WORKING PAPERS

N° 1299

February 2022

"Lifecycle Wages and Human Capital Investments: Selection and Missing Data"

Laurent Gobillon, Thierry Magnac and Sébastien Roux

Toulouse School of Economics

Lifecycle Wages and Human Capital Investments: Selection and Missing Data

Laurent Gobillon[*], Thierry Magnac[†], Sébastien Roux[‡]

First version: June 2018
This version: February 2, 2022[§]

Abstract

We derive wage equations with individual specific coefficients from a structural model of human capital investments over the life-cycle. This model allows for interruptions in labor market participation, and addresses missing data and attrition issues. We further control for selection in a flexible way by using interactive effects. Estimation is based on long administrative panel data of male wages in the private sector in France. A structural function approach shows that interruptions negatively affect average wages. More surprisingly, they also negatively affect the inter-decile range of wages after twenty years, and this is due to interruptions being endogeneous. These results question the popular Missing At Random assumption that is made when assessing the building up of wage inequalities over the life cycle.

JEL Codes: C38, D91, I24, J24, J31

Keywords: Human capital investment, wage inequalities, factor models, missing data

[*]Paris School of Economics-CNRS, 48 Boulevard Jourdan, 75014 Paris, email: laurent.gobillon@psemail.eu, webpage: http://laurent.gobillon.free.fr/.

[†]Toulouse School of Economics, Université de Toulouse Capitole, Toulouse, France

[‡]INSEE.

[§]We are grateful to Stephane Bonhomme, Xavier d'Haultfeuille and Koen Jochmans as well as participants of seminars and conferences at Sciences Po, French Econometrics'18, Luxembourg, IAAE'19, EALE'19, Western Ontario, Marseilles, Venice, Tilburg and Carnegie Mellon for their helpful comments. We acknowledge funding from the Institut Universitaire de France and ANR under grants ANR-17-EURE-0001 and ANR-17-EURE-0010 (Investissements d'Avenir programs). All errors are ours.

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 Blundell, 2014 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, Ejrnaes and Alvarez (2012), Polachek, Das and Thamma-Apiroam (2015) or Magnac and Roux (2021). These authors individualize as much as possible earnings processes by estimating sets of individual specific parameters beyond the permanent effects that appear additively in equations commonly estimated with panel data of earnings (Heckman, Lochner and Todd, 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 is to select wage histories which are sufficiently long and to treat missing observations in histories as random using so called missing at random (MAR) procedures. The missing data issue is particularly important when parameters are individual specific since their estimation uses individual time-series and consistency of those estimates relies on the number of periods being large.

In this paper, we propose a general framework to study the evolution of wage inequalities over the life-cycle which accomodates incomplete individual wage trajectories. Our approach builds upon the structural linear model proposed by Magnac, Pistolesi and Roux (2018) for the logarithm of wages over the life-cycle as a function of four individual specific parameters: the initial level of human capital at entry in the labour market, the returns to human capital, their costs and the terminal value of human capital stocks.

Our first contribution is to extend the model to the case of two sectors in which 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 costs of human capital investments across sectors creates a wedge between the accumulation processes in human capital in the two sectors (see for instance Blundell, Costa-Dias, Meghir and Shaw, 2016, for part-time/full-time evidence). In particular, we expect that interruptions in the career have a sizeable effect on human capital investments (Light and Ureta, 1995). This structural model makes us introduce 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 different impact of potential and actual experience on wages in a homogeneous set-up (Eckstein and Wolpin, 1988, Altug and Miller, 1998).

Our second contribution is an original empirical strategy that deals with selection issues under an assumption much weaker than MAR. We posit a factor structure for the residual structural process of human capital prices and of 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. Such an approach with interactive effects is akin to the one proposed by Aakvik, Heckman and Vytlacil (2005), and squares well with the fact that lots of heterogeneity affect wage histories over the life-cycle. Econometric moment restrictions are further vindicated by a "flat spot" approach introduced by Heckman et al. (1999) which allows 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 followed until 2012 – if they do not leave the panel before. The other sector gathers all other employment and non-employment alternatives. French data provide an attractive observational case because wage inequalities in the population have been quite stable in the past 40 years. This stability of wage inequalities contrasts with the US and the German experiences (Antonczy, DeLeire and Fitzenberger, 2018), and is likely due to policy interventions such as steady increases in the minimum wage, and payroll tax exemptions for the less skilled (Bozio, Breda and Guillot, 2020).

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. We estimate various specifications using Bai (2009)'s least-square method adapted to the presence of missing data, and extended by Song (2013) to the case of individual specific coefficients. We also use as the starting point to the Bai algorithm a consistent estimator proposed by Moon and Weidner (2018). As derived by model selection procedures, our preferred specification includes two factors.

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. Fernandez-Val and Weidner, 2018). We correct biases using methods proposed by Jochmans and Weidner (2019), 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 quantiles, and particularly inter-decile ranges which are more robust.

Results based on our original empirical strategy constitute our third contribution. We first show that omitting interruptions and unobserved factors strongly downward biases returns to experience after 20 years. Second, most of this bias comes from the influence of interruptions on human capital accumulation. In other words, selectivity seems mainly captured by interruptions and not by the additional interactive effects. Third, we estimate average structural functions (Blundell and Powell, 2003) that are constructed by manipulating interruptions. Accordingly, we estimate the causal impact of the existence and timing of interruptions. Provided that identification conditions akin to the ones developed in Chernozhukov et al. (2013) are satisfied, we show that interruptions have a significant negative effect on average wages. More surprisingly they also have a negative effect on the dispersion of wages after 20 years. This is mainly due to an endogeneity issue: individual-specific parameters induce negative correlations in the population between the effects of interruptions and wages reconstructed by setting interruptions to zero.

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 the identifying restrictions of the econometric model. Section 6 presents our estimation strategy and results are reported in Section 7. Section

8 discusses remaining selection issues.

2 Literature review

Earning dynamics We first discuss the very extensive empirical literature on earnings dynamics (see Meghir and Pistaferri, 2010, or Blundell, 2014, for a review). 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) study a large set of countries and shows that experience-wage profiles are twice as steep in rich countries as in poor countries. This literature has mostly remained in a linear framework but there has been a few non-linear alternatives (Browning et al., 2012, Guvenen et al., 2020, Bonhomme and Robin, 2009 or Arellano, Blundell and Bonhomme, 2017).

In a different vein, there is a more economically oriented literature trying to distinguish theories of wage growth, namely human capital, job search or learning by doing. Rubinstein and Weiss (2006) surveys the literature before 2005. There are a few recent papers pursuing this research objective such as Bagger, Fontaine, Postel-Vinay and Robin (2014) or Sorensen and Vejlin (2014) and they are reviewed in Magnac and Roux (2021). As we use a human capital model as a maintained assumption, our analysis is not strictly comparable to theirs. We rather follow Polachek et al. (2015) and Magnac et al. (2018) which set up Ben Porath (1967) human capital model of earnings or wages over the life-cycle in different guises. In their specifications, individual specific parameters governing wage equations have a structural economic interpretation, and they can be related to individual characteristics. These parameters are related to the abilities to learn and to earn of individuals (Browning, Heckman and Hansen, 1999, Rubinstein and Weiss, 2006).

Our paper also studies the impact of interruptions in participation on wages. The issue of actual versus potential experience was dealt with as early as the 70s (see for instance, Polachek, 1975) and revisited in the 90s. More specifically, Light and Ureta (1995) use rich reported information on breaks, and show that these additional variables have explanatory power over and above the quadratic term in experience. Interestingly, the timing of interruptions matters empirically. In our case, we use an admittedly restrictive structural model although it is much richer in terms of individual specific heterogeneity. We also find that interruptions negatively affect average wages and that their timing has a significant impact. We only know of one paper which reports results on the effect of interruptions on wage inequalities (Biewen et al., 2018). They analyze German data between 1985 and 2010 with a focus on the evolution of wage inequalities for the whole population over time, and not for specific cohorts as we do. In a decomposition exercise of exogenous covariate effects, the authors show that inequalities increased with the number of interruptions – which is trending upward over time. This is in sharp contrast with what we obtained with an admittedly different focus on life cycle inequalities within cohorts, using models with lots of heterogeneity, and controlling for endogeneity issues. Indeed, we show that the negative effect on inequalities we obtain stems from the endogenous nature of interruptions. Finally, our paper does not analyze female wages although a huge literature assesses to what extent the gender wage gap can be explained by interruption patterns (see Das and Polachek, 2019 for a survey).

Missing data The pattern of missingness considered in our paper is due to non-observed outcomes (see Bollinger et al., 2019, for an example). In particular, time-series of wages are irregularly observed over time because of interruptions in private sector participation. The missing at random (MAR) assumption, conditional or not on exogenous covariates, is well explored in statistics (see Little and Rubin, 2019). Under this assumption, the focus of GMM literature is the efficiency of estimation (Abowd, Crépon and Kramarz, 2001), or the robust and efficient estimation (Graham, Pinto and Egel, 2012, and Chaudhury and Guilkey, 2016). Attrition using other types of MAR assumptions are also explored by Sasaki (2015) who imposes ingenious non standard exclusion restrictions.

Any Missing At Random (MAR) procedure might induce sizeable biases in wage equation estimates if the degree of attachment to the private sector, as measured by the reciprocal of the number of interruptions in individuals' careers, is somehow associated to individual parameters such as the ones which describe abilities, returns to human capital investments or their costs. Correcting for missing data issues is however difficult in the absence of exogenous variation which would affect entry and exit from the panel without affecting wages i.e. the absence of credible alternative exclusion restrictions. It is difficult to entertain the idea that such instrumental variables can be derived from our administrative data.

Instead of relying on a usual MAR assumption, we impose structure derived from an economic model and a Missing At Random Conditional On Factors (MARCOF) assumption. Our restrictions can then be interpreted in an economic way. The generalization of selection-onunobservable factors was also explored in a difference-in-differences setting by Gobillon and Magnac (2016). Another direction away from MAR taken up recently is the literature on sensitivity (e.g. Kline and Santos, 2013). An intermediate "breakdown" solution between MAR and worst case bounds a la Manski is sought so that substantial results remain (just) significant. Since wages are our outcome of interest, the worst case bounds, in our case, are infinite.

Factor models The development of factor models for panel data started with Holtz-Eakin, Newey and Rosen (1989) and Ahn, Lee and Schmidt (2001). Pesaran (2006) proposes a restrictive framework in which regressors are low rank, i.e. they are equal to the bilinear product of individual specific effects and time varying factors. This framework is not adapted to our setting since explanatory variables related to interruptions are high rank regressors (see Moon and Weidner, 2017). Instead, 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 coefficients. We first 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 (2018) and Beyhum and Gautier (2019) 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, Wainwright and Hu (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.

All individuals born in the first four days of October of an even year are followed over time, and their jobs in the private sector are recorded between 1976 and 2011 except in 1990. 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 in every year. Education as measured by diploma is recovered from *EDP* and the 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 denoted 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 $137,315$ yearly observations involving $7,004$ males. Further details on the sample construction are given in Appendix A.1.

3.2 Descriptive statistics

In Figure 1, we report the profile of statistical summaries of the logarithm of wages as a function of potential experience by education level. We truncate profiles at 20 years of potential experience since the youngest cohort enters in 1992 and the panel ends in 2011.

Figure 1(a) reports the profiles of average log wages for every education group. As expected, they are increasing and concave in potential experience although the profiles are almost linear when experience is large. The slope is steeper for higher education levels, a common finding in most countries (Lagakos et al., 2018).

Mean log wage profiles reported above are the composition of the profile of (log) human capital stocks and (log) human capital prices. We use a “flat spot” approach whose conditions of validity are developed in Heckman, Lochner and Taber (1998) and Bowlus and Robinson (2012) to estimate human capital prices. Its rationale is the following. Because individuals aged

$50 - 55$ have presumably stopped investing in human capital, their median wages by year and education level are consistent estimates of human capital prices conditionally on education level and year if human capital is homogenous within education groups.

Figure 1(b) displays profiles of human capital prices over calendar years. It shows that prices for high school dropouts increased between 1985 and 2011 in France by roughly 80%, mainly because minimum wages have been increasing faster than average wages since at least the 1970s (Cette et al., 2012). For other groups, increases sum up to about 45-50% over the whole period except for high school graduates (about 30%). Albouy and Tavan (2007) document that the supply of these groups increased at the beginning of the 1990s, which might have attenuated their wage increases. Second, policies in France decoupling wages and the costs of labor such as subsidies to lower skill employment help understand the contrasts between education groups that differ greatly in the US (Bozio, Breda and Guillot, 2020).

Net log wages are defined as log wages from which we subtract the logarithm of the yearly price of human capital at each education level, and correspond to individual specific human capital stocks over the life-cycle. These net log wages will be the main outcome variable in our analysis. As shown in Figure 1(c), their averages have profiles different from those obtained for raw wages. The growth of net wages is more pronounced for college graduates and less pronounced for high school dropouts while they are roughly the same for workers with some college education and high school graduates.

To analyze wage dispersion, we display the profile of inter-decile ranges of log wages in Figure 1(d). Interestingly, both levels and slopes differ in a sizable way across education groups. College graduates are characterized by a larger and steeper inter-decile range whereas ranges for high-school graduates and workers with some college education are very close. There is also evidence of a Mincer dip for college graduates (Mincer, 1974) while this dip for other groups is probably masked by the aggregation of different cohorts (see Magnac and Roux, 2021 for a single cohort analysis). We will return below to the significance of Mincer dips when analyzing our complete specification estimates.

Serial correlations between log-human capital stocks at different periods are reported in Table A.1 from which we derive important stylized facts.[1] The one-year lag correlation starts at .83 and grows until 0.94 at the end of the observation period. Human capital stocks are getting more and more persistent when potential experience increases, and this indicates that the variance of idiosyncratic shocks on log-stocks tends to decrease over time (Magnac and Roux, 2021). Non

[1] All Tables and Figures which number is preceded by the A letter are relegated in Appendix.

stationarity is an important element to model in wage dynamics and this will be captured by factors in our empirical analysis. Furthermore, the correlation decreases at longer lags although much less than at a geometric rate. At a 20-year lag, the correlation is equal to 0.28, which is above $(.83)^{20} = 0.024$. This likely denotes the presence of unobserved permanent heterogeneity which will be captured by the factor loadings of observed and unobserved factors in our empirical analysis.

Interruptions in individual participation in the private sector make real and potential experience different and play an important role in our empirical results. In Table 1, we describe these interruptions, including among those, attrition periods. For instance, for an individual exiting the private sector in 2007 and absent until the end of the panel in 2011, we treat as an interruption the periods between 2007 and 2011. This Table shows that the cumulative duration outside the private sector is 3.7 years and the average length of private sector spells is about 21 years. The number of interruptions is equal to 1.44 while the distribution of these interruptions is quite disperse. 523 males have more than 4 interruptions whose cumulative duration reaches about 8 years.

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.

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. For simplicity, we do not index variables or parameters with an individual subscript i in this section although most of them are supposed to be individual specific.

4.1.1 Human capital accumulation in two sectors

Entering the private sector at an initial date, t_0, an individual participates, at each period t, in a labour market sector chosen among two sectors, denoted $s_t \in \{e, n\}$ for either the private or the alternative sector. 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 classified in the alternative sector. Individual wages in the private sector if $s_t = e$, or a wage-equivalent notion if $s_t = n$, are written as:

$$y_t^s = \exp(\delta_t^s) H_t \exp(-\tau), \tag{1}$$

in which H_t is the stock of human capital at the beginning of period t, δ_t^s is the rental rate or "price" of human capital in sector s at time t, and τ is a decision variable such that the term $1 - \exp(-\tau)$ can be interpreted as the fraction of non-leisure time, or alternatively the intensity of effort, devoted to investing in human capital. This fraction is increasing in τ, equal to zero when $\tau = 0$ and equal to one when $\tau = +\infty$. This is why we call, $\tau_t^s \geq 0$, the individual specific investment in human capital at time t in sector s.

The technology of production of human capital in sector s is described by

$$H_{t+1}^s = H_t \exp[\rho^s \tau_t^s - \lambda_t^s], \tag{2}$$

in which ρ^s is the rate of return of human capital investments in sector s (fixed over time but individual specific) and λ_t^s is the depreciation of human capital in sector s at period t. Depreciation λ_t^s embeds individual specific and aggregate shocks that depreciate previous vintages of human capital. Shocks are state-specific if human capital depreciation is larger when in the alternative sector than when in the private sector as on-the-job learning is more likely. Furthermore, the individual rate of learning, ρ^s, differs across sectors. The individual prices of human capital, δ_t^s, and depreciations, λ_t^s, are treated as stochastic processes whose properties are presented below.

We assume that investing in human capital is the only way of smoothing consumption over time. Magnac et al. (2018) derive predictions when relaxing this assumption which requires consumption information that is not available in our data. Without consumption smoothing, period-t utility in sector s is a function of income, effort and participation:

$$\ln y_t^s - c\frac{(\tau_t^s)^2}{2} + \omega_t \mathbf{1}\{s = e\}, \tag{3}$$

in which the variable ω_t is the difference in utility between sectors e and n. Furthermore, the cost of investment is quadratic and indexed by an individual specific parameter, c, that we assume independent of sector s as it is a parameter of the utility function. Moreover, we neglect the linear component of the cost in terms of τ_t^s because it cannot be identified, as log wages in sector s are:

$$\ln y_t^s = \delta_t^s + \ln H_t - \tau_t^s, \tag{4}$$

and the unit in which τ_t^s is expressed, is not identified. Increasing marginal costs fits well with the interpretation of τ_t^s as an exerted effort which decreases current earnings and provides future returns. This is what makes unique the solution in the dynamic programming decision problem.

4.1.2 Timing and value functions

The timing of revelation of shocks, state variables and decisions about sectors and human capital investments is plotted in Figure 2. Our key assumption is that the revelation of sector preference shocks, ω_t, and the choice of sector, s_t, are made before shocks on prices and depreciations of human capital are revealed, and decisions about human capital investments are made. This is a specific version of the Roy model which is known, under conditions developed below, to lead to the absence of selectivity of sector choice on earnings (Heckman and Robb, 1985). In the current paper, this absence of selectivity results from the conditioning on factors and factor loadings, which are unobserved by the econometrician and act as controls for selectivity.

The first row in this figure reports the timing of the revelation of shocks on sector preferences, ω_t, and on price and depreciation of human capital, δ_t^s and λ_t^s. The second row reports the history of the time processes, δ^s, λ^s and ω up to the times described by the first row. In particular Z_t contains the history of ω up to period t and the history of δ^s, λ^s up to period $t-1$. History $Z_{t+1/2}$ completes Z_t with period t information on δ_t^s and λ_t^s. The third line reports the timing of decisions: the choice of sector is made after sector preference shocks are revealed and human capital investments after the revelation of shocks on prices and depreciation. The state variable H_t is inherited from the past according to equation (2) at the very beginning of period t. Below the timeline, the potential wage, y_t, is a function of shocks on prices and depreciation.

Value functions at each stage of this timeline can now be constructed. If V_{t+1} is the value function at the beginning of period $t+1$, its arguments are the state variables, H_{t+1} and Z_{t+1}. At the previous interim stage $t+1/2$, these state variables are H_t, $Z_{t+1/2}$. At time t, because of equations (3) and (4), human capital investments are derived for each sector decision $s \in \{n, e\}$ from the following decision program:

$$\begin{aligned} W_t^s(H_t, Z_{t+1/2}) &= \max_{\tau} \left\{ \delta_t^s + \ln H_t - \left(\tau + c\frac{(\tau)^2}{2} \right) + \beta \mathbb{E}_{t+1/2} \left[V_{t+1}(H_{t+1}^s, Z_{t+1}) \right] \right\} \\ & \quad \text{subject to the human capital accumulation equation (2),} \end{aligned}$$

In this expression, $\mathbb{E}_{t+1/2}(.) = E(. \mid H_t, Z_{t+1/2})$ and β is the discount rate. This means in particular that the delay between t and $t+1/2$ is infinitely smaller than the delay between $t+1/2$ and $t+1$ despite our abusive but clear notation, 1/2.

