde{\lambda}^n_{i,l} 1\{s_{i,l} = n\} - \varpi_{g,t} \tag{124}$$

in which the participation index $1\{s_{i,l} = e\} = 1$ if and only if $\widetilde{\psi}_{i,l} \geq \varphi_{s,l}\theta_{s,i}$ as stated in Lemma 6. As shown in the proof of Lemma 6, $\varphi_{s,t}$ (respectively $\theta_{s,i}$) is a function of the history of factors (resp. of individual specific effects including factor loadings) and enters the list $\varphi^{(t)*}$ (resp. θ_i^*). The first term in $\varepsilon_{i,t}$, current price shock, $\widetilde{\delta}_{i,t}^e$, is independent of current preference shock $\widetilde{\psi}_{i,t}$ (Assumption MARCOF). Next, current preference shock $\widetilde{\psi}_{i,t}$ is independent of past depreciation shocks, $\widetilde{\lambda}_{i,l}^e$, and of past preference shocks $\widetilde{\psi}_{i,l}$ conditional on $\varphi^{(t)*}$ and θ_i^* because of Assumption IND. We thus obtain the first condition stated in the Lemma: $\varepsilon_{i,t} \perp\!\!\!\perp \widetilde{\psi}_{i,t} \mid (\varphi^{(t)*}, \theta_i^*)$. The second condition, $\varepsilon_{i,t} \perp\!\!\!\perp (\varphi^{(t)*}, \theta_i^*)$, is derived from the factor structure (28) and Assumption IND.

S.3.1.3 Proof of Corollary 8

Explanatory variables $x_{i,t}^{(3)}$ and $x_{i,t}^{(4)}$ are exogenous because they are functions of past sectoral choices as stated in equations (18). Past sectoral choices in turn depend on past preference, price and depreciation shocks of which $\varepsilon_{i,t}$ is independent conditional on factors as shown in Lemma 7. This yields:

$$E\left(\varepsilon_{i,t} \middle| x_{i,t}^{(3)}, x_{i,t}^{(4)}, Z_i^{(t-1/2)}, \varphi_t, \eta_i, \theta_i\right) = E\left(\varepsilon_{i,t} \middle| Z_i^{(t-1/2)}, \varphi_t, \eta_i, \theta_i\right) = 0, \tag{125}$$

in which the equality to zero is using the flat spot condition (27).

S.3.2 Complement to Song (2013)'s proof

In this Appendix, we establish the invertibility of a high-dimensional matrix that is used to establish the asymptotic properties of coefficient estimators as given by Proposition 1 in Song (2013). Indeed, the initial proof ignores the fact that this matrix has dimensions that tend to infinity as the number of individuals tends to infinity. This can be an issue as this matrix is inverted whereas its eigenvalues may tend to zero. We establish that this is not the case making use of results given by Su and Ju (2018).

To propose a generic proof in a panel data setting, we use T by abuse of notation for the number of observed years in the panel instead of $T - 1985$ as in our specific context in the previous proofs and text. We can rewrite the equation of Song's page 74 (top of the page) as:

$$\underset{[K,1]}{\Delta_i} = \underset{[K,1]}{\xi_i^*} + \frac{1}{N}\sum_{j=1}^{N} a_{ij} S_{ii}^{-1/2} \underset{[K,K]}{S_{ij}} S_{jj}^{-1/2} \Delta_j + o_P(1) \tag{126}$$

in which:

- $S_{ij} = \frac{x_i' M_\varphi x_j}{T}$, where x_j is a $[T, K]$ matrix and M_φ a $[T, T]$ matrix (notation of Song, page 73).