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

$$s_t = e \text{ iff } \mathbb{E}_t W_t^e(H_t, Z_{t+1/2}) + \omega_t > \mathbb{E}_t W_t^n(H_t, Z_{t+1/2}), \tag{5}$$

where $\mathbb{E}_t(.) = E(. \mid H_t, Z_t)$, which allows us to complete the definition of the recursive equation in V_t:

$$V_t(H_t, Z_t) = \max(\mathbb{E}_t W_t^e(H_t, Z_{t+1/2}) + \omega_t, \mathbb{E}_t W_t^n(H_t, Z_{t+1/2})).$$

As sector choice, denoted by s_t, affects the accumulation of human capital, the optimal level of investment is $\tau_t^{s_t}$. The level of human capital at date $t+1$ is then given by the simplified notation, $H_{t+1} \equiv H_{t+1}^{s_t}$, reflecting that human capital is single dimensional (see Taber and Vejlin, 2020, Lise and Postel-Vinay, 2020, for multidimensional alternatives).

4.1.3 Individual specific terminal conditions

The terminal condition of this decision program could be given by an individual specific date at which investing in human capital stops (Ben Porath, 1967). We use here a dual formulation as described by an individual specific value of human capital stocks at an arbitrary date, $t_0 + d$, in the future.[2] Specifically, we write that at the future date $t_0 + d + 1$ the value function or the discounted value of the utility stream from $t_0 + d + 1$ onwards is given by:

$$V_{t_0+d+1}(H_{t_0+d+1}, Z_{t_0+d+1}) = a_{t_0+d+1}(Z_{t_0+d+1}) + \kappa \ln H_{t_0+d+1}, \tag{6}$$

in which the level a_{t_0+d+1} generically depends on Z_{t_0+d+1}, and parameter κ is the individual specific marginal valuation of log human capital in the final period. The latter commands the horizon effects in wages as shown below. It is not indexed by $t_0 + d + 1$ for notational simplicity, and is assumed to be independent of Z_{t_0+d+1}.

To complete the description of the economic model, we further assume that the distribution of future shocks $(\omega_l, \delta_l^s, \lambda_l^s)_{l \geq t}$ conditionally on $Z_{t-1/2}$ does not depend on the state variable history $H_t, H_{t-1}, ., H_1$.

4.2 Analysis

We now provide the steps leading to the resulting reduced form for log wages, through a sequence of Propositions which are proved in Appendix B.

[2] d is either the length of the working life or of the observed working life in our empirical analysis further on.

4.2.1 Value functions and life-cycle profile of investments

The sequence of investments between $t = t_0$ and the terminal date, $t_0 + d$, is called a life cycle profile of investments. We can analytically solve the dynamic model backwards because of linear assumptions, and the value functions are log-linear in human capital stocks.

Proposition 1 *The sequence of value functions writes:*

$$W_t^s(H_t, Z_{t+1/2}) = a_t^s(Z_{t+1/2}) + \kappa_t \log H_t \text{ for } s = e, n$$

and:

$$V_t(H_t, Z_t) = a_t(Z_t) + \kappa_t \log H_t$$

in which

$$\kappa_t = \frac{1}{1-\beta} + \beta^{t_0+d-t}(\kappa - \frac{1}{1-\beta}).$$

and the constant functions, $a_t^s(Z_{t+1/2})$ and $a_t(Z_t)$ are defined in Proposition 3.

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 $t = t_0$ and $t = t_0 + d$ in each sector s is:*

$$\tau_t^s = \max\{0, \frac{1}{c}(\rho^s \beta \kappa_{t+1} - 1)\} \tag{7}$$

which in turn determines the dynamic equation for the additive terms in the value functions.

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

$$a_t^s(Z_{t+1/2}) = \delta_t^s - \beta\kappa_{t+1}\lambda_t^s + c\frac{(\tau_t^s)^2}{2} + \beta\mathbb{E}_{t+1/2}\left[a_{t+1}(Z_{t+1})\right].$$

in which τ_t^s is the optimal value of human capital investment when being in sector s as defined in equation (7).

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:*

$$\begin{aligned} s_t &= e \text{ iff} \\ \omega_t + \mathbb{E}_t\left(\delta_t^e - \beta\kappa_{t+1}\lambda_t^e + c\frac{(\tau_t^e)^2}{2}\right) &\geq \mathbb{E}_t\left(\delta_t^n - \beta\kappa_{t+1}\lambda_t^n + c\frac{(\tau_t^n)^2}{2}\right). \end{aligned} \tag{8}$$

This is the equation that determines the structure of selection that we entertain below. We now turn to our main object of interest, the profile of log wages in the private sector.

4.3 The reduced form

Consider a worker who is in sector e at period t. As observations consist in wage histories starting in the private sector beginning at time t_0, $s_{t_0} = e$. Denote $t_1 = \min\{l; l \geq t_0, s_l = n\} \geq t_0$ the first period in sector n, t_2 the first return in sector e i.e. $t_2 = \min\{l; l > t_1, s_l = e\} > t_1 + 1$ and so forth by induction, and K_t the overall number of spells in sector n before period $t > t_0$. The vector $(t_0, t_1, t_2, ., t_{2K_t})$ is the sequence of transition dates into sector e (even index values) and into sector n (odd index values). We deduce from this setting that the mapping between date $l \leq t$ and sectoral choice is given by:

$$\begin{aligned} s_l &= e \text{ for } t_{2k} \leq l \leq t_{2k+1} - 1, \\ &= n \text{ for } t_{2k+1} \leq l \leq t_{2k+2} - 1, \text{ for } k \leqslant K_t. \end{aligned}$$

Proposition 5 *Consider a worker in sector e at date $t \in \{t_0, ., t_0 + d\}$ and assume that $\tau_l^{s_l} > 0$ for any $t_0 \leq l < t_0 + d + 1$. Log wages are:*

$$\ln y_t = \eta_0 + \eta_1 t + \eta_2 \beta^{-t} + \eta_3 x_t^{(3)} + \eta_4 x_t^{(4)} + \underbrace{\delta_t^e - \sum_{l=t_0}^{t-1} \lambda_l^{s_l}}_{v_t} \tag{9}$$

in which $\mathsf{H} = (\eta_0, \eta_1, \eta_2, \eta_3, \eta_4)$ are functions of parameters $(\rho^s, c, \beta, \kappa)$ as well as of the initial value of human capital stock $\ln H_{t_0}$. Variables $x_t^{(3)}$ and $x_t^{(4)}$ are defined by:

$$x_t^{(3)} = \sum_{k=0}^{K_t-1} (t_{2k+2} - t_{2k+1}); \qquad x_t^{(4)} = \sum_{k=0}^{K_t-1} \left(\beta^{-t_{2k+2}+1} - \beta^{-t_{2k+1}+1}\right). \tag{10}$$

This reduced form is the wage equation we estimate in our empirical application.

5 Econometric model

In our empirical analysis, we use panel data on males from period $t_0 = 1$ to T and rely on information on wages observed during spells of employment in the private sector to estimate equation (9). We allow structural parameters implicit in this equation to be individual specific. It concerns returns in both sectors, ρ_i^e, ρ_i^n, the cost of effort, c_i, the terminal value of human capital, κ_i, and the initial value of (log) human capital stocks, $\log(H_{i1})$. We restrict however the discount factor β to be homogeneous, as is commonly assumed. The log wage equation can thus be written as

$$\ln y_{it} = \eta_{i0} + \eta_{i1} t + \eta_{i2} \beta^{-t} + \eta_{i3} x_{it}^{(3)} + \eta_{i4} x_{it}^{(4)} + v_{it} = x_{it} \eta_i + v_{it},$$

where $v_{it} = \delta^e_{it} - \sum_{l=t_0}^{t-1} \lambda^{s_{il}}_{il}$ in which s_{il} is the sector chosen by individual i at period l, $x_{it} = (1, t, \beta^{-t}, x^{(3)}_{it}, x^{(4)}_{it})$ and $\eta_i = (\eta_{ij})_{j=0,.,4}$.[3]

We first analyze the identification of parameters, η_i, when selection in private sector employment is exogenous. We then turn to stating the conditions under which selection is conditionally exogenous.

5.1 Identification under exogenous selection

Assume that selection in the private sector is exogenous. Note that the number of structural parameters and the number of reduced form parameters are both equal to 5 for each individual.[4] Yet, a necessary condition for point identification is that there is enough individual mobility across sectors. Indeed, consider an individual i who is employed during the whole period in sector e, or who moves only once out of sector e to sector n, so that $x^{(3)}_{it} = x^{(4)}_{it} = 0$ 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, $x^{(3)}_{it} = (t_{2i} - t_{1i})\, 1_{\{t \geqslant t_{2i}\}}$ and $x^{(4)}_{it} = \left(\beta^{-t_{2i}} - \beta^{-t_{1i}}\right) 1_{\{t \geqslant t_{2i}\}}$, and the two variables $x^{(3)}_{it}$ and $x^{(4)}_{it}$ are proportional to $1_{\{t \geqslant t_{2i}\}}$ 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 $\eta_{i3}\,(t_{2i} - t_{1i}) + \eta_{i4}\left(\beta^{-t_{2i}} - \beta^{-t_{1i}}\right)$ 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) that parameters η_{i3} and η_{i4} are identified separately. Note that underidentification of parameters η_{i3} and η_{i4} does not affect the identification of the other parameters η_{i0}, η_{i1} and η_{i2}. This issue is akin to the one identified by Chernozhukov et al. (2013) in a treatment set-up.

5.2 Missing at random conditionally on factors (MARCOF) restriction

We now discuss the identifying assumptions that we adopt and that make selection exogeneous when we impose the structural model and in particular the participation equation (8). First, stochastic processes ω_{it} (i.e. private sector preference), δ^s_{it} (i.e. human capital price) and λ^s_{it} (i.e.

[3] From now on, we drop the superscript, e, to refer to the private sector by default. Otherwise, the superscript gives the sector.

[4] The derivation of structural parameters from **the reduced form**, as well as sufficient conditions for this derivation, is available upon request.

depreciation), are specified using factor structures:

$$\omega_{it} = \varphi_t^{(\omega)}\theta_i^{(\omega)} + \widetilde{\omega}_{it}, \tag{11}$$

$$\delta_{it}^s = \varphi_t^{(\delta),s}\theta_i^{(\delta),s} + \widetilde{\delta}_{it}^s, \tag{12}$$

$$\lambda_{it}^s = \varphi_t^{(\lambda),s}\theta_i^{(\lambda),s} + \widetilde{\lambda}_{it}^s. \tag{13}$$

in which residual random shocks are assumed to be mean independent of factors and factor loadings. In other words, they satisfy the following orthogonality restrictions for $t \geq 1, s \in \{n, e\}$:

$$E\left(\widetilde{\omega}_{it} \left| \Upsilon_t, \theta_i \right.\right) = E\left(\widetilde{\delta}_{it}^s \left| \Upsilon_{t+1/2}, \theta_i \right.\right) = E\left(\widetilde{\lambda}_{it}^s \left| \Upsilon_{t+1/2}, \theta_i \right.\right) = 0.$$

in which we denote $\Upsilon_t = \left\{\varphi_t^{(\omega)}, \Upsilon_{t-1}\right\}$ to mimic the construction of history Z_t for factors, and we define $\Upsilon_{t+1/2}$ accordingly, that is $\Upsilon_{t+1/2} = \left\{\varphi_t^{(\delta),e}, \varphi_t^{(\delta),s}, \varphi_t^{(\lambda),e}, \varphi_t^{(\lambda),n}, \Upsilon_t\right\}$, consistently with the timing of Figure 2. We also denote $\tilde{\theta}_i = \left\{\eta_i, \theta_i^{(\omega)}, \theta_i^{(\delta),e}, \theta_i^{(\delta),n}, \theta_i^{(\lambda),e}, \theta_i^{(\lambda),n}\right\}$.

We now strengthen this assumption into independence and mean independence restrictions that condition on information available just before the revelation of those shocks:

Assumption M(issing)A(t)R(andom)C(onditionally)O(n)F(actors):

$$\Pr(\widetilde{\omega}_{it} \leq \omega \mid \tilde{Z}_{t-1/2}, \Upsilon_t, \tilde{\theta}_i) = \Pr(\widetilde{\omega}_{it} \leq \omega \mid \Upsilon_t, \tilde{\theta}_i), \tag{14}$$

$$E((\widetilde{\delta}_{it}^s, \widetilde{\lambda}_{it}^s) \mid \tilde{Z}_t, \Upsilon_{t+1/2}, \tilde{\theta}_i) = 0. \tag{15}$$

in which we extend the notation Z_t and $Z_{t+1/2}$ in a natural way to $\tilde{Z}_t$ and $\tilde{Z}_{t+1/2}$ which now include the histories of residual random shocks $\widetilde{\omega}_{it}$, $\widetilde{\delta}_{it}^s$ and $\widetilde{\lambda}_{it}^s$.[5] Note that this assumption implies that $\{\widetilde{\omega}_{it}\}_{t\geq 1}$ and $\{(\widetilde{\delta}_{it}^s, \widetilde{\lambda}_{it}^s)\}_{t\geq 1}$ are independent and that they are both independently distributed over time.

We now prove that these assumptions in a linear factor setting imply that selection is exogenous and that experience variables $x_{it}^{(3)}$ and $x_{it}^{(4)}$ are exogenous. First rewrite the wage equation under assumptions (11)-(13):

$$\ln y_{it} = x_{it}\eta_i + \varphi_t^{(\delta)}\theta_i^{(\delta)} - \left[\sum_{l=t_0}^{t-1} \varphi_l^{(\lambda),n} 1_{\{s_{il}=n\}}\right]\theta_i^{(\lambda),n} - \left[\sum_{l=t_0}^{t-1} \varphi_l^{(\lambda),e} 1_{\{s_{il}=e\}}\right]\theta_i^{(\lambda),e} + \widetilde{v}_{it} \tag{16}$$

where $\widetilde{v}_{it} = \widetilde{\delta}_{it}^e - \sum_{l=t_0}^{t-1} \widetilde{\lambda}_l^{s_{il}}$.

Second, the sectoral choice equation (8) can be rewritten as:

$$\tilde{\omega}_{it} + \mathbb{E}_t\left(\tilde{\delta}_{it}^e - \beta\kappa_{it+1}\tilde{\lambda}_{it}^e\right) - \mathbb{E}_t\left(\tilde{\delta}_{it}^n - \beta\kappa_{it+1}\tilde{\lambda}_{it}^n\right) \geq f(\Upsilon_t, \tilde{\theta}_i)$$

[5] Variables $\tilde{\theta}_i$ and factors implicitly determine H_{it}.

in which the notation, $\mathbb{E}_t$, is defined in Section 4.1.2, and conditions on available information, and $f(\Upsilon_t, \tilde{\theta}_i)$ is a function of factors and factor loadings which subsumes investment terms like $c_i \frac{(\tau_{it}^s)^2}{2}$.[6] Because of condition (15) we have that $\mathbb{E}_t\left(\tilde{\delta}_{it}^s - \beta \kappa_{it+1} \tilde{\lambda}_{it}^s\right) = 0$ for $s = e, n$ and the selection equation rewrites as:

$$\tilde{\omega}_{it} \geq f(\Upsilon_t, \tilde{\theta}_i). \tag{17}$$

Furthermore, conditions (14) and (15) imply that:

- $\tilde{\omega}_{it}$ is independent of $\widetilde{\delta}_{it}$ given factors and factor loadings $(\Upsilon_{t+1/2}, \tilde{\theta}_i)$,
- $\tilde{\omega}_{it}$ is independent of the history of depreciation shocks, $\widetilde{\lambda}_i^s$, $s \in \{n, e\}$, up to date $t-1$, given factors and factor loadings $(\Upsilon_{t+1/2}, \tilde{\theta}_i)$,
- $\tilde{\omega}_{it}$ is independent of the history of sector preferences, $\widetilde{\omega}_i$, up to date $t-1$, given factors and factor loadings $(\Upsilon_t, \tilde{\theta}_i)$.

and this in turn implies that $\tilde{\omega}_{it}$ and $\widetilde{v}_{it}$ are independent given factors and factor loadings $(\Upsilon_{t+1/2}, \tilde{\theta}_i)$. This proves that under conditions (14) and (15), selection is exogenous.[7]

Moreover, explanatory variables $x_{it}^{(3)}$ and $x_{it}^{(4)}$ are exogenous under the same conditions. Indeed, these two variables can be written as functions of past sectoral choices as stated in equations (10). We evaluate $E\left(\widetilde{v}_{it} \middle| x_{it}^{(3)}, x_{it}^{(4)}, \Upsilon_{t+1/2}, \tilde{\theta}_i\right)$ as given by equation (16). First, $(x_{it}^{(3)}, x_{it}^{(4)})$ and $\widetilde{\delta}_{it}^e$ are mean independent because of condition (15). Moreover, the second term of $\widetilde{v}_{it}$ is such that:

$$\begin{aligned}
& E\left[\left(\sum_{l=t_0}^{t-1} \widetilde{\lambda}_{il}^s 1_{\{s_{il}=s\}}\right) \mid x_{it}^{(3)}, x_{it}^{(4)}, \Upsilon_t, \tilde{\theta}_i\right] \\
= \; & E\left[E\left(\sum_{l=t_0}^{t-1} \widetilde{\lambda}_{il}^s 1_{\{s_{il}=s\}} \mid \omega_{it-1}, ., \omega_{i1}, \Upsilon_{t+1/2}, \tilde{\theta}_i\right) \mid x_{it}^{(3)}, x_{it}^{(4)}, \Upsilon_t, \tilde{\theta}_i\right] \\
= \; & E\left[\left(\sum_{l=t_0}^{t-1} E(\widetilde{\lambda}_{il}^s \mid \omega_{it-1}, ., \omega_{i1}, \Upsilon_{t+1/2}, \tilde{\theta}_i) 1_{\{s_{il}=s\}}\right) \mid x_{it}^{(3)}, x_{it}^{(4)}, \Upsilon_t, \tilde{\theta}_i\right] \\
= \; & 0
\end{aligned}$$

because the processes $\{\widetilde{\omega}_{it}\}_{t \geq 1}$ and $\{(\widetilde{\delta}_{it}^s, \widetilde{\lambda}_{it}^s)\}_{t \geq 1}$ are independent over time conditional on factors and factor loadings. This is why we obtain that covariates are exogenous since:

$$E\left(\widetilde{v}_{it} \middle| x_{it}^{(3)}, x_{it}^{(4)}, \Upsilon_{t+1/2}, \tilde{\theta}_i\right) = 0.$$

[6] Indeed, the sector choice depends on the terms τ_{it}^s as shown by equation (8), which depend themselves on ρ_i^s through equation (7).

[7] These conditions are sufficient but far from necessary, and they are stated this way for the sake of simplicity.

5.3 Structural functions and counterfactuals

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 in the 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 the structural objects and then explain how to compute them in practice. We drop individual index i for simplicity.