- $\xi_i^* = \frac{1}{\sqrt{T}} S_{ii}^{-1/2} x_i' M_\varphi \varepsilon_i = S_{ii}^{-1/2} \xi_i$ (the latter using notation of Song, page 73). S_{ii} is invertible because of Song's Assumptions B.ii and B.iii uniformly in i (eigenvalue bound)

- the random vector $\Delta_i = \sqrt{T} S_{ii}^{1/2} (\hat{\eta}_i - \eta_i)$ (our notation)

- the scalar, $a_{ij} = \theta_j' (\frac{\Theta\Theta'}{N})^{-1} \theta_i$ in which $\Theta = (\theta_1, ., \theta_n)$ and the matrix $A = [a_{ij}]$ (our notation)

The issue at stake is the invertibility of this linear system of equations (126) with unknowns $\Delta = (\Delta_1, ., \Delta_N)$ that we can write as:

$$\Delta = \xi + \Gamma\Delta, \tag{127}$$

where

$$\Gamma = \text{Block matrix}[\Gamma_{ij}]_{i,j} \tag{128}$$

in which $\Gamma_{ij} = \frac{a_{ij}}{N} S_{ii}^{-1/2} \underset{[K,K]}{S_{ij}} S_{jj}^{-1/2}$. The issue is the invertibility of $I - \Gamma$.

First, approximate $\frac{\Theta\Theta'}{N} = \Sigma_\theta + o_P(1)$. Thus the random variable

$$\Xi_i := \theta_i' (\frac{\Theta\Theta'}{N})^{-1} \theta_i = \theta_i' (\Sigma_\theta)^{-1} \theta_i + o_P(1), \tag{129}$$

is well defined since all eigenvalues of Σ_θ are bounded from below. Set $\theta_i^* = \Sigma_\theta^{-1/2} \theta_i$, and observe that $\mathbb{E}(\theta_i^*) = 0$, $V(\theta_i^*) = I_r$ as well as $a_{ii} = \theta_i^{*\prime} \theta_i^* + o_P(1) = \theta_i' (\Sigma_\theta)^{-1} \theta_i + o_P(1)$. Note that $\mathbb{E}(\Xi_i) = r + o(1)$ and $V(\Xi_i) < \infty$ since by Assumption A2i, we have that $E \|\theta_i\|^4 < \infty$.

Second, we follow the same technique of proof as Su and Ju (2018, page 3 in their Online Appendix) and write $\Gamma = C_1 + C_1' - C_d$ in which:

$$C_1 = N^{-1} \begin{pmatrix} a_{11} I_K & a_{12} S_{11}^{-1/2} S_{12} S_{22}^{-1/2} & \cdots & a_{1N} S_{11}^{-1/2} S_{1N} S_{NN}^{-1/2} \\ 0 & a_{22} I_K & \cdots & a_{2N} S_{22}^{-1/2} S_{2N} S_{NN}^{-1/2} \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & a_{NN} I_K \end{pmatrix} \tag{130}$$

and $C_d = N^{-1} A \otimes I_K$. Denote by $\mu_{\max}(M)$ the maximal eigenvalue of matrix M. Using the fact that eigenvalues of a block upper/lower triangular matrix are the combined eigenvalues of its diagonal block matrices, as well as Weyl's inequality, we get:

$$\mu_{\max}(\Gamma) \leq 2\mu_{\max}(C_1) - \mu_{\min}(C_d) \tag{131}$$

$$\leq 2N^{-1} \max_{1\leq i\leq N}(a_{ii}) = 2N^{-1}(\max_{1\leq i\leq n}(\theta_i^{*\prime}\theta_i^*) + o_P(1)) \tag{132}$$

$$= 2N^{-1} \max_{1\leq i\leq N}(\|\theta_i^*\|^2) + o_P(N^{-1}) = o_P(N^{-3/4}). \tag{133}$$

since $\max_{1\leq i\leq N}(\|\theta_i^*\|^2) = o_P(N^{1/4})$ by the Markov inequality and strengthening Assumption A2.i in Song into Assumption A1.ii of Su and Ju (2018). Therefore $I - \Gamma$ is invertible and equation (126) leads to:

$$\underset{[KN,1]}{\Delta} = (I - \Gamma)^{-1} \underset{[KN,1]}{\xi^*} + o_P(1). \tag{134}$$

S.3.3 The iterative procedure

Our iteration algorithm runs as follows. We use $[k]$ as a superscript for parameters at step k. To obtain initial values at step 0, we follow Moon and Weidner (2019), and first recover regularized estimators of parameters η_i denoted $\eta_i^{[0]}$, by minimizing the nuclear-norm of residuals, a convex program that has a unique solution. In contrast, the least squares minimization program is not convex (e.g. Jiang *et al.*, 2021) because of interactive effects, and may yield several local solutions. Starting with consistent initial conditions $\eta_i^{[0]}$ however makes least squares minimization yield unique consistent estimates and their asymptotic convergence rate is faster (see Hsiao, 2018). We also conduct a principal component analysis of $\ln y_{i,t} - x_{i,t}\eta_i^{[0]}$ (whose value is set to zero when $y_{i,t}$ is not observed), and we get initial factor values $\varphi^{[0]}$ imposing the normalization $\frac{\varphi^{[0]}(\varphi^{[0]})'}{T} = I$.

Updating from step $k-1$ to step k proceeds in three inner steps (M-E-M) as follows:

1. We regress $\ln y_{i,t}$ on $x_{i,t}$ and $\varphi_t^{[k-1]}$ for each individual, considering only years at which they are observed, and we recover the estimators $\eta_i^{[k]}$ and $\theta_i^{[k]}$.[14]

2. We predict the values of $\ln y_{i,t}$ when they are not observed using the formula: $\widehat{\ln y_{i,t}} = x_{i,t}\eta_i^{[k]} + \varphi_t^{[k-1]}\theta_i^{[k]}$.

[14]Note that we retain the estimator of θ_i at this step rather than the one from Bai's procedure at step 3 of previous iteration to avoid using imputed values of $\ln y_{i,t}$ to estimate θ_i. This makes the algorithm converge faster. Note also that even if $\theta_i^{[k]}\theta_i^{[k]\prime}/N$ is not diagonal by construction at each iteration of our algorithm, it becomes diagonal as the algorithm converges since estimated parameters converge to the least square solution as shown in Supplementary Appendix S.3.5.

3. We estimate the factor model: $\ln y_{i,t} - x_{i,t}\eta_i^{[k]} = \varphi_t\theta_i + v_{i,t}$, and recover the estimator $\varphi_t^{[k]}$ using Bai (2009)'s approach.

Supplementary Appendix S.3.4.1 details the adaptation of the EM algorithm to a multiple cohort setting. The stopping rule of the iterative procedure is detailed in Supplementary Appendix S.3.4.2. In Supplementary Appendix S.3.5, we further show that this EM algorithm is valid using Heyde and Morton (1996). It delivers the pseudo-ML estimators of parameters.

S.3.4 Computation details

S.3.4.1 Factor normalization

A slight adaptation of the flat spot restriction (27) is needed because we use different cohorts which are observed over overlapping spells from the first cohort defined by $t_{i,0} = 1985$ up until the last one defined by $t_{i,0} = 1992$, and because we need to normalize factors as is necessary for all interactive effect models. Denote the matrix of calendar time factors as $\varphi^{(T)} = (\varphi'_{1985}, ..., \varphi'_T)$, and introduce the expanded sequence of explanatory variables, all of dimension $T-1984$, $e^{(0)} = (1, ., 1)'$, $e^{(1)} = (1, 2, \cdots, T-1984)'$ and $e^{(2)} = (\beta^{-1}, \cdots, \beta^{-(T-1984)})'$. Stacking $\{\theta_i\}_{i=1,.,N}$ into matrix Θ, we normalize factors and factor loadings as:

$$\frac{\varphi^{(T-1984)}(\varphi^{(T-1984)})'}{T-1984} = I, \;\; \Theta\Theta'/N \text{ is diagonal} \tag{135}$$

and we impose that all elements of φ_1 are positive. We also rewrite the flat spot restriction (27) with respect to factors as:

$$\forall k = 1, 2, 3, \;\; \varphi^{T-1984} e^{(k)} = 0, \tag{136}$$

while there is no such restriction related to factor loadings θ_i that can be freely correlated with the terms η_i.