These structural objects are obtained by manipulating the history of interruptions along the life-cycle while keeping constant the individual structural parameters $\eta = (\eta_0, \eta_1, \eta_2, \eta_3, \eta_4)$. First, denote S_t the counterfactual individual choice of a sector at time t and $S^{(t)} = \{S_1, ., S_t\}$ the counterfactual history.[8] Second, denote $x_t^{(3)}(S^{(t-1)})$ and $x_t^{(4)}(S^{(t-1)})$ the reduced-form variables as defined by equations (10) but as functions of potential history $S^{(t-1)}$. By extension, the observed values are $x_t^{(3)} = x_t^{(3)}(s^{(t-1)})$ and $x_t^{(4)} = x_t^{(4)}(s^{(t-1)})$. The list of counterfactual explanatory variables in equation (9) are defined as $x_t(S^{(t-1)}) = (1, t, \beta^{-t}, x_t^{(3)}(S^{(t-1)}), x_t^{(4)}(S^{(t-1)}))$. In consequence, define the counterfactual outcomes as $x_t(S^{(t-1)})\eta$, the expectation of net log wages of an individual given by a specific realization of η and $S^{(t-1)}$. Idiosyncratic shocks are integrated out using the fact that the interactive effects controlling for selection have mean zero.

Given the distribution of η, and the counterfactual distribution of $S^{(t-1)}$, average structural functions are defined as:

$$E_{\eta, S^{(t-1)}}(\Phi\left(x_t(S^{(t-1)})\eta\right)), \tag{18}$$

in which function $\Phi(.)$ can be of various types: the identity function to recover means, squares to recover variances, or indicator functions to recover interquartile or inter-decile ranges. For instance, if $S^{(t-1)}$ stands for the history of continuous participation in sector e, $S^{(t-1)} = (e, .., e)$, this expression defines the average counterfactual net log wages at time t as if participation had been continuous.

Two extensions are worth considering. First, structural functions can be defined conditionally on observed participation in the private sector, e.g:

$$E_{\eta, S^{(t-1)}}(\Phi\left(x_t(S^{(t-1)})\eta\right) \mid s_t = e\right),$$

is the average counterfactual net wage for those who are working in sector e at time t and setting the potential history to $S^{(t-1)}$. Another extension involves several periods in which instead of being interested in the net log wage at a single period only, $x_t(S^{(t-1)})\eta$, we could be interested

[8] Recall that s_t is the observed sector **at date** t and by extension $s^{(t)}$ is the realized history of sectoral choices.

in the discounted sum of net log wages as in

$$\sum_{\tau=1}^{t} \beta^{\tau} x_t(S^{(t-1)})\eta. \tag{19}$$

We now review in more detail specific structural effects. We will have to keep in mind that under-identification of (η_3, η_4) in the subpopulation with fewer than two interruptions makes unidentified any counterfactual in which the individual specific numbers of interruptions increase in that population.

5.3.1 Selection and interruption effects

We compute structural functions (18) contrasting the observed and counterfactual situations where selection effects on wages are neutralized. We distinguish selection effects due to individuals being out of the private sector at a given date, and interruption effects due to past spells out of the private sector. We thus compute equation (18) in the following four cases:

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

$$\Phi_t^{(0)} = E_{\eta, s^{(t-1)}}(\Phi\left(x_t(s^{(t-1)})\eta \mid s_t = e\right).$$

- The situation in which interruption effects are neutralized, i.e. $S^{(t-1)} = (e, ., e)$. There is no career interruption and $x_t^{(3)}(S^{(t-1)}) = x_t^{(4)}(S^{(t-1)}) = 0$:

$$\Phi_t^{(1)} = E_{\eta}(\Phi\left(x_t(e, ., e)\eta\right) \mid s_t = e).$$

- The situation in which selection effects are neutralized:

$$\Phi_t^{(2)} = E_{\eta, s^{(t-1)}}(\Phi\left(x_t(s^{(t-1)})\eta\right).$$

- The situation in which both selection and interruption effects are neutralized:

$$\Phi_t^{(3)} = E_{\eta}(\Phi\left(x_t(e, ., e)\eta\right)).$$

We can contrast $\{\Phi_t^{(j)}\}_{j=1,.,3}$ with benchmark $\Phi_t^{(0)}$. The same experiment can be performed using discounted values as in equation (19).

5.3.2 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 and the sum of discounted wages since yearly wages are partly determined by all past spells out of employment. We restrict our attention to individuals whose parameters related to spells out of private sector are identified, i.e. to individuals experiencing at least two such spells followed by employment spells ($K_T \geq 2$).

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. The other counterfactual exercises consist in reassigning interruptions either at the end of the observed life cycle (imposing at least one year of presence in the private sector) or at the beginning. Comparing these two counterfactuals to the benchmark, we can measure the wage change 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). More details are given in Appendix A.3.

6 Estimation strategy

Our estimation strategy is driven by our available data which consist in employment status and wage histories when employed in the private sector for cohorts of individuals entering the labour market between 1985 and 1992 in France and potentially observed until 2012. No information is available when individuals are not employed by the private sector.

6.1 Estimation procedure

We estimate the model pooling all cohorts together and making the simplifying assumption that factors and factor loadings associated with the depreciation rate of human capital are the same in both sectors: $\varphi_t^{(\lambda),e} = \varphi_t^{(\lambda),n} \equiv \varphi_t^{(\lambda)}$ and $\theta_i^{(\lambda),e} = \theta_i^{(\lambda),n} \equiv \theta_i^{(\lambda)}$. The wage equation becomes:

$$\begin{aligned}
\ln y_{it} &= x_{it}\eta_i + \varphi_t^{(\delta)}\theta_i^{(\delta)} - \sum_{l=t_{0i}}^{t-1} \varphi_l^{(\lambda)}\theta_i^{(\lambda)} + \widetilde{v}_{it} \\
&= -\left(\sum_{l=t_{0i}}^{T} \varphi_l^{(\lambda)}\right)\theta_i^{(\lambda)} + x_{it}\eta_i + \varphi_t^{(\delta)}\theta_i^{(\delta)} + \left(\sum_{l=t}^{T} \varphi_l^{(\lambda)}\right)\theta_i^{(\lambda)} + \widetilde{v}_{it}
\end{aligned}$$

Without loss of generality, we make the normalization $\sum_{l=t_{0i}}^{T} \varphi_l^{(\lambda)} = 0$ (since this term enters the additive individual fixed effect) and, denoting $\widetilde{\varphi}_t^{(\lambda)} = \sum_{l=t}^{T} \varphi_l^{(\lambda)}$, we get:

$$\ln y_{it} = x_{it}\eta_i + \varphi_t^{(\delta)}\theta_i^{(\delta)} + \widetilde{\varphi}_t^{(\lambda)}\theta_i^{(\lambda)} + \widetilde{v}_{it}$$

Interactive terms associated with the rental price and depreciation rate of human capital enter additively in a similar way in the wage equation and they are thus undistinguishable. Without loss of generality, we relabel $\varphi_t^{(\delta)}\theta_i^{(\delta)} + \varphi_t^{(\lambda)}\theta_i^{(\lambda)}$ as $\varphi_t\theta_i$:

$$\ln y_{it} = x_{it}\eta_i + \varphi_t\theta_i + \widetilde{v}_{it} \tag{20}$$

in which:

$$x_{it} = \left(1, t, \beta^{-t}, x_{it}^{(3)}, x_{it}^{(4)}\right); \qquad \eta_i = \{\eta_{i0}, \eta_{i1}, \eta_{i2}, \eta_{i3}, \eta_{i4}\}'$$

To identify individual specific effects η_{i0}, η_{i1} and η_{i2} interacting with observed individual-invariant explanatory variables, t and β^{-t}, a restriction on factors is needed, and provided by the flat spot approach:

$$\varphi \perp \left(x^{(0)}, x^{(1)}, x^{(2)}\right) \tag{21}$$

with $\varphi = (\varphi_1, ..., \varphi_T)'$, $x^{(0)} = (1, ..., 1)'$, $x^{(1)} = (1, ..., T)'$ and $x^{(2)} = (1, ..., \beta^{-T})'$. Note that this restriction yields, in particular, that the time average of factors is zero: $\sum_{t=1}^{T} \varphi_t = 0$ while there is no such restriction on factor loadings θ_i that can be freely correlated with the terms η_i.

Our approach consists in minimizing the sum of squares of residuals for observations for which wages are observed which is equivalent to maximizing the pseudo-likelihood function of normal disturbances. 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 wages with their linear predictions at dates at which workers have not yet entered the labor market or are not employed (sector n). In the maximization step, we maximize the pseudo-likelihood for observations corresponding to all individuals and dates.

Our iteration algorithm runs as follows. We use (k) as a superscript for parameters at step k. To obtain initial values, we follow Moon and Weidner (2018), 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. By contrast, the least squares minimization program is not, and may yield several local solutions (se Hsiao, 2018). Second, we conduct a principal component analysis of $\ln y_{it} - x_{it}\eta_i^{(0)}$ (whose value is imputed to zero when y_{it} is not observed), and we get initial factor values $\varphi^{(0)}$ such that $\frac{\varphi^{(0)}(\varphi^{(0)})'}{T} = I$. The updating from step $k-1$ to step k is the following:

1. We regress y_{it} on x_{it} and $\varphi_t^{(k-1)}$ for each individual, considering only periods at which they are observed, and we recover the estimators $\eta_i^{(k)}$ and $\theta_i^{(k)}$.[9]

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

3. We estimate the factor model: $\ln y_{it} - x_{it}\eta_i^{(k)} = \varphi_t\theta_i + \widetilde{\widetilde{v}}_{it}$, and recover the estimator $\varphi_t^{(k)}$ using Bai (2009)'s approach. Stacking $\{\varphi_t\}_{t=1,.,T}$ into matrix φ and $\{\theta_i\}_{i=1,.,N}$ into matrix θ, we impose the usual identification restrictions that $\varphi\varphi'/T = I$, $\theta\theta'/N$ is diagonal, and the first element of each row of φ_1 is positive.[10] For the additional identification restrictions (21) to be verified, we project $\varphi^{(k)}$ on the space orthogonal to $x^{(0)}$, $x^{(1)}$ and $x^{(2)}$. We then re-normalize the projection within this space such that the identification restriction $\varphi^{(k)}\varphi^{(k)\prime}/T = I$ is still verified and such that $\varphi_1 > 0$.

The stopping rule of the iterative procedure is detailed in Appendix D.1. In Appendix D.2, we further show that this EM algorithm is valid using Heyden and Morton (1996). It delivers the pseudo-ML estimators of parameters.

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. We complete them in Appendix C by adding the proof of the invertibility of a matrix which was missing. Note though 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 periods for every individual i.

6.2 Bias correction of counterfactuals

Using estimated parameters, we compute structural functions of potential outcomes as defined in Section 5.3. Their empirical counterparts, however, generically suffer from the incidental parameter issue. Variances and other summary statistics like quantiles are biased (Fernandez-Val and Weidner, 2018). Biases of variances and covariances can be corrected when the covariance

[9] 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 y_{it} 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 Appendix D.2.

[10] Alternatively, regressing $y_{it} - x_{it}\eta_i^{(k)}$ on $\theta_i^{(k)}$ under the constraint $\varphi\varphi'/T = I$ would deliver another estimate of φ.

matrix of idiosyncratic errors is restricted as shown by Arellano and Bonhomme (2012) (see Appendix E.1). For quantiles and interquantile ranges, we resort to the bias-correction procedure based on Taylor expansions proposed by Jochmans and Weidner (2019) as developed further in Appendix E.2.

Bias corrections rely on asymptotic formulas established when the number of individuals and the number of periods during which they are employed in the private sector tend to infinity. Some individuals are employed during 15 periods only whereas the model involves up to 7 individual parameters capturing the individual unobserved heterogeneity. 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 Appendix F. 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 (2019). This is why we focus mostly on estimating means, deciles and inter-decile ranges in our empirical application. Monte Carlo results also show that bias-correction for centiles and inter-centile differences work better when disturbances, v_{it}, are homoskedastic. This is because bias correction terms involve individual variances of residuals, and some of them can be poorly estimated. We therefore assume that disturbances are homoskedastic when computing estimated standard errors.

7 Estimation Results

In this section, we first present estimation results of different specifications of the model, and justify our preferrence for the specification given by equation (20) and including two unobserved factors. We then characterize counterfactual wage profiles in the four cases introduced in Section 5.3.1, and in the case of random, early or late interruptions introduced in Section 5.3.2.

7.1 Model selection and comparisons

We estimated 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 $3)$.[11] 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

[11] Appendix A.2 discusses results of the homogenous Mincer equation (20).

coefficients of interruption variables and factor loadings in no-, one- and two-factor specifications rejects that those are equal to zero; (2) Four 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 Table 2); (3) Estimates for the three-factor specification are quite unstable signaling possible identification issues and overfitting.

We now 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 log wage i.e. as if all spells in the private sector were uninterrupted. Note that the potential log wage only depends on the estimates of the first three individual specific coefficients, i.e. $\eta_{i0}, \eta_{i1}, \eta_{i2}$.

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 either interruptions or selection into the private sector have significant effects on wages, and that ignoring them downward biases returns to potential experience.

Figure 3(c) displays the profiles of the uncorrected variance of potential log wages and the bias-corrected estimates are displayed in Figure 3(d). The comparison between them shows how large the bias in variances is. Furthermore, these graphs show that variance estimates are larger for the main specification than 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. The profile of high-return workers, who invest more in human capital at the beginning of their life-cycle, crosses after a few years the profile of low-return workers. The crossing point is estimated at about 5 years.

Monte Carlo experiments taught us, however, that biases in corrected variances might remain sizable. This is why we now turn to the profile of the inter-decile range of potential log wages as defined as the difference between the 90% quantile and the 10% quantile (Figure 3(e)). Correcting the bias for the inter-decile range mildly affects these profiles by at most 10% (Figure 3(f)). The inter-decile range for the main specification is hovering between 90% and 140%, 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 5 years of potential experience.

Overall, comparisons between model specifications have shown that omitting interruptions and factors dampens the dispersion of log wages. We shall return to this important point when analyzing counterfactual results in Subsection 7.3.

7.2 Estimated coefficients and the components of wages

We present descriptive statistics on the distributions of estimated parameters for our main specification.[12] Table A.4 reports means as well as uncorrected and bias-corrected variances and quantiles. The corrected median of individual specific parameters, capturing growth, η_{i1}, and curvature, η_{i2}, are respectively positive and negative. This means that potential experience has a positive effect on wages but its return decreases with the number of years consistently with our theoretical model. These results on medians however mask important heterogeneity. For instance, the 90% quantile of parameter, η_{i1}, is 6.2 times larger than the median when both are bias-corrected. In the bottom, the estimate of the first quartile is negative indicating that the estimated growth is negative for a non-negligible share of the population. In the same vein, the bias-corrected estimate of the 10% quantile of the curvature parameter, η_{i2}, is around 10 times larger in absolute value than the median estimate.

Parameters, η_{i3} (years of interruptions), and, η_{i4} (curvature in interruptions) are not identified – and set to zero in Table A.4 – in the sub-sample of individuals who have no or only one interruption during the observation period (see Section 5.1). This is why we report results on their distribution in the subsample of individuals with 2 or more interruptions in Table A.5. Bias-corrected medians of the effects of interruptions, η_{i3}, and the related curvature terms, η_{i4}, are very close to zero although the heterogeneity in estimates of those parameters is even larger than for the η_{i1} and η_{i2} estimates. Interestingly, the distributions of these effects seem quite symmetric around zero.[13]

Components of wages Given parameter estimates, we decompose log wages into 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 explained already, we instead report the more robust inter-decile ranges and rank correlations. Results on inter-decile ranges in Table A.3 show that the potential experience component is the largest but the interruption component is sizable. Factors play a role albeit a minor one. Remarkably, the potential experience and interruptions components are highly negatively rank-correlated. This

[12] Factor estimates are displayed in Figure S.1.

[13] We can also compare uncorrected and **bias-corrected values**. Medians of parameters are not affected by correction. By contrast, bias-correction changes the values of quantiles, and the spread of parameter distributions decreases. In particular, bias-corrected 90% (resp. 10%) quantile of parameter η_{i1} (resp. η_{i3} in absolute terms) is 20% (resp. 27%) smaller. Such effect of bias correction can also be observed for η_{i2} and η_{i4}.

can be mostly explained by the negative correlation between the linear coefficients, η_{i1} and η_{i3}. Their Spearman rank correlation is equal to -0.32.

7.3 The counterfactual structural impacts of interruptions and current-period selection

Using the results of the main specification and bias-correction, we now contrast different structural objects as defined in Section 5.3 to assess the economic importance of interruptions and participation on the profile of log wages. We define the benchmark as the profile of log wages for private sector employees when log wages are predicted including interruptions and excluding factors.

We compare profiles of summary statistics in this benchmark and in three counterfactual situations defined in the absence of current-period participation selection ("no selection"), in the absence of interruptions ("no interruption", that is the potential wage) or both ("no selection, no interruption"). The last one corresponds to the potential log wages in the absence of current-period selection that we studied above (see Figure 3, "Main").

First, in every graph of medians, means, variances or inter-decile ranges that we report in Figure 4, current-period selection does not have a significant effect. This agrees well with the small magnitude of the inter-decile range of interactive effects due to factors that we found in Table A.3. By contrast, interruptions have a strong and significant effect. Those results highlight the importance of taking into account interruptions when predicting wage profiles. In other words, both potential and real experience matter (e.g. Light and Ureta, 1995, Das and Polachek, 2019).

Figures 4(a) and 4(b) show that potential experience increases log wages by around 65% in 20 years. This result squares well with other studies which cover many countries and use homogeneous Mincer equations (e.g. Lagakos et al., 2018). In addition, the average cost of interruptions after 20 years is about 10%.

The impact on dispersion of wages, which has not been documented so far in the literature, is shown in Figure 4(d) through the lens of inter-decile ranges (see also Figure 4(c) for variances). After 20 years, the average duration of interruptions is 2.47 years, and interruptions decrease dispersion by -0.52 (-38%). This effect plays both at the 90% quantile and the 10% quantile. It is stronger at the 90% quantile (-0.34 after 20 years) than at the 10% quantile ($+0.18$) as shown by Figure A.2. These results could be due to very different causes: minimum wage constraints at the bottom of the distribution of wages, self-employment or employment abroad at the top of the distribution. Absent further information, we are not able to investigate those causes more precisely.

These results on the impact of interruptions on dispersion stem from the negative rank correlation between the potential wage and the interruptions effects that we mentioned when commenting Table A.3. To go further, we can analyse counterfactuals in order to disentangle the effects of the timing and length of interruption spells (i.e. $x_{it}^{(3)}$ and $x_{it}^{(4)}$), and the impact of differential sector-specific returns to investments (i.e. parameters η_{i3} and η_{i4}).

Finally, we replicated results in this Section for every education group and the stylized facts on the profiles of means, medians and inter-decile ranges were similar.[14]

7.4 The counterfactual timing of interruptions

For our additional counterfactuals, we have to restrict the population to workers with at least two interruptions since parameters η_{i3} and η_{i4} are identified only for those. The impact of interruptions on dispersion shown by Figure 4(d) persists and if anything is larger (Figure A.3).

The first counterfactual experiment we analyze regards the estimated effect of interruptions on wages when years of interruption are randomly assigned over time for each worker. 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 definite attrition (the "Random" case). We compare these counterfactual wage profiles to those for which years of interruptions are the observed ones, and participation selection is absent (the "No selection" case). Results reported in Figure 5 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 the likely endogenous choices of interruptions. By contrast, inter-decile ranges start diverging after 5 years (Figure 5(c)) when wage dispersion increases more quickly in the random case than in the no selection one. This divergence comes from changes in both the first and last deciles (Figure 5(d)).