S.3.4.2 The stopping rule of the iterative procedure

The stopping rule of the iterative procedure that we use is a combination of two rules concerning factors and factor loadings. In the principal components approach, factors can be recovered as the K normalized eigenvectors corresponding to the K largest eigenvalues of matrix $\sum_{i=1}^{N}\left(\ln y_i - x_i\eta_i^{[k]}\right)\left(\ln y_i - x_i\eta_i^{[k]}\right)'$ in which $\ln y_i = (\ln y_{i,1}, ..., \ln y_{i,T})'$ and $x_i = \left(x'_{i,1}, ..., x'_{i,T}\right)'$ so that the estimated space spanned by these eigenvectors converges to the true value. Our first criterium to assess convergence is thus: $C_1 \equiv \left\| M_{\varphi^{[k-1]}}\varphi^{[k]}\right\| / RT$. Second, as it is very demanding to have each factor loading converge, we evaluate convergence through studentized averages and covariance matrices. Formally, our second and third criteria are:

$$C_2 = N\left(\overline{\theta}^{[k]} - \overline{\theta}^{[k-1]}\right)' V\left(\theta_i^{[k-1]}\right)^{-1}\left(\overline{\theta}^{[k]} - \overline{\theta}^{[k-1]}\right) \tag{137}$$

with $\overline{\theta}^{[k-1]} = \sum_{i=1}^{N} \theta_i^{[k-1]}/N$ (the inverse of variance $V\left(\theta_i^{[k-1]}\right)$ being used to give less weight to averages of factor loadings estimated with more uncertainty), and:

$$C_3 = tr\left[\left(V\left(\theta_i^{[k]}\right) - V\left(\theta_i^{[k-1]}\right)\right)\left(V\left(\theta_i^{[k]}\right) - V\left(\theta_i^{[k-1]}\right)\right)'\right] / tr\left[V\left(\theta_i^{[k-1]}\right)\right] \tag{138}$$

using the fact that $tr\left[(A-B)'(A-B)\right]$ is a distance between matrices A and B, and dividing by $tr\left[V\left(\theta_i^{[k-1]}\right)\right]$ as a normalization. Our overall stopping rule requires to have $C_1 < 1.e-9$, $C_2 < 1.e-8$ and $C_3 < 1.e-4$.

S.3.5 Convergence of the iterative estimation procedure

We use a specific iterative procedure to find the solution of the sum-of-squares minimization program. We show in this section that our iterative procedure converges to the solution of this program as the number of iterations tends to infinity. Doing so, we follow Heyde and Morton (1996) (see also Dominitz and Sherman, 2005, for a general framework).

The sum of squares we consider is given by:

$$C(\theta, \varphi, \eta) = \sum_{i,t|s_{i,t}=1} \left(\ln y_{i,t} - x_{i,t}\eta_i - \varphi_t\theta_i\right)^2 \tag{139}$$

For a given set of parameters, say for instance η_i, we denote by $\eta_i^{[k]}$ the value of the estimates at the k^{th} iteration.

As explained in the text, the first stage of our algorithm consists in minimizing $C\left(\theta, \varphi^{[k-1]}, \eta\right)$ with respect to θ and η – maintaining $\varphi^{[k-1]}$ constant. We denote the values of the arguments of the minimizer as $\eta^{[k]} = (\eta_i^{[k]})_{i=1,.,n}$ and $\theta^{[k]} = (\theta_i^{[k]})_{i=1,.,n}$.

At the second stage, we impute wages that are not observed using the formula:[15]

$$\ln y_{i,t}^{[k]} = x_{i,t}\eta_i^{[k]} + \varphi_t^{[k-1]}\theta_i^{[k]} \tag{140}$$

[15] A few workers are more than 50 years old and according to the flat-spot approach we assume that they no longer accumulate human capital. We also replace their wages by their linear prediction after 50 as a mere statistical device to balance the panel.

At the third stage, we recover values of θ and φ – fixing the values of $y_{i,t}^{[k]}$ and $\eta_i^{[k]}$ – that minimize the sum of squares:

$$\widetilde{C}\left(\theta,\varphi,\eta^{[k]}\right) = C\left(\theta,\varphi,\eta^{[k]}\right) + \sum_{i,t|s_{i,t}=n} \left(\ln y_{i,t}^{[k]} - x_{i,t}\eta_i^{[k]} - \varphi_t\theta_i\right)^2 \quad (141)$$

using Bai's algorithm and we denote these values, $\widetilde{\theta}^{[k]}$ and $\varphi^{[k]}$.

We now show that the sum of squares decreases at each iteration of our algorithm.

Lemma 9

$$C\left(\widetilde{\theta}^{[k]},\varphi^{[k]},\eta^{[k]}\right) \leqslant C\left(\widetilde{\theta}^{[k-1]},\varphi^{[k-1]},\eta^{[k-1]}\right). \quad (142)$$

Proof. From the first stage of our algorithm, we have that:

$$C\left(\theta^{[k]},\varphi^{[k-1]},\eta^{[k]}\right) \leqslant C\left(\widetilde{\theta}^{[k-1]},\varphi^{[k-1]},\eta^{[k-1]}\right), \quad (143)$$

since $\theta^{[k]},\eta^{[k]}$ are minimizers of the left-hand side. Using the definition of $y_{i,t}^{[k]}$, we also have that

$$\widetilde{C}\left(\theta^{[k]},\varphi^{[k-1]},\eta^{[k]}\right) = C\left(\theta^{[k]},\varphi^{[k-1]},\eta^{[k]}\right), \quad (144)$$

since the sum of squares on the right hand side of equation (141) is equal to zero. The third stage of our algorithm yields, by minimization:

$$\widetilde{C}\left(\widetilde{\theta}^{[k]},\varphi^{[k]},\eta^{[k]}\right) \leqslant \widetilde{C}\left(\theta^{[k]},\varphi^{[k-1]},\eta^{[k]}\right). \quad (145)$$

and we get, using equations (145), (144) and (143) successively:

$$\begin{aligned}
C\left(\widetilde{\theta}^{[k]},\varphi^{[k]},\eta^{[k]}\right) &\leqslant \widetilde{C}\left(\widetilde{\theta}^{[k]},\varphi^{[k]},\eta^{[k]}\right) && (146)\\
&\leqslant \widetilde{C}\left(\theta^{[k]},\varphi^{[k-1]},\eta^{[k]}\right) = C\left(\theta^{[k]},\varphi^{[k-1]},\eta^{[k]}\right) && (147)\\
&\leqslant C\left(\widetilde{\theta}^{[k-1]},\varphi^{[k-1]},\eta^{[k-1]}\right). && (148)
\end{aligned}$$

■

This shows that the sum of squares is decreasing at each iteration. In fact, it is strictly decreasing as a consequence of the following lemma:

Lemma 10

$$C\left(\widetilde{\theta}^{[k]},\varphi^{[k]},\eta^{[k]}\right) = C\left(\widetilde{\theta}^{[k-1]},\varphi^{[k-1]},\eta^{[k-1]}\right) \Longrightarrow \left(\widetilde{\theta}^{[k]},\varphi^{[k]},\eta^{[k]}\right) = \left(\widetilde{\theta}^{[k-1]},\varphi^{[k-1]},\eta^{[k-1]}\right). \quad (149)$$

Proof. The left-hand side equality implies that:

$$C\left(\widetilde{\theta}^{[k]},\varphi^{[k]},\eta^{[k]}\right)=\widetilde{C}\left(\widetilde{\theta}^{[k]},\varphi^{[k]},\eta^{[k]}\right) \tag{150}$$

according to equation (148). Using (141), this yields:

$$\sum_{i,t|s_{i,t}=n}\left(\varphi_t^{[k-1]}\theta_i^{[k]}-\varphi_t^{[k]}\widetilde{\theta}_i^{[k]}\right)^2=0 \tag{151}$$

and thus $\varphi_t^{[k-1]}\theta_i^{[k]}=\varphi_t^{[k]}\widetilde{\theta}_i^{[k]}$ for all i,t such that $s(i,t)=0$. Considering also that there are identification restrictions on parameters, we then have generically $\varphi_t^{[k-1]}=\varphi_t^{[k]}$ and $\widetilde{\theta}_i^{[k]}=\theta_i^{[k]}$ for all i, t. From equation (148), we also have that:

$$C\left(\theta^{[k]},\varphi^{[k-1]},\eta^{[k]}\right)=C\left(\widetilde{\theta}^{[k-1]},\varphi^{[k-1]},\eta^{[k-1]}\right). \tag{152}$$

As C is strictly concave, the solution in the first step is unique for a given $\varphi^{[k-1]}$, and we get that $\theta^{[k]}=\widetilde{\theta}^{[k-1]}$ and $\eta^{[k]}=\eta^{[k-1]}$. Putting all the equalities on parameters together, we obtain $\left(\widetilde{\theta}^{[k]},\varphi^{[k]},\eta^{[k]}\right)=\left(\widetilde{\theta}^{[k-1]},\varphi^{[k-1]},\eta^{[k-1]}\right)$. ■