This result allows the economic interpretation of Figure 4(d) to be refined. It showed the new stylized fact that interruptions in private sector participation shrink the dispersion of human capital stocks after 20 years. According to Figures 5(c) and 5(d), an explanation of this stylized fact is the correlation between the timing of interruptions and potential log wages, and not the correlation between the estimated coefficients of interruptions and potential log wages. Surprisingly, endogenously-chosen interruptions smooth inequalities over time.

[14] Results are available upon request.

We can also estimate other counterfactuals related to the structure of interruptions by reassigning interruptions either at the beginning or at the end of the observed life-cycle as was studied by Light and Ureta (1995). Again, we contrast those counterfactuals with the counterfactual in which selection is absent ("No selection"). Results are reported in Figure 6.

Reassigning interruptions at the beginning of the working life 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 catch up what they have initially lost while median log wages do. In contrast, when reassigning interruptions at the end of observed life-cycle, mean log wages increase above what is observed absent selection, for any number of years of experience. Effects are smaller and insignificant for median log wage profiles.

Interestingly, reassigning interruptions at the beginning of observed life-cycle largely increases the inter-decile range over the whole life-cycle with respect to the benchmark (Figure 6(c)). This increase is larger at the beginning of the life-cycle as expected, and it slowly fans out after 6 years, presumably because fewer and fewer interruptions are reallocated between the two scenarii.

This widening of the inter-decile range is due to both a higher 90% quantile and a lower 10% quantile (Figure 6(d)). A higher 90% quantile can be explained by exits to self-employment or employment abroad which enable a faster accumulation of human capital than when being employed in the private sector. This gain does not disappear over time. In contrast, a lower 10% quantile can be explained by a lower human capital accumulation at this quantile when being out of the private sector. This is presumably due to unemployment or non-employment spells with low human capital investments.

In the same Figure, we also report the results of the experiment when artificially moving interruptions to the periods preceding the last period of observation of individual histories before definite attrition. The rise in the inter-decile range first parallels the trend observed in Figure 5(d) before taking off quite steeply after 15 years. This is partly due to interruptions having a stronger effect because of the geometric terms which enter the construction of $x_{it}^{(4)}$ whose coefficient is η_{i4}.

8 Discussion

Our working sample excludes workers who have severely incomplete histories, since we selected out workers for whom we have fewer than fifteen observations.

One can wonder whether such a selection has an effect on our results about the dispersion of wages. For instance, within our working sample of 15+ observations, the dispersion seems to decrease with the number of non-missing observations (Tables S.1, S.2 and S.3).[15] There are two explanations: The first one is the remaining asymptotic bias in $1/T^2$ that our estimation method entails; The second one is the substantive fact that the more incomplete observations are, the more dispersed the wage profiles.

We can first assess to what extent our selection that wages are observed at least 15 years alters observed wage profiles. The mean log wage and inter-decile range profiles for our restricted sample displayed in Figure A.1 are found to be very similar to those for the whole population but the power of this omnibus test might be low.

A more robust exercise whose statistical properties are left for further research is the following. Consider two samples: The first one comprises individuals who are observed between 10 and 14 years and who are excluded from our working sample; The second one comprises individuals in our working sample who are observed more than 20 years. In the following, we set the values of factors to those estimated in the working sample, and we thus assume that factors remain the same when we extend the model to the additional population $10-14$.

In the second sample (20+), we randomly draw the number of periods of observations for every individual so that the marginal distribution of this number is the same as its marginal distribution in the first sample $(10-14)$. In other words, given the number of observations for each individual i, say T_i, we set $S_i^{(1)} = T_i$ in the first sample $(10-14)$. In the second sample, we draw $S_i^{(2)} < T_i$ in a way that respects the two conditions; (1) Marginal distributions of $S_i^{(1)}$ and $S_i^{(2)}$ are the same (2) The first period observation for every individual is always included while the $S_i^{(2)} - 1$ further wage observations are randomly retained. Explanatory variables $x_{it}^{(3)}$ and $x_{it}^{(4)}$ remain the same as in the original sample and it is only the dependent variable, log wage, of which more instances are considered as "missing".

We then estimate the individual specific coefficients using the first incomplete sample (10-14), and we re-estimate these parameters using the second censored sample. Both sample estimates are then compared to the second sample original estimates. Biases due to the number of observations should then be neutralized if our selection of sample (10-14) and censored sample (20+) is exogenous. We evaluate these biases by comparing profiles of log wages along the life-cycle.

Formally, consider wage observations, $\ln y_i = (\ln y_{i1}, ., \ln y_{it}, ., ., \ln y_{iT})$ and participation $s_i = (s_{i1}, ., s_{it}, ., ., s_{iT})$. In the original samples, non missing observations are $\ln y_i * s_i =$

[15] All Tables and Figures which number is preceded by the S letter are relegated in the Online Supplementary Appendix.

$(\ln y_{i1}s_{i1}, ., \ln y_{it}s_{it}, ., ., \ln y_{iT}s_{\underset{T}{i}})$ and the final data are $(\ln y_i * s_i, x_i)$. Denote parameter estimates, $\hat{\eta}_i^{(1)}$ (respectively $\hat{\eta}_i^{(2)}$ in the sample (10-14) (resp. (20+)). When censoring sample (20+), we replace s_i by $s_i^{(3)}$ and the data is transformed into $(\ln y_i * s_i^{(3)}, x_i)$. Denote the corresponding estimate $\hat{\eta}_i^{(3)}$. The experiment we performed is to compare the distributions of three predicted variables, $\{x_i^c \hat{\eta}_i^{(j)}\}_{j=1,.,3}$ while including or not interruptions in the definition of counterfactual covariates, x_i^c.

Figure 7(a) shows that median profiles for the uncensored and censored samples 20+ are similar as well as inter-decile ranges displayed in Figure 7(b), although ranges are slightly more dispersed when potential experience is greater than 7 in the censored sample. In contrast, Figure 7(c) and 7(d) show that profiles differ greatly between the samples (10-14) and censored (20+). In particular, there are sizable median gaps increasing with potential experience between the benchmark and no-interruption cases for the sample (10-14) but not for the censored sample (20+). This can be explained by low returns to human capital outside the private sector for sample (10-14). A larger heterogeneity of returns in sample (10-14) explains the larger inter-decile range for sample (10-14) as well as its larger increase when interruptions are set to zero.

Overall, selection of the working sample seems to matter and the results which were derived above are thus restricted to the working sample that we considered.

9 Conclusion

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

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

We show how strong the dynamic selection effects are in the location and dispersion summaries of wage profiles whereas current-period selection effects are much weaker. Furthermore, past interruptions in participation in the private sector decrease mean and median wages as expected whereas surprisingly, the more interrupted participation is, the less dispersed wages are.

Moreover, the latter result can be mainly 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 have produced as well other statistics of interest such as the discounted sums of log wages, e.g. the integral of wage profiles as in Magnac and Roux (2021). There also remains the issue of external validity regarding selection since our working sample is restricted to individuals who participated in the private sector during at least 15 years. Further research should explore the empirical strategy of using restrictions on heterogeneity that would allow to weaken selection issues. Trade-offs exist however because these restictions are affecting precisely what we want to measure, i.e. wage inequalities.

REFERENCES

Aakvik, A, J.J., Heckman and E., 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.

Abowd, J. M., Crépon, B., & Kramarz, F., 2001, Moment estimation with attrition: an application to economic models. *Journal of the American statistical Association,* 96(456), 1223-1231.

Ahn, S. C., Lee, Y. H., & Schmidt, P., 2001, "GMM estimation of linear panel data models with time-varying individual effects", *Journal of Econometrics,* 101(2), 219-255.

Albouy, V., & C. Tavan , 2007, "Accès à l'enseignement supérieur en France: une démocratisation réelle mais de faible ampleur", *Économie et Statistique*, 410:3-22.

Altuğ, S., & Miller, R. A. (1998). "The effect of work experience on female wages and labour supply", *The Review of Economic Studies*, 65(1), 45-85.

Antonczyk, D., DeLeire, T., & Fitzenberger, B. (2018), "Polarization and rising wage inequality: comparing the US and Germany", *Econometrics*, 6(2), 20.

Arellano M., & S., Bonhomme, 2012, "Identifying Distributional Characteristics in Random Coefficient Panel Data Models", *The Review of Economic Studies*, 79(3): 987-1020.

Arellano, M., R., Blundell and S., Bonhomme, 2017, "Earnings and consumption dynamics: a nonlinear panel data framework", *Econometrica*, 85(3), 693-734.

Bai, J., & Ng, S., 2002, "Determining the number of factors in approximate factor models" , *Econometrica*, 70(1), 191-221.

Baker, M., 1997, "Growth-rate heterogeneity and the covariance structure of life-cycle earnings", *Journal of Labor Economics* 15(2):338-375.

Balakrishnan, S., Wainwright, M. J., & Yu, B., 2017, Statistical guarantees for the EM algorithm: From population to sample-based analysis. *The Annals of Statistics*, 45(1), 77-120.

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

Beyhum, J., & Gautier, E., 2019, "Square-root nuclear norm penalized estimator for panel data models with approximately low-rank unobserved heterogeneity", TSE Working Paper, n° 19-1008.

Biewen, M., Fitzenberger, B., & de Lazzer, J., 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.

Blundell, R., 2014, "Income Dynamics and Life-cycle inequality: Mechanisms and Controversies", *The Economic Journal*, 124:289-318.

Blundell, R., M., Costa-Dias, C., Meghir and J., Shaw, 2016, "Female Labor Supply, Human Capital, and Welfare Reform", *Econometrica*, 84(5):1705–1753.

Blundell, R. W., & Powell, J. L., 2003, "Endogeneity in nonparametric and semiparametric regression models", *Advances in Economics and Econometrics: Theory and Applications,* II: 312 & 357.(M. Dewatripont, LP Hansen, SJ Turnovsky, Eds.).

Bollinger, C. R., Hirsch, B. T., Hokayem, C. M., & Ziliak, J. P., 2019, "Trouble in the Tails? What We Know about Earnings Nonresponse Thirty Years after Lillard, Smith, and Welch.", *Journal of Political Economy*, 127(5), 2143-2185.

Bowlus, A. and C. Robinson, 2012, "Human Capital Prices, Productivity, and Growth", *American Economic Review*, 102(7), 3483-3515.

Bozio, A., Breda, T., & Guillot, M. (2020), "The Contribution of Payroll Taxation to Wage Inequality in France", unpublished manuscript.

Browning, M., Ejrnaes, M., & J., Alvarez, J., 2012, “Modelling income processes with lots of heterogeneity”, *Review of Economic Studies*, 77(4):1353-1381

Browning, M., Hansen, L. P., & J. J., Heckman, 1999, “Micro data and general equilibrium models”, in J.B. Taylor and M.Woodford (eds), *Handbook of Macroeconomics*, Elsevier Science, Amsterdam.

Cette, G., Chouard, V., & G., Verdugo, 2012, “Les effets des hausses du Smic sur le salaire moyen”, *Économie et Statistique*, 448-449:3-28.

Chaudhuri, S., & Guilkey, D. K., 2016, GMM with multiple missing variables. *Journal of Applied Econometrics*, 31(4), 678-706.

Chernozhukov, V., Fernández-Val, I., Hahn, J., & Newey, W. (2013), "Average and quantile effects in nonseparable panel models", *Econometrica*, 81(2), 535-580.

Das, T., & S., Polachek, 2019, "Microfoundations of Earnings Differences" in *The Palgrave Handbook of Economic Performance Analysis*, Palgrave Macmillan, Cham, p. 9-76.

Dominitz, J., & Sherman, R. P., 2005,. Some convergence theory for iterative estimation procedures with an application to semiparametric estimation. *Econometric Theory*, 21(4), 838-863.

Eckstein, Z., & Wolpin, K. I., 1989, "Dynamic labour force participation of married women and endogenous work experience", *The Review of Economic Studies*, 56(3), 375-390.

Fernández-Val, I., & Weidner, M., 2018, "Fixed effects estimation of large-T panel data models", *Annual Review of Economics*, 10, 109-138.

Gobillon, L., & T., Magnac, 2016, "Regional policy evaluation: Interactive fixed effects and synthetic controls", *Review of Economics and Statistics*, 98(3), 535-551.

Graham, B. S., de Xavier Pinto, C. C., & Egel, D., 2012, Inverse probability tilting for moment condition models with missing data. *The Review of Economic Studies*, 79(3), 1053-1079.

Graham, B. S., & Powell, J. L., 2012, "Identification and estimation of average partial effects in “irregular” correlated random coefficient panel data models", *Econometrica*, 80(5), 2105-2152.

Guvenen, F., Karahan, F., Ozkan, S., & Song, J., 2021, "What Do Data on Millions of US Workers Reveal about Life-Cycle Earnings Risk?", *Econometrica*, 89(5), 2303–2339

Heckman, J.J., L. Lochner, & C., 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, J. J., Lochner, L. L., & Todd, P.E., 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.

Heckman, J. J., & R., Robb, 1985, "Alternative methods for evaluating the impact of interventions: An overview", *Journal of Econometrics*, 30(1), 239-267.

Heyde, C., & G., Morton, 1996, “Quasi-likelihood and generalizing the EM algorithm”, *Journal of the Royal Statistical Association* B, 59(2), pp. 317-327.

Hoffmann, F., 2019, "HIP, RIP and the Robustness of Empirical Earnings Processes", *Quantitative Economics*, 10, 1279-1315.

Hsiao, C., 2018, "Panel models with interactive effects", *Journal of Econometrics*, 206(2), 645-673.

Jochmans, K., & M., Weidner, 2019, "Inference on a distribution from noisy draws", CEMMAP Working Paper 44/19.

Keane, M.P., and N., Wasi, 2016, "Labour supply: the roles of human capital and the extensive margin" *The Economic Journal*, 126(592), 578-617.

Kline, P., & Santos, A., 2013, Sensitivity to missing data assumptions: Theory and an evaluation of the US wage structure. *Quantitative Economics*, 4(2), 231-267.

Light, A., & Ureta, M. 1995, "Early-career work experience and gender wage differentials". *Journal of Labor Economics*, 13(1), 121-154.

Lise, J., & Postel-Vinay, F. (2020). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, 110(8), 2328-76.

Little, R. J., & Rubin, D. B., 2019. *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

Magnac, T., N., Pistolesi & S., Roux, 2018, “Post Schooling human capital investments and the life-cycle variance of earnings”, *Journal of Political Economy*, 126(3), 1219–1249.

Magnac, T., & S., Roux, 2021, “Heterogeneity and Wage Inequalities over the Life-Cycle”, *European Economic Review,* 134, 103715.

Meghir, C. & L., Pistaferri, 2010, “Earnings, consumption and life-cycle choices”, in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. 4b, Elsevier Science, Amsterdam, pp. 773-854.

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

Moon, H. R., & M., Weidner, 2017, "Dynamic linear panel regression models with interactive fixed effects", *Econometric Theory*, 33(1), 158-195.

Moon, H. R., & M., Weidner, 2018, "Nuclear norm regularized estimation of panel regression models", arXiv preprint arXiv:1810.10987.

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

Polachek, S., 1975, “Differences in Expected Post-School Investments as a Determinant of Market Wage Differentials”, *International Economic Review*, 16(2), 451-70.

Polachek, S., Das, T., & R., Thamma-Apiroam, 2015, “Heterogeneity in the Production of Human Capital”, *Journal of Political Economy*, 123(6), 1410-1455.

Rubinstein, Y., Y., & 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, pp. 1-67.

Sasaki, Y., 2015, "Heterogeneity and selection in dynamic panel data". *Journal of Econometrics*, 188(1), 236-249.

Song, M., 2013, "Asymptotic Theory for Dynamic Heterogenous Panels with Cross Section Dependence and Its Apllications", unpublished manuscript.

Taber, C., & Vejlin, R. (2020). Estimation of a Roy/search/compensating differential model of the labor market. *Econometrica*, 88(3), 1031-1069.

Table 1: Descriptive statistics on interruptions

Number of interruptions	Number of individuals	Proportion in interruption	Cumulated duration in interruption	Average number of interruptions
All	7004	0.154	3.7	1.44
0	1219	0.000	0.0	0.00
1	2279	0.110	2.6	1.00
2	1933	0.196	4.7	2.00
3	1050	0.261	6.3	3.00
4	383	0.321	7.7	4.00
5	118	0.355	8.5	5.00
6	22	0.378	9.3	6.00

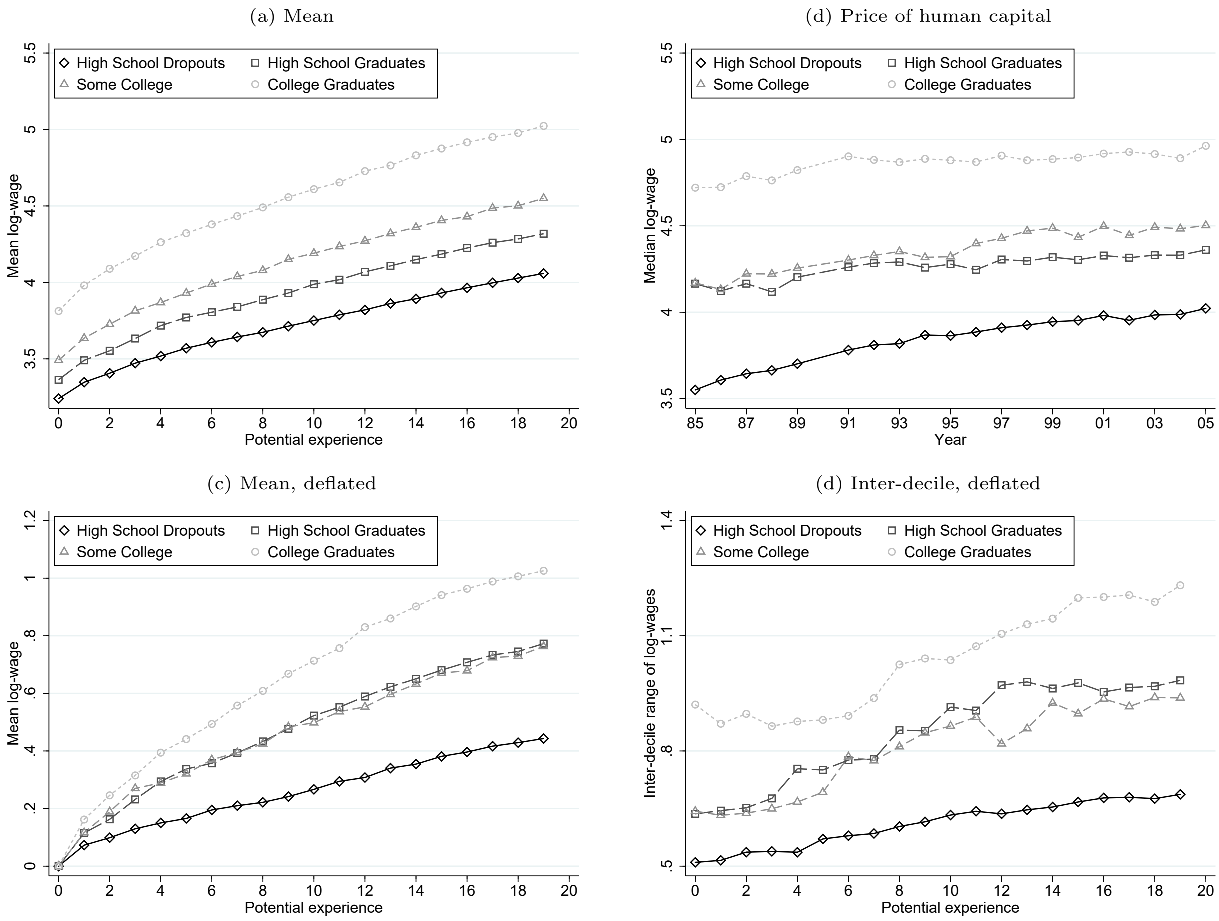
Note: For a given individual, observations after the last year employed in the private sector are treated as interruptions.