Using the contraposition of the Lemma and equation (142), we have that:

$$\left(\widetilde{\theta}^{[k]},\varphi^{[k]},\eta^{[k]}\right)\neq\left(\widetilde{\theta}^{[k-1]},\varphi^{[k-1]},\eta^{[k-1]}\right)\Longrightarrow C\left(\widetilde{\theta}^{[k]},\varphi^{[k]},\eta^{[k]}\right)<C\left(\widetilde{\theta}^{[k-1]},\varphi^{[k-1]},\eta^{[k-1]}\right), \tag{153}$$

which shows that the sum of squares is strictly decreasing at each iteration. As it is bounded below by zero, it converges to a value $\overline{C}$ and parameters converge to the value of its minimizers $\left(\widehat{\widetilde{\theta}},\widehat{\varphi},\widehat{\eta}\right)$. As $\theta^{[k]}$ minimizes $C\left(\theta,\varphi^{[k-1]},\eta^{[k]}\right)$, and $\varphi^{[k-1]}$ and $\eta^{[k]}$ converge respectively to $\widehat{\varphi}$ and $\widehat{\eta}$, $\theta^{[k]}$ converges to the value of θ denoted $\widehat{\theta}$ that minimizes $C\left(\theta,\widehat{\varphi},\widehat{\eta}\right)$. We also have that $\widetilde{\theta}^{[k]}$ is the value that minimizes:

$$\widetilde{C}\left(\theta,\widehat{\varphi},\widehat{\eta}\right)=C\left(\theta,\widehat{\varphi},\widehat{\eta}\right)+\sum_{i,t|s_{it}=n}\left(\widehat{\varphi}\left(\widehat{\theta}_i-\theta_i\right)\right)^2 \tag{154}$$

As $C\left(\theta,\widehat{\varphi},\widehat{\eta}\right)$ is minimum in $\widehat{\theta}$, and the second right-hand side term is positive but zero for $\theta=\widehat{\theta}$, then $\widetilde{C}\left(\theta,\widehat{\varphi},\widehat{\eta}\right)$ is minimized at $\widehat{\theta}$ and we have $\widehat{\widetilde{\theta}}=\widehat{\theta}$. Overall, step 1 yields that $\widehat{\theta}$ and $\widehat{\eta}$ verify the least squares first-order conditions, and step 3 makes $\widehat{\varphi}$ verify the least squares first-order conditions. Hence, $\left(\widehat{\theta},\widehat{\eta},\widehat{\varphi}\right)$ is the least squares solution.

S.3.6 Bias correction and small sample issues for counterfactuals

The asymptotic properties of consistency and asymptotic normality of our estimates are obtained in a balanced panel data setting such that N and T tend to infinity. Proofs of Bai (2009) are extended by Song (2013) to the case of individual specific coefficients of covariates. Note also that individual observations are incomplete because of non participation, and we need to assume that T_i/T tends to an individual specific positive constant where T_i is the number of observed years for every individual i.

Using estimated individual specific parameters, we compute structural functions of potential outcomes as defined in Section 6. Their empirical counterparts generically suffer from the incidental parameter issue, which causes variances and other summary statistics like quantiles to be asymptotically biased (Fernández-Val and Weidner, 2018). As detailed in the Online Appendix E of Gobillon *et al.* (2022), these biases can be corrected. This is the case for biases of variances and covariances when the covariance matrix of idiosyncratic errors is restricted as shown by Arellano and Bonhomme (2012). For quantiles and interquantile ranges, we resort to the bias-correction procedure based on Taylor expansions proposed by Jochmans and Weidner (2024).

Bias corrections rely on asymptotic formulas established when the number of individuals and the number of years during which they are employed in the private sector tend to infinity. Some individuals are employed during 15 years only whereas the model involves up to 7 individual parameters capturing the individual unobserved heterogeneity in our preferred specification. Finite sample properties of estimators are thus not granted and need to be investigated. For that purpose, we conducted Monte-Carlo simulations whose results are presented in detail in Online Appendix F of Gobillon *et al.* (2022). As expected, these simulations show that the means of individual parameters and of structural functions are barely biased. Estimated variances are strongly biased however, and the bias-correction procedure removes part of the bias only. By contrast, estimated quantiles are characterized by smaller biases and those can be corrected satisfactorily using Jochmans and Weidner (2024).