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

Number of factors	1	2	3
Criteria			
PC_{p1}	0.009936	0.006610	0.003839
PC_{p2}	0.012717	0.012172	0.012181
PC_{p3}	0.012714	0.012165	0.012172
IC_{p1}	-4.654	-5.015	-5.348
IC_{p2}	-4.307	-4.321	-4.306
IC_{p3}	-4.307	-4.322	-4.307
Quantities used to compute criteria			
N	7004	7004	7004
$T - DF$	16.000	16.000	16.000
$V(k.\widehat{\varphi}_t)$	0.011	0.009	0.008
$\widehat{\sigma}^2$	0.008	0.008	0.008
$C^2_{N(T-DF)}$	16.000	16.000	16.000

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 avergae 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 or 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 1: Mean and inter-decile of log-wages as a function of potential experience for individuals in our sample, by diploma



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. The price of human capital is measured with the median log-wage for individuals aged 50-55 in the whole population of individuals. In panels (c) and (d), "deflated" means that log-wages are deflated with prices of human capital.

Figure 2: Timing of the model

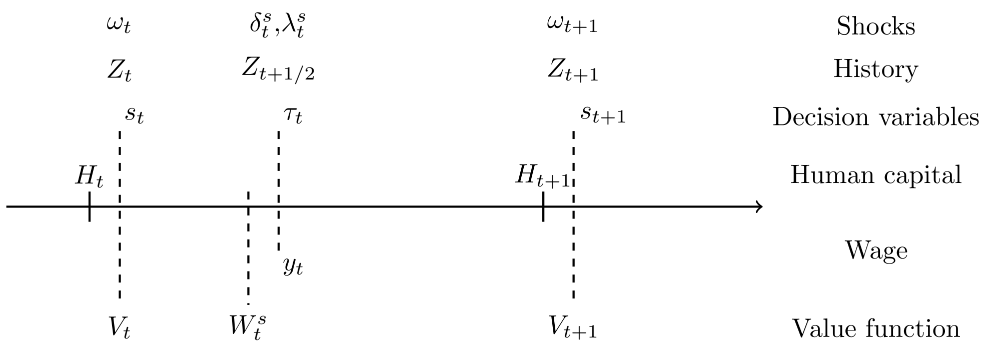
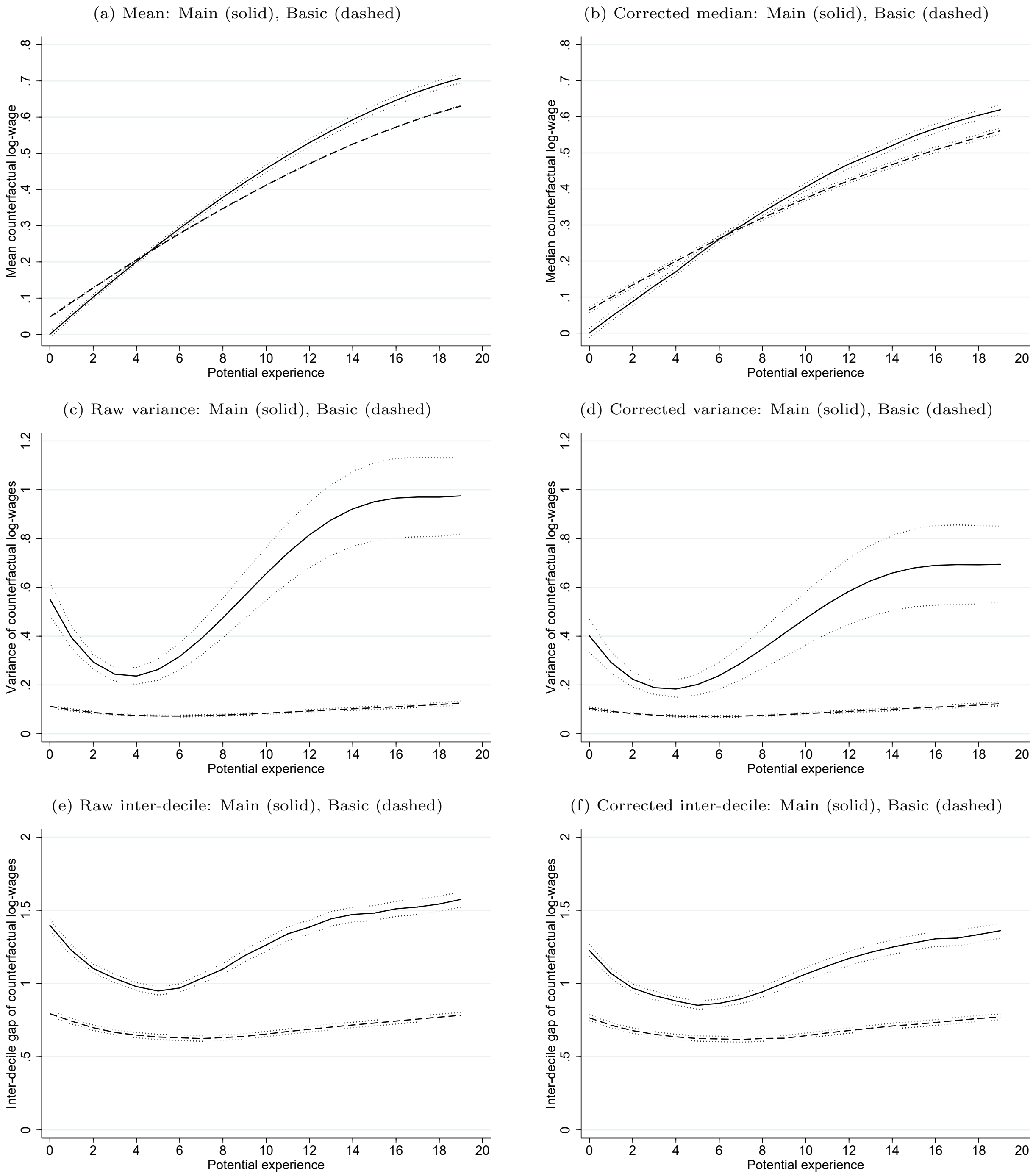
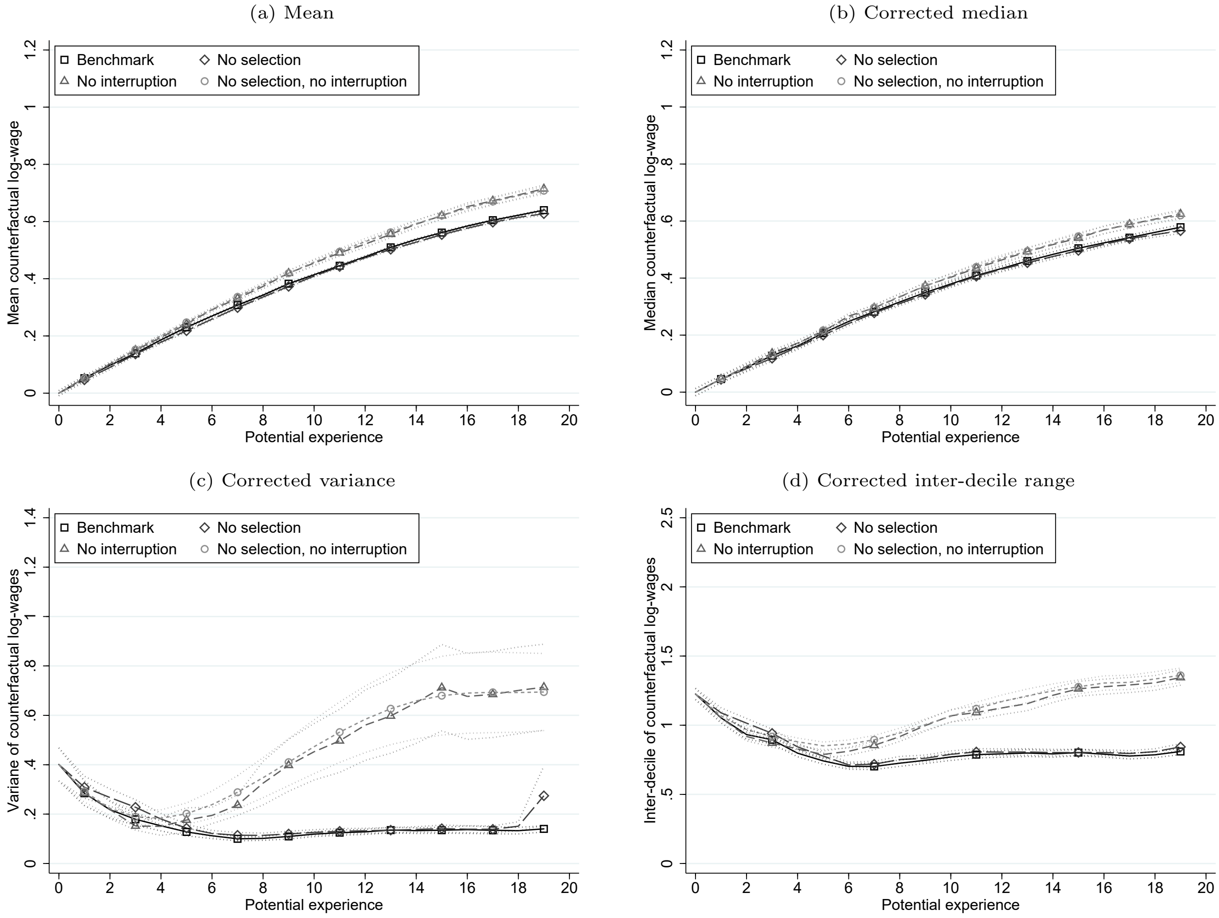


Figure 3: Mean, median, variance and inter-decile range of counterfactual log-wages as a function of potential experience, main and basic specifications



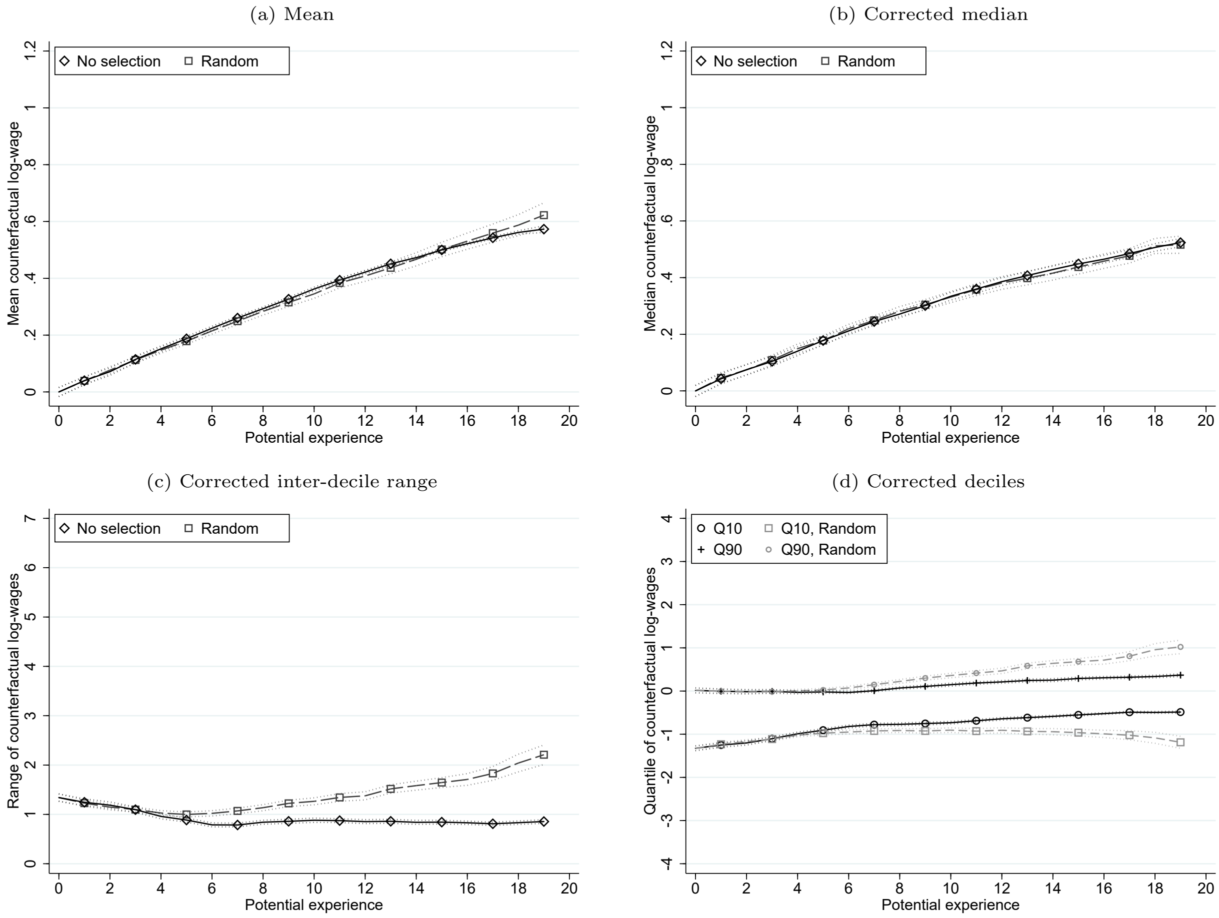
Note: "Corrected" statistics are obtained after bias correction as described in the Online Appendix. "Main": main specification that includes variables x_{i1}, x_{i2}, x_{i3} and x_{i4} as well as the additive individual effect and two interactive factors; "Basic": specification that includes only variables x_{i1} and x_{i2}, and the additive individual effect. In panels (a) and (b), the levels of mean and corrected median counterfactual log-wages are normalized using the value at period zero of the main specification.

Figure 4: Mean, median, variance and inter-decile range of counterfactual log-wage as a function of potential experience, counterfactual scenarii 1-4



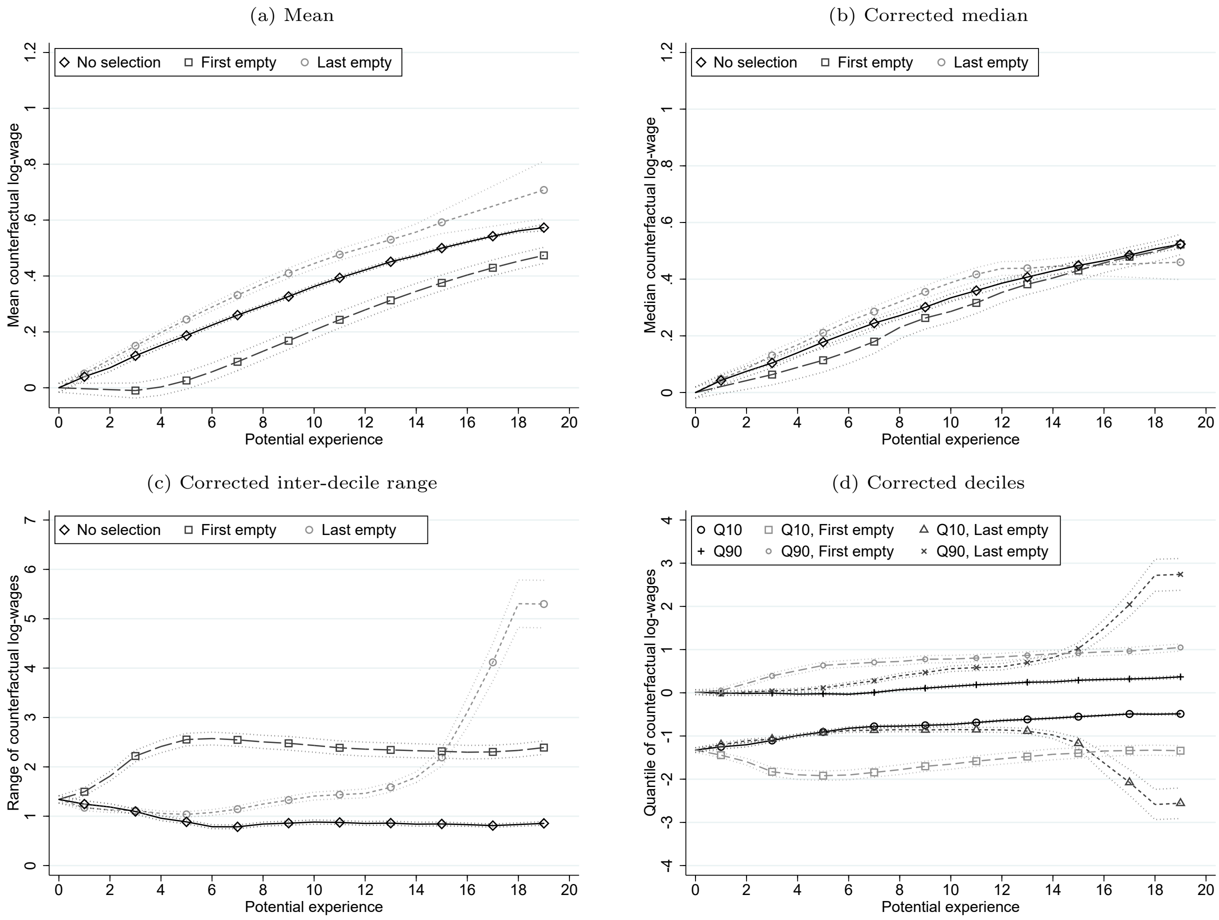
Note: "Corrected" statistics are obtained after bias correction as described in the Online Appendix. The levels of mean and corrected median counterfactual log-wages are normalized using the value at period zero of the benchmark specification.

Figure 5: Mean, median, deciles and inter-decile range of counterfactual log-wages as a function of potential experience, counterfactual scenario 5 (non-employment in random years), sample of individuals with two interruptions or more



Note: "Corrected" statistics are obtained after bias correction as described in the Online Appendix. In panels (a) and (b), the levels of mean and corrected median counterfactual log-wages are normalized using the value at period zero of the "No Selection" specification. In the "Random" case, non-employment years are drawn randomly in the period limited by the last year that an individual is observed.)