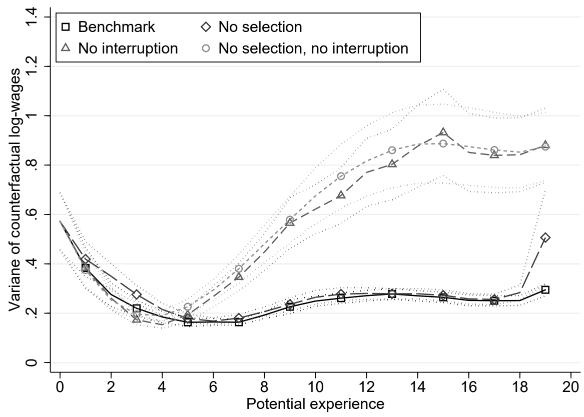
This is why we focus mostly on means, deciles and inter-decile ranges in our empirical application. Monte Carlo results also show that, when disturbances $v_{i,t}$ are heteroskedastic rather than homoskedastic, bias-correction for centiles and inter-centile differences are farther away from their true values. This is because individual variances of residuals are poorly estimated. We assume that disturbances are homoskedastic when computing estimated standard errors.

Table S.2: Minimization criteria used to select the number of factors

Number of factors	0	1	2	3
Criteria				
PC_{p1}	.014071	.012745	.012015	.011875
PC_{p2}	.014071	.012652	.011829	.011596
PC_{p3}	.014071	.012939	.012404	.012460
IC_{p1}	-4.263	-4.307	-4.335	-4.330
IC_{p2}	-4.263	-4.319	-4.359	-4.366
IC_{p3}	-4.263	-4.282	-4.285	-4.254
Quantities used to compute criteria				
N	7339	7339	7339	7339
$T - DF$	15.575	15.575	15.575	15.575
$V(k.\widehat{\varphi}_t)$	.0140	.0115	.0096	.0083
$\widehat{\sigma}^2$	.0110	.0091	.0076	.0065
$C^2_{N(T-DF)}$	15.575	15.575	15.575	15.575

Note: This table reports the values of six minimization criteria introduced by Bai and Ng (2002) to determine the number of factors, PC_{pj} and IC_{pj}, with $j \in \{1, 2, 3\}$. We also report quantities that are used to construct these criteria. N is the number of individuals in our sample and T is the average number of periods per individual. We correct for the average number of degrees of freedom: We consider $T - DF$ instead of T, were DF is the average number of individual-specific coefficients for the explanatory variables introduced in our specification. $\widehat{\sigma}^2$ is the estimated variance of residuals. Other quantities $V(k.\widehat{\varphi}_t)$, with k the number of factors and $C^2_{N(T-DF)}$ are given in Bai and Ng (2002, p. 201).

Figure S.2: Corrected variances of counterfactual adjusted log-wage as a function of potential experience, counterfactual scenarii 1-4



Note: Adjusted log-wages are computed from raw wages by (i) changing the timing of cohorts such that all cohorts enter the labour market in 1984; (ii) deflating wages with skill-specific prices of human capital; (iii) adding time-constant skill-specific prices of human capital. "Corrected" statistics are obtained after bias correction as described in Supplementary Appendix S.3.6.

Table S.3: Descriptive statistics on counterfactual adjusted log-wages, when neutralizing static and dynamic selections, different numbers of days full-time in the private sector for entry