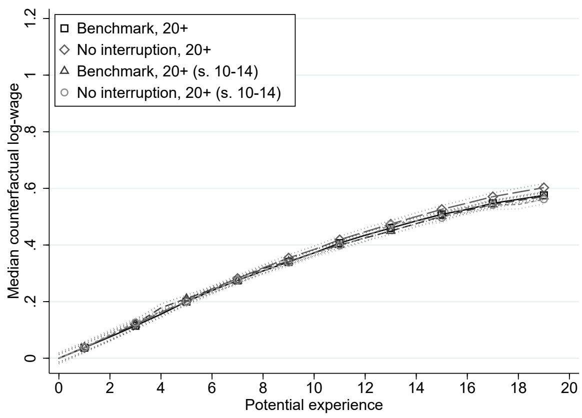
Figure 6: Mean, median, deciles and inter-decile range of counterfactual log-wages as a function of potential experience, counterfactual scenario 5 (non-employment in the first or last years), sample of individuals with two interruptions or more



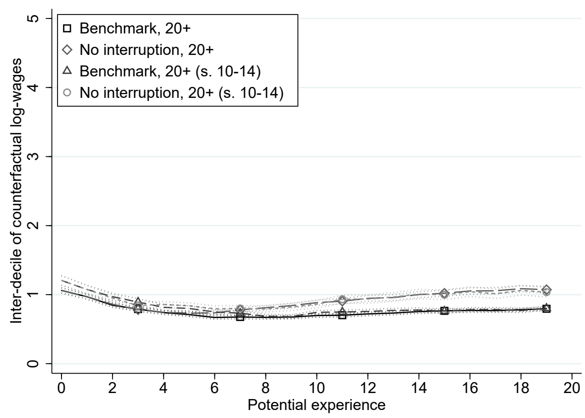
Note: "Corrected" statistics are obtained after bias correction as described in the Online Appendix. In panels (a) and (b), the levels of mean and corrected median counterfactual log-wages are normalized using the value at period zero of the "No Selection" specification. "First empty" corresponds to the case where all non-employment years are assigned to the first years of observations (except the very first one). "Last empty" corresponds to the case where non-employment years are assigned to the last years of the period during which the individual is observed.

Figure 7: Corrected median and inter-decile range of counterfactual log-wages as a function of potential experience, counterfactual scenarii 1 and 3 (benchmark and no interruption), individuals in employment 10-14 or 20 years and more (without or with sampling of employment years consistently with individuals in employment for 10-14 years)

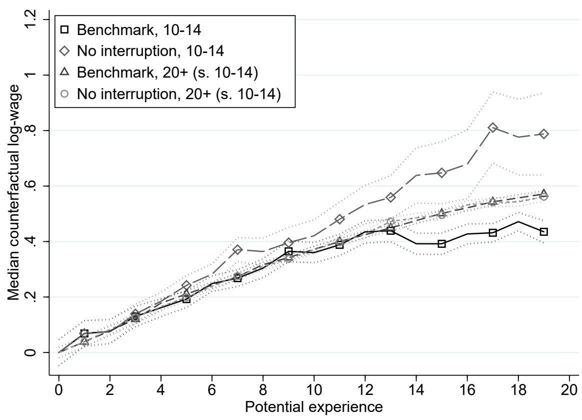
(a) Corrected median, 20+ and 20+ (sampling 10-14)



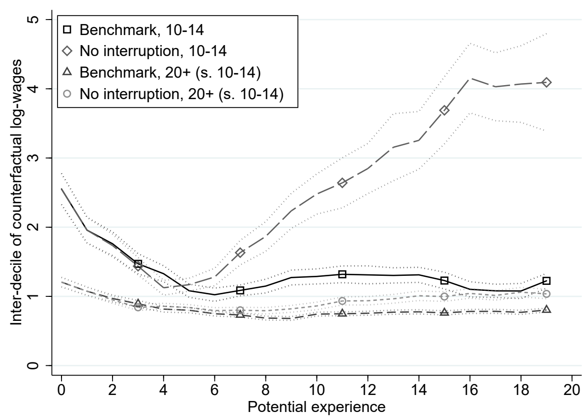
(b) Corrected inter-decile, 20+ and 20+ (sampling 10-14)



(c) Corrected median, 10-14 and 20+ (sampling 10-14)



(d) Corrected inter-decile, 10-14 and 20+ (sampling 10-14)



Note: "Corrected" statistics are obtained after bias correction as described in the Online Appendix. In panels (a) and (c), the levels of corrected median counterfactual log-wages are normalized using the value at period zero of the benchmark case for the considered subsample ("10-14": individuals in employment 10-14 years, ' '20+": 20 years and more, or "20+ (s. 10-14)": 20 years and more after sampling 10-14 years according to the distribution of employment years for individuals employed 10-14 years).

APPENDIX

A Data appendix

A.1 Data construction

In the raw data, there are $4,884,767$ person-job-year observations in the public and private sector over the 1976-2012 period corresponding to individuals born in the first four days of October. For individuals born an odd year, there is no information before 2002. When restricting the sample to males, we are left with $2,658,470$ observations. For consistency across time, we restrict our attention to individuals born on even years over the whole period, and this makes the sample size drop to $2,017,624$ observations. When considering only jobs in the private sector, we are left with $1,772,511$ observations, and when considering only full-time positions, the sample size decreases to $1,520,615$ observations. We also delete jobs for workers on a training period and apprentices, and this leaves us with $1,492,091$ observations. Once jobs are aggregated per individual-year, we end up with $1,365,837$ observations. We ignore overlaps of job spans because they are exceptional for full-time jobs.

We then restrict the sample to jobs such that 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 (ie. 39 hours per week). After 1998, some firms change their number of working hours to 151.67 (ie. 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.[A.1] 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 $1,354,104$ observations.

We keep only individual-year observations such that the total amount of working days is larger than 6 months, and the sample then includes $1,192,377$ observations corresponding to $102,425$ males. We keep only observations for individuals entering the labor market over the 1985-1992 period (ie. individuals observed for the first time in the panel during that period), and we are left with $200,756$ observations corresponding to $15,039$ are males. After restricting the sample to individuals aged $16-30$, our sample includes $178,111$ observations corresponding to $12,216$ males. We delete individuals for whom the education level is missing (4 individuals) and this leaves us with $178,098$ observations corresponding to $12,212$ male individuals. Finally, we keep individuals who were present at least 15 years, which leaves us with $7,004$ individuals with $137,315$ observations.

The education level is defined as the highest diploma obtained by individuals. Using French diploma names, high-school drop-outs includes no diploma, CAP, BEPC and CEP; high-school graduates includes 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

[A.1] We use as proportions for every year over the 1999-2006 period: 10%, 20%, 30%, 40%, 60%, 70%, 80% and 90%.

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.

A.2 Flat spots and homogenous Mincer equations

The method of flat spots achieves the separation between human capital stocks and prices (Figure 1(b)). In contrast with the US (Bowlus and Robinson, 2012), the increase in human capital prices is lower for high skill groups (about 45% for college graduates and about 50% for some college) than for the low skill group (about 80%) while price increases by only 30% for high school graduates. These prices are nominal, and the INSEE Consumer Price Index over the period increases by about 65%.

As a descriptive device, we ran homogeneous Mincer regressions with and without correction for 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 A.2. Coefficient estimates of the interruption variables ($x^{(3)}$ and $x^{(4)}$) are significant and negative even when the selection correction term is introduced, and this selection is not significant. 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 indicates that returns to human capital investments are lower when outside the private sector; (ii) interruptions move the Mincer dip to a lower value of potential experience; (iii) the effect of current-period selection is weak.

A.3 Counterfactuals

First, because our manipulations of potential histories to construct counterfactuals are related to potential and actual experience impacts only, we harmonize them across the cohorts that enter the private sector from 1985 to 1992. We change the timeline for each cohort to make them start artificially in 1985 as for the first cohort we consider. Parameters are re-scaled according to structural formulas (B.11)-(B.15) and verify:

$$(\widetilde{\eta}_{i0}, \widetilde{\eta}_{i1}, \widetilde{\eta}_{i2}, \widetilde{\eta}_{i3}, \widetilde{\eta}_{i4}) = (\eta_{i0} + \eta_{i1}(t_{0i} - 1), \eta_{i1}, \eta_{i2}\beta^{-(t_{0i}-1)}, \eta_{i3}, \eta_{i4}\beta^{-(t_{0i}-1)}).$$

Second, in the counterfactuals developed in Section **5.3.2**, either the first year or the last year out of the private sector is replaced by a year in the private sector. In the former case, the counterfactual first date of interruption is given by $t_1^c = t_1 + 1$ and counterfactual variables denoted $\mathsf{x}_t^{(j)}$ for $j = 3$ or 4 are defined as:

$$\begin{aligned}
\mathsf{x}_t^{(3)} &= 0 \text{ and } \mathsf{x}_t^{(4)} = 0 \text{ for } t \leqslant t_1^c, \\
\mathsf{x}_t^{(3)} &= x_t^{(3)} - 1 \text{ and } \mathsf{x}_t^{(4)} = \sum_{k=1}^{K_t - 1} \left(\beta^{-t_{2k+2}} - \beta^{-t_{2k+1}}\right) + \left(\beta^{-t_2} - \beta^{-t_1^c}\right) \text{ for } t > t_1^c.
\end{aligned}$$

After the counterfactual first date out of the private sector, the duration of interruptions is one year lower than in the benchmark situation at all dates, and the curvature variable ($\mathsf{x}_t^{(4)}$) is modified.

In the case where the last year out of the private sector is replaced by a year in, the counterfactual last date of entry into the private sector is given by $t_{2K_t+2}^c = t_{2K_t+2} - 1$ and we have:

$$
\begin{aligned}
\mathsf{x}_t^{(3)} &= x_t^{(3)} \text{ and } \mathsf{x}_t^{(4)} = x_t^{(4)} \text{ for } t \leqslant t^c_{2K_t+2} \\
\mathsf{x}_t^{(3)} &= x_t^{(3)} - 1 \text{ and } \mathsf{x}_t^{(4)} = \left(\beta^{-t^c_{2K_t+2}} - \beta^{-t_{2K_t-1}}\right) + \sum_{k=0}^{K_t-2} \left(\beta^{-t_{2k+2}} - \beta^{-t_{2k+1}}\right) \text{ for } t > t^c_{2K_t+2}
\end{aligned}
$$

The career interruption and curvature terms are the same as in the benchmark case until the counterfactual last date of entry into the private sector. Afterwards, the duration of interruptions is one year lower than in the benchmark situation and the curvature term is modified by the use of the counterfactual last date of entry into the private sector instead of the observed one.

B Proofs of Section 4.2

B.1 Proof of Proposition 1

Consider an individual who evaluates the consequences of working in sector s and choosing human capital investments, τ_t^s, 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 with respect to the level of human capital. Using the envelope theorem if τ_t^s is an interior solution, or replacing with the corner solution, $\tau_t^s = 0$, we have that for any H_t:

$$
\begin{aligned}
\frac{\partial W_t^s}{\partial H_t} &= \frac{1}{H_t} + \beta \left\{ \exp\left(\rho^s \tau_t^s - \lambda_t^s\right) \mathbb{E}_{t+1/2} \left[\frac{\partial V_{t+1}}{\partial H_{t+1}} \right] \right\} \\
&= \frac{1}{H_t} + \frac{H_{t+1}}{H_t} \beta \mathbb{E}_{t+1/2} \left[\frac{\partial V_{t+1}}{\partial H_{t+1}} \right]
\end{aligned}
$$

since we have $\frac{H_{t+1}}{H_t} = \exp\left(\rho^s \tau_t^s - \lambda_t^s\right)$. This expression is equivalent to:

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

and implies that:

$$
H_t \mathbb{E}_t \frac{\partial W_t^s}{\partial H_t} = 1 + \beta \mathbb{E}_t \left[H_{t+1} \frac{\partial V_{t+1}}{\partial H_{t+1}} \right]. \tag{B.1}
$$

This shows that derivatives do not depend on s i.e. $E_t \frac{\partial W_t^e}{\partial H_t} = E_t \frac{\partial W_t^n}{\partial H_t}$ and this proves that:

$$
H_t \frac{\partial V_t}{\partial H_t} = H_t \frac{\partial}{\partial H_t} \left(\max(\mathbb{E}_t W_t^e + \omega_t, \mathbb{E}_t W_t^n) \right) = 1 + \beta \mathbb{E}_t \left[H_{t+1} \frac{\partial V_{t+1}}{\partial H_{t+1}} \right] \tag{B.2}
$$

For $t = t_0 + d + 1$, specification (6) writes:

$$
\frac{\partial V_{t_0+d+1}}{\partial H_{t_0+d+1}} = \frac{\kappa}{H_{t_0+d+1}} \Longrightarrow H_{t_0+d+1} \frac{\partial V_{t_0+d+1}}{\partial H_{t_0+d+1}} = \kappa. \tag{B.3}
$$

Denote:

$$
\kappa_t = H_t \frac{\partial V_t}{\partial H_t} \tag{B.4}
$$

By backward induction, using equation (B.2) and the initial condition (B.3), all values κ_t are deterministic, that is, independent of Z_t. We obtain that:

$$\kappa_t = 1 + \beta\kappa_{t+1} \Longrightarrow \kappa_t - \frac{1}{1-\beta} = \beta(\kappa_{t+1} - \frac{1}{1-\beta}) \tag{B.5}$$

so that by backward induction:

$$\kappa_t = \frac{1}{1-\beta} + \beta^{t_0+d+1-t}(\kappa - \frac{1}{1-\beta}). \tag{B.6}$$

By integration of equations (B.1) and (B.4), we obtain the value functions of the Proposition in which the arbitrary constants of integration, $a_t^s(Z_{t+1/2})$ and $a_t(Z_t)$ are further defined below. ■

B.2 Proof of Proposition 2

The first order condition of the maximization problem for $t \in [t_0, t_0+d]$ with respect to the level of investment τ_t is

$$-(1 + c\tau_t^s) + \beta\rho^s\mathbb{E}_{t+1/2}\left[H_{t+1}\frac{\partial V_{t+1}}{\partial H_{t+1}}\right] = 0, \tag{B.7}$$

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

$$\beta\rho^s\mathbb{E}_{t+1/2}\left[H_{t+1}\frac{\partial V_{t+1}}{\partial H_{t+1}}\right] > 1. \tag{B.8}$$

Using equation (B.4), this condition is equivalent to $\beta\rho^s\kappa_{t+1} > 1$ and equation (B.7) yields the optimal investment which verifies:

$$(1 + c\tau_t^s) = \beta\rho^s\kappa_{t+1}, \tag{B.9}$$

and the second term in equation (7) follows. When $\beta\rho^s\kappa_{t+1} \leq 1$, we obtain that $\tau_t^s = 0$. Furthermore, as the second left hand side term in (B.7) is constant, the second order condition is satisfied if and only if $c > 0$.

B.3 Proof of Proposition 3

Using Proposition 1:

$$\begin{aligned}
W_t^s(H_t, Z_{t+1/2}) &= \delta_t^s + \ln H_t - \left(\tau_t^s + c\frac{(\tau_t^s)^2}{2}\right) + \beta\mathbb{E}_{t+1/2}\left[V_{t+1}\right] \\
&= \delta_t^s + \ln H_t - \left(\tau_t^s + c\frac{(\tau_t^s)^2}{2}\right) + \beta\mathbb{E}_{t+1/2}\left[a_{t+1}(Z_{t+1}) + \kappa_{t+1}\log H_{t+1}\right] \\
&= \delta_t^s + \ln H_t - \left(\tau_t^s + c\frac{(\tau_t^s)^2}{2}\right) + \beta\mathbb{E}_{t+1/2}\left[a_{t+1}(Z_{t+1}) + \kappa_{t+1}(\ln H_t + \rho^s\tau_t^s - \lambda_t^s)\right].
\end{aligned}$$

By identifying constant terms and using **equation** (B.9) and Proposition **2**, we get:

$$\begin{aligned}
a_t^s(Z_{t+1/2}) &= \delta_t^s + \left(\beta\kappa_{t+1}\rho^s\tau_t^s - \tau_t^s - c\frac{(\tau_t^s)^2}{2}\right) - \beta\kappa_{t+1}\lambda_t^s + \beta\mathbb{E}_{t+1/2}\left[a_{t+1}(Z_{t+1})\right], \\
&= \delta_t^s + c\frac{(\tau_t^s)^2}{2} - \beta\kappa_{t+1}\lambda_t^s + \beta\mathbb{E}_{t+1/2}\left[a_{t+1}(Z_{t+1})\right].
\end{aligned}$$

B.4 Proof of Proposition 4

By equation (5) we have:

$$
\begin{aligned}
& \omega_t + \mathbb{E}_t\left[\delta_t^e + c\frac{(\tau_t^e)^2}{2} - \beta\kappa_{t+1}\lambda_t^e\right] + \beta\mathbb{E}_t\left[a_{t+1}(Z_{t+1})\right] + \kappa_t\log(H_t) \\
\geq\ & \mathbb{E}_t\left[\delta_t^n + c\frac{(\tau_t^n)^2}{2} - \beta\kappa_{t+1}\lambda_t^n\right] + \beta\mathbb{E}_t\left[a_{t+1}(Z_{t+1})\right] + \kappa_t\log(H_t).
\end{aligned}
$$

and we note that neither initial conditions H_t nor terminal conditions $E_t\left[a_{t+1}(Z_{t+1})\right]$ depend on current sector choice (absent any transition costs) and we obtain condition (8). It also yields:

$$
a_t(Z_t) = \max(\omega_t + \mathbb{E}_t(\delta_t^s - \beta\kappa_{t+1}\lambda_t^s + c\frac{(\tau_t^s)^2}{2}), \mathbb{E}_t(\delta_t^n - \beta\kappa_{t+1}\lambda_t^n + c\frac{(\tau_t^n)^2}{2})) + \beta\mathbb{E}_t\left[a_{t+1}(Z_{t+1})\right].
$$

B.5 Proof of Proposition 5

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

$$
\begin{aligned}
H_t &= H_{t_{2K_t}} \exp\left[\sum_{l=t_{2K_t}}^{t-1} \rho^e\tau_l^e - \sum_{l=t_{2K_t}}^{t-1} \lambda_l^e\right] \\
&= H_{t_{2K_{t-1}}} \exp\left[\sum_{l=t_{2K_t}}^{t-1} \rho^e\tau_l^e - \sum_{l=t_{2K_t}}^{t-1} \lambda_l^e + \sum_{l=t_{2K_{t-1}}}^{t_{2K_t}-1} \rho^n\tau_l^n - \sum_{l=t_{2K_{t-1}}}^{t_{2K_t}-1} \lambda_l^n\right] \\
& \dots \\
&= H_{t_0} \exp\left[\sum_{l=t_0}^{t-1} \rho^{s_l}\tau_l^{s_l} - \sum_{l=t_0}^{t-1} \lambda_l^{s_l}\right]
\end{aligned}
$$

At each date, we have that:

$$
\tau^{s_l} = \max\left\{0, \frac{1}{c}\left(\rho^{s_l}\beta\kappa_{l+1} - 1\right)\right\}
$$

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

$$
\begin{aligned}
\ln H_t &= \ln H_{t_0} - \sum_{l=t_0}^{t-1} \lambda_l^{s_l} + \sum_{l=t_0}^{t-1} \rho^{s_l}\tau_l^{s_l} \\
&= \ln H_{t_0} - \sum_{l=t_0}^{t-1} \lambda_l^{s_l} + \sum_{l=t_0}^{t-1} \frac{\rho^{s_l}}{c}\left(\rho^{s_l}\beta\kappa_{l+1} - 1\right)
\end{aligned}
$$