	No selection, no interruption				No selection			
	1	5	10	20	1	5	10	20
Number of days=90								
Mean	3.174	3.387	3.569	3.857	3.174	3.360	3.510	3.766
Inter-decile	1.766	1.035	1.573	1.892	1.766	1.048	1.192	1.349
Number of days=180, baseline								
Mean	3.214	3.401	3.558	3.841	3.214	3.390	3.522	3.770
Inter-decile	1.298	0.937	1.272	1.635	1.298	0.951	1.089	1.274
Number of days=210								
Mean	3.226	3.408	3.561	3.843	3.226	3.395	3.526	3.779
Inter-decile	1.296	0.936	1.308	1.692	1.296	0.955	1.095	1.278
Number of days=240								
Mean	3.232	3.412	3.563	3.839	3.232	3.401	3.532	3.774
Inter-decile	1.265	0.934	1.258	1.615	1.265	0.922	1.075	1.252
Number of days=270								
Mean	3.225	3.415	3.575	3.849	3.225	3.402	3.541	3.774
Inter-decile	1.293	0.935	1.282	1.631	1.293	0.915	1.103	1.264
Number of days=360								
Mean	3.263	3.441	3.587	3.838	3.263	3.430	3.557	3.775
Inter-decile	1.019	0.897	1.134	1.431	1.019	0.890	1.005	1.180

Note: "No selection, no interruption": Log-wages are computed as $\eta_{i0} + \eta_{i1}d + \eta_{i2}\beta^{-d}$ where d is the potential experience. "No selection": Log-wages are computed as $\eta_{i0} + \eta_{i1}d + \eta_{i2}\beta^{-d} + \eta_{i3}x^{(3)}_{i,t_{i,0}+d} + \eta_{i4}x^{(4)}_{i,t_{i,0}+d}$. In both cases, log-wages are adjusted by (i) changing the timing of cohorts such that all cohorts enter the labour market in 1984; (ii) deflating wages with skill-specific prices of human capital; (iii) adding time-constant skill-specific prices of human capital. Wage statistics are computed on the whole working sample of individuals at each value of potential experience (whether they are employed or not). "Corrected" statistics are obtained after bias correction as described in Supplementary Appendix S.3.6.

Table S.4: Descriptive statistics on distributions of counterfactual adjusted log-wages, when neutralizing static and dynamic selections, different β values

	No selection, no interruption				No selection			
	1	5	10	20	1	5	10	20
$\beta = .93$								
Mean	3.221	3.399	3.553	3.843	3.221	3.390	3.519	3.774
Inter-decile	1.316	0.951	1.276	1.696	1.316	0.966	1.098	1.297
$\beta = .94$								
Mean	3.218	3.400	3.555	3.842	3.218	3.390	3.521	3.772
Inter-decile	1.315	0.941	1.289	1.653	1.315	0.944	1.097	1.288
$\beta = .95$, baseline								
Mean	3.214	3.401	3.558	3.841	3.214	3.390	3.522	3.770
Inter-decile	1.298	0.937	1.272	1.635	1.298	0.951	1.089	1.274
$\beta = .96$								
Mean	3.211	3.402	3.561	3.840	3.211	3.390	3.524	3.769
Inter-decile	1.267	0.921	1.269	1.636	1.267	0.939	1.078	1.260
$\beta = .97$								
Mean	3.207	3.403	3.564	3.839	3.207	3.390	3.525	3.767
Inter-decile	1.230	0.947	1.285	1.591	1.230	0.935	1.073	1.242
$\beta = .98$								
Mean	3.204	3.404	3.566	3.839	3.204	3.390	3.527	3.765
Inter-decile	1.209	0.936	1.284	1.572	1.209	0.927	1.065	1.224
$\beta = .99$								
Mean	3.200	3.406	3.569	3.839	3.200	3.390	3.528	3.764
Inter-decile	1.190	0.944	1.294	1.569	1.190	0.917	1.046	1.207

Note: "No selection, no interruption": Log-wages are computed as $\eta_{i0} + \eta_{i1}d + \eta_{i2}\beta^{-d}$ where d is the potential experience. "No selection": Log-wages are computed as $\eta_{i0} + \eta_{i1}d + \eta_{i2}\beta^{-d} + \eta_{i3}x^{(3)}_{i,t_{i,0}+d} + \eta_{i4}x^{(4)}_{i,t_{i,0}+d}$. In both cases, log-wages are adjusted by (i) changing the timing of cohorts such that all cohorts enter the labour market in 1984; (ii) deflating wages with skill-specific prices of human capital; (iii) adding time-constant skill-specific prices of human capital. Wage statistics are computed on the whole working sample of individuals at each value of potential experience (whether they are employed or not). "Corrected" statistics are obtained after bias correction as described in Supplementary Appendix S.3.6.