Using the sequence of periods in every sector and replacing κ_{l+1} by its expression $\kappa_{l+1} = \frac{1}{1-\beta} + \beta^{t_0+d-l}(\kappa - \frac{1}{1-\beta})$ (see Proposition 1), the term $\sum_{l=t_0}^{t-1} \rho^{s_l} \frac{1}{c} \left(\rho^{s_l}\beta\kappa_{l+1} - 1\right)$ can be decomposed into:

$$
\begin{aligned}
\sum_{l=t_0}^{t-1} \frac{\rho^{s(l)}}{c}\left(\rho^{s(l)}\beta\kappa_{l+1} - 1\right) \quad = \quad & \sum_{k=0}^{K_t-1}\sum_{l=t_{2k}}^{t_{2k+1}-1} \frac{\rho^e}{c}\left(\rho^e\frac{\beta}{1-\beta} - 1 + \rho^e\beta^{t_0+d+1-l}\left(\kappa - \frac{1}{1-\beta}\right)\right) \\
& + \sum_{l=t_{2K_t}}^{t-1} \frac{\rho^e}{c}\left(\rho^e\frac{\beta}{1-\beta} - 1 + \rho^e\beta^{t_0+d+1-l}\left(\kappa - \frac{1}{1-\beta}\right)\right) \\
& + \sum_{k=0}^{K_t-1}\sum_{l=t_{2k+1}}^{t_{2k+2}-1} \frac{\rho^n}{c}\left(\rho^n\frac{\beta}{1-\beta} - 1 + \rho^n\beta^{t_0+d+1-l}\left(\kappa - \frac{1}{1-\beta}\right)\right)
\end{aligned}
$$

where 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 she is in sector n. Since

$$
\begin{aligned}
& \sum_{l=s_0}^{s_1-1} \frac{\rho^e}{c}\left(\rho^e\frac{\beta}{1-\beta} - 1 + \rho^e\beta^{t_0+d+1-l}\left(\kappa - \frac{1}{1-\beta}\right)\right) \\
= \quad & \frac{\rho^e}{c}\left(\rho^e\frac{\beta}{1-\beta} - 1\right)(s_1 - s_0) + \frac{(\rho^e)^2}{c}\beta^{t_0+d+1-s_0}\left(\kappa - \frac{1}{1-\beta}\right)\sum_{l=0}^{s_1-s_0-1}\beta^{-l} \\
= \quad & \frac{\rho^e}{c}\left(\rho^e\frac{\beta}{1-\beta} - 1\right)(s_1 - s_0) + \frac{(\rho^e)^2}{c}\frac{\beta^{t_0+d+2}}{1-\beta}\left(\kappa - \frac{1}{1-\beta}\right)\left(\beta^{-s_1} - \beta^{-s_0}\right)
\end{aligned}
$$

the term $\sum_{l=t_0}^{t-1} \frac{\rho^{s_l}}{c}\left(\rho^{s_l}\beta\kappa_{l+1} - 1\right)$ simplifies into:

$$
\begin{aligned}
& \frac{\rho^e}{c}\left(\rho^e\frac{\beta}{1-\beta} - 1\right)\sum_{k=0}^{K_t-1}(t_{2k+1} - t_{2k}) + \frac{(\rho^e)^2}{c}\left(\kappa - \frac{1}{1-\beta}\right)\frac{\beta^{t_0+d+2}}{1-\beta}\sum_{k=0}^{K_t-1}\left(\beta^{-t_{2k+1}} - \beta^{-t_{2k}}\right) \\
& + \frac{\rho^e}{c}\left(\rho^e\frac{\beta}{1-\beta} - 1\right)(t - t_{2K_t}) + \frac{(\rho^e)^2}{c}\left(\kappa - \frac{1}{1-\beta}\right)\frac{\beta^{t_0+d+2}}{1-\beta}\left(\beta^{-t} - \beta^{-t_{2K_t}}\right) \\
& + \frac{\rho^n}{c}\left(\rho^n\frac{\beta}{1-\beta} - 1\right)\sum_{k=0}^{K_t-1}(t_{2k+2} - t_{2k+1}) + \frac{(\rho^n)^2}{c}\left(\kappa - \frac{1}{1-\beta}\right)\frac{\beta^{t_0+d+2}}{1-\beta}\sum_{k=0}^{K_t-1}\left(\beta^{-t_{2k+2}} - \beta^{-t_{2k+1}}\right)
\end{aligned}
$$

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:

$$\begin{aligned}
& \sum_{l=t_0}^{t-1} \frac{\rho^{s_l}}{c} \left(\rho^{s_l} \beta \kappa_{l+1} - 1\right) \\
= & \frac{\rho^e}{c} \left(\rho^e \frac{\beta}{1-\beta} - 1\right) \left\{ \sum_{k=0}^{K_t - 1} \left[(t_{2k+1} - t_{2k}) + (t_{2k+2} - t_{2k+1})\right] + (t - t_{2K_t}) \right\} \\
& + \left[\frac{\rho^n}{c} \left(\rho^n \frac{\beta}{1-\beta} - 1\right) - \frac{\rho^e}{c} \left(\rho^e \frac{\beta}{1-\beta} - 1\right) \right] \sum_{k=0}^{K_t - 1} (t_{2k+2} - t_{2k+1}) \\
& + \frac{(\rho^e)^2}{c} \frac{\beta^{t_0+d+2}}{1-\beta} \left(\kappa - \frac{1}{1-\beta}\right) \left\{ \sum_{k=0}^{K_t-1} \left[\beta^{-t_{2k+1}} - \beta^{-t_{2k}} + \beta^{-t_{2k+2}} - \beta^{-t_{2k+1}}\right] + \beta^{-t} - \beta^{-t_{2K_t}} \right\} \\
& + \frac{(\rho^n)^2 - (\rho^e)^2}{c} \frac{\beta^{t_0+d+2}}{1-\beta} \left(\kappa - \frac{1}{1-\beta}\right) \sum_{k=0}^{K_t-1} \left(\beta^{-t_{2k+2}} - \beta^{-t_{2k+1}}\right) \\
= & \frac{\rho^e}{c} \left(\rho^e \frac{\beta}{1-\beta} - 1\right) (t - t_0) + \frac{(\rho^e)^2}{c} \frac{\beta^{t_0+d+2}}{1-\beta} \left(\kappa - \frac{1}{1-\beta}\right) \left(\beta^{-t} - \beta^{-t_0}\right) \\
& + \left[\frac{\rho^n}{c} \left(\rho^n \frac{\beta}{1-\beta} - 1\right) - \frac{\rho^e}{c} \left(\rho^e \frac{\beta}{1-\beta} - 1\right) \right] \sum_{k=0}^{K_t - 1} (t_{2k+2} - t_{2k+1}) \\
& + \frac{(\rho^n)^2 - (\rho^e)^2}{c} \frac{\beta^{t_0+d+2}}{1-\beta} \left(\kappa - \frac{1}{1-\beta}\right) \sum_{k=0}^{K_t-1} \left(\beta^{-t_{2k+2}} - \beta^{-t_{2k+1}}\right)
\end{aligned}$$

Defining

$$\begin{aligned}
x_t^{(3)} &= \sum_{k=0}^{K_t-1} (t_{2k+2} - t_{2k+1}) \\
x_t^{(4)} &= \sum_{k=0}^{K_t-1} \left(\beta^{-t_{2k+2}} - \beta^{-t_{2k+1}}\right)
\end{aligned} \tag{B.10}$$

and

$$\begin{aligned}
\eta_3 &= \frac{\rho^n}{c} \left(\rho^n \frac{\beta}{1-\beta} - 1\right) - \frac{\rho^e}{c} \left(\rho^e \frac{\beta}{1-\beta} - 1\right) \\
\eta_4 &= \frac{1}{c} \frac{\beta^{T+2}}{1-\beta} \left(\kappa - \frac{1}{1-\beta}\right) \left((\rho^n)^2 - (\rho^e)^2\right)
\end{aligned}$$

Human capital at date t has the following expression:

$$\begin{aligned}
\ln H_t = & \ln H_{t_0} - \sum_{l=t_0}^{t-1} \lambda_l^{s_l} + \eta_3 x_t^{(3)} + \eta_4 x_t^{(4)} \\
& + \frac{\rho^e}{c} \left(\rho^e \frac{\beta}{1-\beta} - 1\right) (t - t_0) \\
& + \frac{(\rho^e)^2}{c(1-\beta)} \beta^{t_0+d+2} \left(\kappa - \frac{1}{1-\beta}\right) \left(\beta^{-t} - \beta^{-t_0}\right)
\end{aligned}$$

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

$$
\begin{aligned}
\ln y_t &= \delta_t + \ln H_t - \tau_t \\
&= \delta_t + \ln H_{t_0} - \sum_{l=t_0}^{t-1} \lambda_l^{s_l} + \eta_3 x_t^{(3)} + \eta_4 x_t^{(4)} \\
&\quad + \frac{\rho^e}{c}\left(\rho^e \frac{\beta}{1-\beta} - 1\right)(t-t_0) + \frac{(\rho^e)^2}{c\,(1-\beta)}\beta^{t_0+d+2}\left(\kappa - \frac{1}{1-\beta}\right)\left(\beta^{-t} - \beta^{-t_0}\right) \\
&\quad - \frac{1}{c}\left(\frac{\rho^e\beta}{1-\beta} + \rho^e\beta^{t_0+d+1-t}(\kappa - \frac{1}{1-\beta}) - 1\right) \\
&= \ln H_{t_0} - \frac{\rho^e t_0 + 1}{c}\left(\frac{\rho^e\beta}{1-\beta} - 1\right) - \frac{(\rho^e)^2\beta}{c}\frac{\beta^{d+1}}{1-\beta}\left(\kappa - \frac{1}{1-\beta}\right) \\
&\quad + \frac{\rho^e}{c}\left(\rho^e\frac{\beta}{1-\beta} - 1\right)t \\
&\quad + \frac{\rho^e}{c}\left(\frac{\rho^e\beta}{1-\beta} - 1\right)\beta^{t_0+d+1}\left(\kappa - \frac{1}{1-\beta}\right)\beta^{-t} \\
&\quad + \delta_t - \sum_{l=t_0}^{t-1}\lambda_l^{s_l} + \eta_3 x_t^{(3)} + \eta_4 x_t^{(4)}
\end{aligned}
$$

We can then set

$$\eta_0 = \ln H_{t_0} - \frac{\rho^e t_0 + 1}{c}\left(\frac{\rho^e\beta}{1-\beta} - 1\right) - \frac{(\rho^e)^2\beta}{c}\frac{\beta^{d+1}}{1-\beta}\left(\kappa - \frac{1}{1-\beta}\right) \tag{B.11}$$

$$\eta_1 = \frac{\rho^e}{c}\left(\rho^e\frac{\beta}{1-\beta} - 1\right) \tag{B.12}$$

$$\eta_2 = \beta^{t_0+d+1}\frac{\rho^e}{c}\left(\kappa - \frac{1}{1-\beta}\right)\left(\frac{\rho^e\beta}{1-\beta} - 1\right) \tag{B.13}$$

$$\eta_3 = \left(\frac{\rho^n}{c}\left(\frac{\rho^n\beta}{1-\beta} - 1\right) - \frac{\rho^e}{c}\left(\frac{\rho^e\beta}{1-\beta} - 1\right)\right) \tag{B.14}$$

$$\eta_4 = \frac{1}{c}\left((\rho^n)^2 - (\rho^e)^2\right)\left(\kappa - \frac{1}{1-\beta}\right)\frac{\beta^{t_0+d+1}}{1-\beta} \tag{B.15}$$

and obtain the reduced form **given by** equation (9).

C 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).

We can rewrite the equation of Song's page 74 (top of the page) as:

$$\underset{[K,1]}{\Delta_i} = \underset{[K,1]}{\xi_i^*} + \frac{1}{N}\sum_{j=1}^{N} a_{ij} S_{ii}^{-1/2} \underset{[K,K]}{S_{ij}} S_{jj}^{-1/2}\Delta_j + o_P(1) \tag{C.16}$$

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 Assumption B.ii and B.iii page 7 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 (C.16) with unknowns $\Delta = (\Delta_1, ., \Delta_N)$ that we can write as:

$$\Delta = \xi + \Gamma \Delta,$$

where

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

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

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

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 $\mathbb{E}\theta_i^* = 0$, $\mathbb{V}(\theta_i^*) = I_r$ 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 $\mathbb{V}\Xi_i < \infty$ since by Assumption A2i, we have that $\mathbb{E} \|\theta_i\|^4 < \infty$.

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

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

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:

$$\begin{aligned} \mu_{\max}(\Gamma) &\leq 2\mu_{\max}(C_1) - \mu_{\min}(C_d) \\ &\leq 2N^{-1} \max_{1 \leq i \leq N} (a_{ii}) = 2N^{-1} (\max_{1 \leq i \leq n} (\theta_i^{*\prime} \theta_i^*) + o_P(1)) \\ &= 2N^{-1} \max_{1 \leq i \leq N} (\|\theta_i^*\|^2) + o_P(N^{-1}) = o_P(N^{-3/4}). \end{aligned}$$

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 (C.16) leads to:

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

D Computational Appendix

D.1 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 $\ln y_i = (\ln y_{i1}, ..., \ln y_{iT})'$ and $x_i = (x_{i1}', ..., x_{iT}')'$ 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 criterium is: $C_2 \equiv \min\left(c_1, c_2\right)$ where:

$$c_1 = N\left(\overline{\theta}^{(k)} - \overline{\theta}^{(k-1)}\right)' V\left(\theta_i^{(k-1)}\right)^{-1}\left(\overline{\theta}^{(k)} - \overline{\theta}^{(k-1)}\right)$$

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:

$$c_2 = tr\left[\left(V\left(\theta_i^{(k)}\right) - V\left(\theta_i^{(k-1)}\right)\right)\left(V\left(\theta_i^{(k)}\right) - V\left(\theta_i^{(k-1)}\right)\right)'\right] / tr\left[V\left(\theta_i^{(k-1)}\right)\right]$$

using the fact that $tr\left[(A-B)'(A-B)\right]$ is a distance between matrices A and B, and dividing by $tr\left[V\left(\theta_i^{(\delta)}\right)\right]$ as a normalization. Our overall stopping rule is based on $C = \min\left(C_1, C_2\right)$ such that there is convergence when $C < 1e-4$.

D.2 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 Heyden and Morton (1996) (see also Dominitz and Sherman (2005) for a general framework).

The sum of squares we consider is given by:

$$C\left(\theta, \varphi, \eta\right) = \sum_{i,t|s_{it}=1}\left(\ln y_{it} - x_{it}\eta_i - \varphi_t\theta_i\right)^2 \tag{D.17}$$

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:[A.2]

$$\ln y_{it}^{(k)} = x_{it}\eta_i^{(k)} + \varphi_t^{(k-1)}\theta_i^{(k)} \tag{D.18}$$

[A.2] 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 $y_{it}^{(k)}$ and $\eta_i^{(k)}$ – that minimize the sum of squares:

$$\widetilde{C}\left(\theta, \varphi, \eta^{(k)}\right) = C\left(\theta, \varphi, \eta^{(k)}\right) + \sum_{i,t|s_{it}=n} \left(\ln y_{it}^{(k)} - x_{it}\eta_i^{(k)} - \varphi_t\theta_i\right)^2 \tag{D.19}$$

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 6

$$C\left(\widetilde{\theta}^{(k)}, \varphi^{(k)}, \eta^{(k)}\right) \leqslant C\left(\widetilde{\theta}^{(k-1)}, \varphi^{(k-1)}, \eta^{(k-1)}\right). \tag{D.20}$$

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

$$C\left(\theta^{(k)}, \varphi^{(k-1)}, \eta^{(k)}\right) \leqslant C\left(\widetilde{\theta}^{(k-1)}, \varphi^{(k-1)}, \eta^{(k-1)}\right), \tag{D.21}$$

since $\theta^{(k)}, \eta^{(k)}$ are minimizers of the left-hand side. Using the definition of $y_{it}^{(k)}$, we also have that

$$\widetilde{C}\left(\theta^{(k)}, \varphi^{(k-1)}, \eta^{(k)}\right) = C\left(\theta^{(k)}, \varphi^{(k-1)}, \eta^{(k)}\right), \tag{D.22}$$

since the sum of squares on the right hand side of equation (D.19) is equal to zero. The third stage of our algorithm yields, by minimization:

$$\widetilde{C}\left(\widetilde{\theta}^{(k)}, \varphi^{(k)}, \eta^{(k)}\right) \leqslant \widetilde{C}\left(\theta^{(k)}, \varphi^{(k-1)}, \eta^{(k)}\right). \tag{D.23}$$

and we get, using equations (D.23), (D.22) and (D.21) successively:

$$\begin{aligned} C\left(\widetilde{\theta}^{(k)}, \varphi^{(k)}, \eta^{(k)}\right) &\leqslant \widetilde{C}\left(\widetilde{\theta}^{(k)}, \varphi^{(k)}, \eta^{(k)}\right) && \text{(D.24)} \\ &\leqslant \widetilde{C}\left(\theta^{(k)}, \varphi^{(k-1)}, \eta^{(k)}\right) = C\left(\theta^{(k)}, \varphi^{(k-1)}, \eta^{(k)}\right) && \text{(D.25)} \\ &\leqslant C\left(\widetilde{\theta}^{(k-1)}, \varphi^{(k-1)}, \eta^{(k-1)}\right). && \text{(D.26)} \end{aligned}$$

■

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

Lemma 7

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

Proof. The left-hand side equality implies that:

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

according to equation (D.26). Using (D.19), this yields

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

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 (D.26), we also have that

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

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 (D.20), we have that

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

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 $C(\theta, \widehat{\varphi}, \widehat{\eta})$. We also have that $\widetilde{\theta}^{(k)}$ is the value that minimizes:

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

As $C(\theta, \widehat{\varphi}, \widehat{\eta})$ is minimum in $\widehat{\theta}$, and the second (positive) right-hand side term is positive but zero for $\theta = \widehat{\theta}$, then $\widetilde{C}(\theta, \widehat{\varphi}, \widehat{\eta})$ 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.

Table A.1: Correlation of log-wages deflated with prices of human capital at two values of potential experience

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1.00																			
2	0.83	1.00																		
3	0.75	0.85	1.00																	
4	0.70	0.79	0.87	1.00																
5	0.66	0.74	0.79	0.86	1.00															
6	0.63	0.70	0.75	0.81	0.87	1.00														
7	0.59	0.66	0.73	0.78	0.83	0.89	1.00													
8	0.55	0.62	0.68	0.73	0.79	0.84	0.90	1.00												
9	0.50	0.58	0.64	0.69	0.75	0.79	0.84	0.89	1.00											
10	0.46	0.54	0.59	0.65	0.70	0.75	0.78	0.83	0.88	1.00										
11	0.43	0.51	0.56	0.63	0.67	0.72	0.75	0.81	0.86	0.90	1.00									
12	0.42	0.50	0.55	0.60	0.65	0.69	0.73	0.78	0.82	0.86	0.90	1.00								
13	0.37	0.45	0.51	0.57	0.61	0.64	0.70	0.75	0.79	0.82	0.86	0.90	1.00							
14	0.37	0.43	0.50	0.56	0.60	0.65	0.69	0.73	0.77	0.80	0.83	0.87	0.91	1.00						
15	0.32	0.40	0.46	0.51	0.56	0.61	0.65	0.70	0.74	0.76	0.81	0.84	0.88	0.91	1.00					
16	0.31	0.39	0.45	0.50	0.55	0.60	0.63	0.68	0.72	0.75	0.78	0.82	0.85	0.89	0.91	1.00				
17	0.31	0.38	0.45	0.51	0.54	0.60	0.63	0.67	0.71	0.73	0.77	0.81	0.84	0.86	0.89	0.92	1.00			
18	0.29	0.36	0.43	0.50	0.53	0.57	0.62	0.66	0.69	0.71	0.74	0.79	0.81	0.84	0.87	0.90	0.93	1.00		
19	0.28	0.36	0.42	0.48	0.51	0.57	0.60	0.64	0.68	0.70	0.73	0.77	0.80	0.83	0.86	0.88	0.91	0.93	1.00	
20	0.28	0.35	0.40	0.48	0.50	0.56	0.59	0.64	0.67	0.69	0.72	0.76	0.79	0.81	0.83	0.85	0.89	0.91	0.94	1.00

Note: The first year of potential experience is the first year individuals are employed in the private sector and appear in our sample. x- and y- axes give values of potential experience.

Table A.2: Mincer regression in line with the theoretical model

	(1)	(2) 2^{nd} stage	(2) 1^{st} stage (probit)
$x_{it}^1 = t$	0.058***	0.058***	0.037***
	(0.001)	(0.001)	(0.005)
$x_{it}^2 = \beta^{-t}$	-0.295***	-0.289***	-1.511***
	(0.009)	(0.020)	(0.055)
x_{it}^3	-0.009***	-0.008*	-0.409***
	(0.002)	(0.005)	(0.009)
x_{it}^4	-0.178***	-0.208**	9.014***
	(0.027)	(0.099)	(0.115)
Married			-0.071***
			(0.017)
Marriage			0.006***
tenure			0.002
Having a			-0.010
child			(0.018)
Number of			-0.055***
children 3+			(0.012)
Number of			0.017**
children 18+			(0.009)
$\lambda(\hat{p}_{it}^*)$		-0.003	
		(0.008)	
Cohort fixed effects	X	X	X
N	138,447	138,447	158,194
R^2	0.234	0.234	

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).

Table A.3: Corrected inter-centile ranges and rank correlations of the effects, main specification

	Inter-decile	Rank correlation			
		Log-wage	Potential experience effect	Non-employment effect	Factors effect
Log-wage	0.876	1.000	0.579	0.000	0.057
Potential experience effect	1.275	0.579	1.000	-0.570	-0.211
Non-employment effect	0.674	0.000	-0.570	1.000	-0.048
Effect of factors	0.323	0.057	-0.211	-0.048	1.000

Note: "Potential experience effect": sum of all effects related to potential experience and the individual additive effect: $\eta_{i0} + \eta_{i1}t + \eta_{i2}\beta^{-t}$; "Non-employment effect": sum of all effects related to being absent from the panel: $\eta_{i3}x_{it}^{(3)} + \eta_{i4}x_{it}^{(4)}$; " Factors effect": .

Table A.4: Descriptive statistics on distributions of uncorrected and corrected parameters, estimation sample

	η_{i0}	η_{i1}	η_{i2}	η_{i3}	η_{i4}	δ_1	δ_2
Mean	-0.50	0.07	-0.36	-0.05	0.33	-0.03	-0.01
Variance	1.65	0.11	6.22	1.53	166.58	0.11	0.01
Corrected Variance	1.20	0.07	4.24	1.14	126.18	0.07	0.01
Q5	-2.04	-0.38	-3.72	-1.40	-11.17	-0.45	-0.14
Q10	-1.53	-0.20	-2.49	-0.75	-5.11	-0.29	-0.10
Q25	-0.94	-0.02	-1.09	-0.19	0.00	-0.12	-0.05
Median	-0.51	0.06	-0.29	0.00	0.00	-0.03	-0.01
Q75	-0.08	0.17	0.38	0.15	0.00	0.06	0.02
Q90	0.45	0.35	1.78	0.64	5.29	0.25	0.07
Q95	1.00	0.53	3.13	1.20	12.41	0.44	0.11
Corrected Q5	-1.76	-0.26	-2.99	-1.10	-8.15	-0.37	-0.12
Corrected Q10	-1.39	-0.14	-2.03	-0.58	-3.92	-0.24	-0.08
Corrected Q25	-0.87	-0.01	-0.94	-0.16	0.00	-0.11	-0.04
Corrected Median	-0.49	0.06	-0.29	0.00	0.00	-0.03	-0.01
Corrected Q75	-0.15	0.15	0.26	0.14	0.00	0.05	0.02
Corrected Q90	0.23	0.29	1.38	0.46	4.11	0.20	0.05
Corrected Q95	0.84	0.44	2.30	0.89	9.32	0.35	0.09
N	7004	7004	7004	7004	7004	7004	7004

Note: For individuals with fewer than 2 interruptions, parameters η_{i3} and η_{i4} are normalized as they are not identified, and this normalization contaminates descriptive statistics. "Corrected" statistics are obtained after bias correction as described in the Online Appendix.

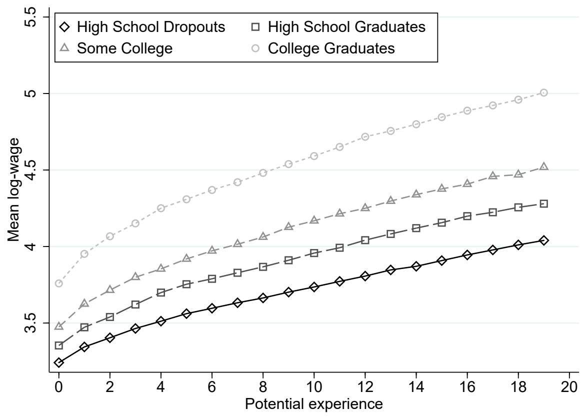
Table A.5: Descriptive statistics on distributions of uncorrected and corrected parameters, individuals with two interruptions and more in the estimation sample

	η_{i0}	η_{i1}	η_{i2}	η_{i3}	η_{i4}	δ_1	δ_2
Mean	-0.52	0.07	-0.32	-0.10	0.81	-0.03	-0.02
Variance	2.03	0.15	8.25	4.19	470.87	0.15	0.01
Corrected Variance	1.36	0.10	5.37	3.00	338.01	0.10	0.01
Q5	-2.25	-0.48	-4.19	-3.14	-27.91	-0.54	-0.17
Q10	-1.70	-0.26	-2.82	-1.89	-15.88	-0.35	-0.12
Q25	-1.05	-0.04	-1.18	-0.72	-5.77	-0.14	-0.06
Median	-0.52	0.06	-0.26	0.01	-0.12	-0.03	-0.01
Q75	0.00	0.18	0.54	0.64	6.04	0.08	0.03
Q90	0.63	0.40	2.15	1.64	19.35	0.31	0.08
Q95	1.26	0.59	3.86	2.69	30.27	0.50	0.14
Corrected Q5	-1.81	-0.36	-3.30	-2.80	-26.54	-0.43	-0.14
Corrected Q10	-1.51	-0.18	-2.29	-1.48	-12.14	-0.28	-0.10
Corrected Q25	-0.94	-0.03	-1.04	-0.58	-4.61	-0.13	-0.05
Corrected Median	-0.49	0.06	-0.26	0.01	-0.04	-0.03	-0.01
Corrected Q75	-0.13	0.16	0.39	0.48	4.97	0.07	0.02
Corrected Q90	0.36	0.33	1.66	1.26	15.08	0.27	0.07
Corrected Q95	0.98	0.49	2.98	2.24	25.60	0.39	0.11
N	1795	1795	1795	1795	1795	1795	1795

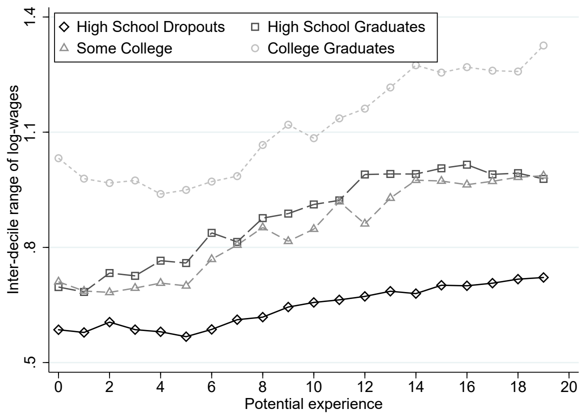
Note: "Corrected" statistics are obtained after bias correction as described in the Online Appendix.

Figure A.1: Mean and inter-decile of log-wages as a function of potential experience for all individuals entering the labour market between 1985 and 1992, by diploma

(a) Mean

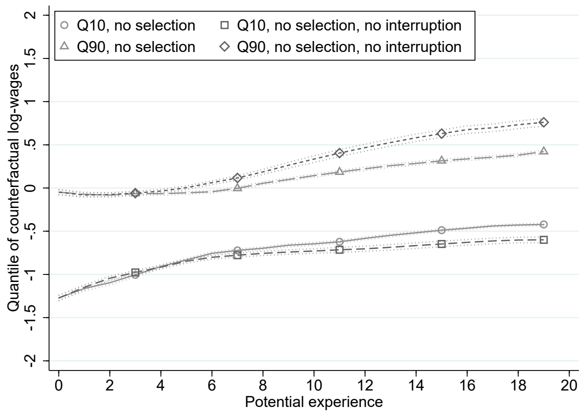


(b) Inter-decile



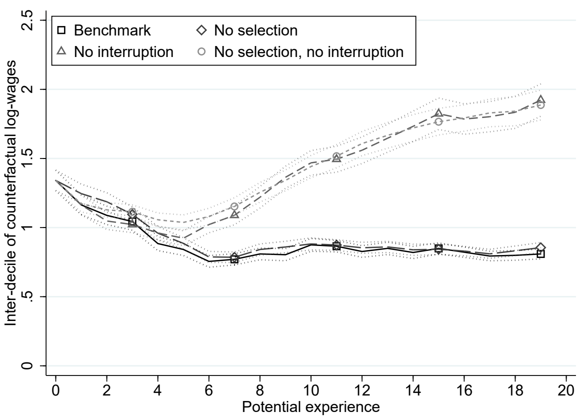
Note: Summary statistics presented here are computed using the sample of all individuals entering the labour market between 1985 and 1992.

Figure A.2: Corrected deciles of counterfactual log-wage as a function of potential experience, counterfactual scenarii 1-4



Note: "Corrected" statistics are obtained after bias correction as described in the Online Appendix. .

Figure A.3: Corrected inter-deciles of counterfactual log-wage as a function of potential experience, counterfactual scenarii 1-4, two interruptions and more



Note: "Corrected" statistics are obtained after bias correction as described in the Online Appendix.

Lifecycle Wages and Human Capital Investments: Selection and Missing Data Online Appendix

Laurent Gobillon*, Thierry Magnac†, Sébastien Roux‡

February 2, 2022

*Paris School of Economics-CNRS, 48 Boulevard Jourdan, 75014 Paris, email: laurent.gobillon@psemail.eu, webpage: http://laurent.gobillon.free.fr/.

†Toulouse School of Economics, Université de Toulouse Capitole

‡INSEE.

E Computation of counterfactual measures and their variances

In this appendix, we consider the specification of the paper:

$$\ln y_{it} = x_{it}\eta_i + \varphi_t\theta_i + \sigma\varepsilon_{it} \tag{S.1}$$

where we assume that terms ε_{it} are identically and independently distributed, and verify $E(\varepsilon_{it}) = E(\varepsilon_{it}^3) = 0$, and we denote $E(\varepsilon_{it}^2) = \sigma^2 < 0$ and $E(\varepsilon_{it}^4) = \varsigma/\sigma^4 < +\infty$. We focus on the unfeasible estimation when φ_t is observed since replacing φ_t by $\hat{\varphi}_t$ induces an additional but negligible bias.

E.1 Counterfactual means

E.1.1 Estimator

We want to estimate counterfactual means of the form $\mathcal{M}_t = E(X_t^c \mathsf{H} \,|\Theta_t^c)$ where H are random variables corresponding to individual effects, X_t^c are counterfactual random explanatory variables and Θ_t^c is a subset of individuals, for instance those who are employed at date t in a counterfactual situation. We do not observe individual effects but we can recover their unbiased OLS estimators which corresponding random variables are denoted $\widehat{\mathsf{H}}$. These estimators are obtained from the Frisch-Waugh theorem applied to equation (**??**) and verify:

$$\begin{aligned} \widehat{\eta}_i &= (x_i' M_{\varphi_i} x_i)^{-1} x_i' M_{\varphi_i} \ln y_i \\ &= \eta_i + \sigma (x_i' M_{\varphi_i} x_i)^{-1} x_i' M_{\varphi_i} \varepsilon_i \end{aligned} \tag{S.2}$$

where, for a given quantity q_{it}, we denote $q_i = \left(q'_{it_{i1}}, ..., q'_{it_{iT_i}}\right)'$ where t_{i1},t_{i2},...,t_{iT_i} are the dates at which individual i is employed where T_i is the number of periods individual i is employed, and $\varphi_i = \left(\varphi'_{t_{i1}}, ..., \varphi'_{t_{iT_i}}\right)'$. Our estimator of the counterfactual mean is given by:

$$\widehat{\mathcal{M}}_t = \frac{1}{N_t^c} \sum_{i \in \Theta_t^c} x_{it}^c \widehat{\eta}_i \tag{S.3}$$

with $N_t^c = Card\ \Theta_t^c$. Using equation (S.2), note that our estimator of the counterfactual mean (**??**) is consistent since we have $E(\widehat{\eta}_i \,|x_{it}^c, \eta_i) = \eta_i$. Indeed, we can apply the law of large numbers and we get:

$$\begin{aligned} \widehat{\mathcal{M}}_t &\to E\left(X_t^c \widehat{\mathsf{H}} \,|\Theta_t^c\right) = E\left[X_t^c E\left(\widehat{\mathsf{H}} \,|X_t^c, \mathsf{H}\right) |\Theta_t^c\right] \\ &= E(X_t^c \mathsf{H} \,|\Theta_t^c) = \mathcal{M}_t \end{aligned}$$

E.1.2 Variance of the estimator

We now turn to the computation of the variance of the mean estimator. We have:

$$\begin{aligned}
V\left(\widehat{\mathcal{M}}_t \left| x_t^c \right.\right) &= V\left(\frac{1}{N_t^c} \sum_{i \in \Theta_t^c} x_{it}^c \widehat{\eta}_i \left| x_t^c \right.\right) \\
&= \frac{1}{N_t^{c2}} \sum_{i,j \in \Theta_t^c} cov\left(x_{it}^c \widehat{\eta}_i, x_{jt}^c \widehat{\eta}_j \left| x_t^c \right.\right)
\end{aligned}$$

where $x_t^c = \left(x_{i_1 t}^{c\prime} ..., x_{i_{N_t^c} t}^{c\prime}\right)'$. Using equation (S.2), covariances in the right-hand side sum are given by:

$$cov\left(x_{it}^c \widehat{\eta}_i, x_{jt}^c \widehat{\eta}_j \left| x_t^c \right.\right) = E\left[x_{it}^c \left(x_i' M_{\varphi_i} x_i\right)^{-1} x_i' M_{\varphi_i} \varepsilon_i \varepsilon_j' M_{\varphi_j} x_j \left(x_j' M_{\varphi_i} x_j\right)^{-1} x_{jt}^{c\prime} \left| x_t^c \right.\right]$$

These covariances are zero for $i \neq j$ and we get:

$$V\left(\widehat{\mathcal{M}}_t \left| x_t^c \right.\right) = \frac{\sigma^2}{N_t^{c2}} \sum_{i \in \Theta_t^c} x_{it}^c \left(x_i' M_{\varphi_i} x_i\right)^{-1} x_i^{c\prime}$$

and an estimator of this variance is obtained by replacing φ_i with its estimator $\widehat{\varphi}_i$, and σ^2 with its unbiased OLS estimator verifying

$$\widehat{\sigma}^2 = \sum_i \widehat{\widehat{v}}_i' \widehat{\widehat{v}}_i / \left[\left(T_i - K_i - L\right) N\right] \tag{S.4}$$

where $\widehat{\widehat{v}}_i$ is the unfeasible OLS estimator of the residual, K_i is the number of identified parameters of individual variables (between 3 and 5) and L is the number of unobserved factors.

E.2 Bias-corrected counterfactual covariances and variances

E.2.1 Estimator

We show here how to compute the bias-corrected empirical counterpart of the covariance statistic $cov\left(X_{t1}^c \mathsf{H}, X_{t2}^c \mathsf{H} \left| S_t = e\right.\right)$ where X_{tj}^c are explanatory variables in counterfactual situation j. For readability, we ignore the conditioning with respect to $S_t = e$ but we explain how formula are modified when taking it into account. We also introduce the notations: $M_{tj} = X_t \mathsf{H}$ where X_t are the observed explanatory variables, $M_{tj} = X_{tj}^c \mathsf{H}$ and $\widehat{M}_{tj} = X_{tj}^c \widehat{\mathsf{H}}$ where $\widehat{\mathsf{H}}$ is an unbiased OLS estimator of H, and $m_{itj} = x_{itj}^c \eta_i$ and $\widehat{m}_{itj} = x_{itj}^c \widehat{\eta}_i$, where x_{itj}^c are realization of X_{tj}^c and $\widehat{\eta}_i$ is given by equation S.2.

We have $cov\left(M_{t1}, M_{t2}\right) = cov\left(M_{t1} - \mathsf{E}\left(M_{t1}\right), M_{t2} - \mathsf{E}\left(M_{t2}\right)\right)$. Hence, we can consider the case where M_{tj} are centered (and center $\widehat{m}_{itj}$ with respect to their mean). This is inocuous to compute an estimator of the covariance statistics, but it will be useful to compute a variance of this estimator. The covariance statistic embeds the variance which is obtained when fixing $M_{t1} = M_{t2}$. We have:

$$\begin{aligned} cov\left(\widehat{M}_{t1}, \widehat{M}_{t2} \left| S_t = e \right.\right) &= E\left[cov\left(\widehat{M}_{t1}, \widehat{M}_{t2} \left| M_t, M_{t1}, M_{t2} \right.\right)\right] \\ &+cov\left[E\left(\widehat{M}_{t1} \left| M_t, M_{t1} \right.\right), E\left(\widehat{M}_{t2} \left| M_t, M_{t2} \right.\right)\right] \end{aligned}$$

Using equation (S.2), we get $E\left(\widehat{M}_{tj} \left| M_t, M_{tj} \right.\right) = M_{tj}$ and we then have:

$$cov\left(M_{t1}, M_{t2}\right) = cov\left(\widehat{M}_{t1}, \widehat{M}_{t2}\right) - E\left[cov\left(\widehat{M}_{t1}, \widehat{M}_{t2} \left| M_t, M_{t1}, M_{t2} \right.\right)\right] \tag{S.5}$$

We now explain how to construct estimators of the two right-hand side terms. Denote by Φ_t the subset of individuals who are employed at date t. An estimator of the first right-hand side term is given by:

$$\widehat{cov}\left(\widehat{M}_{t1}, \widehat{M}_{t2}\right) = \frac{1}{N-1}\sum_i\left(\widehat{m}_{it1} - \frac{1}{N}\sum_i \widehat{m}_{it1}\right)\left(\widehat{m}_{it2} - \frac{1}{N}\sum_i \widehat{m}_{it2}\right) \tag{S.6}$$

Using equation (S.2), the second right-hand side term of equation (**??**) can be rewritten as:

$$\begin{aligned} E\left[cov\left(\widehat{M}_{t1}, \widehat{M}_{t2} \left| M_t, M_{t1}, M_{t2} \right.\right)\right] &= cov\left(\sigma x^c_{it1}\left(x_i' M_{\varphi_i} x_i\right)^{-1} x_i' M_{\varphi_i}\varepsilon_i, \sigma x^c_{it2}\left(x_i' M_{\varphi_i} x_i\right)^{-1} x_i' M_{\varphi_i}\varepsilon_{ii}\right) \\ &= \sigma^2 E\left[x^c_{it1}\left(x_i' M_{\varphi_i} x_i\right)^{-1} x^{c\prime}_{it2} \left| x_{it}, x^c_{it1}, x^c_{it2} \right.\right] \end{aligned} \tag{S.7}$$

An estimator of expression (S.7) is then given by:

$$\widehat{E}\left[cov\left(\widehat{M}_{t1}, \widehat{M}_{t2} \left| M_t, M_{t1}, M_{t2} \right.\right)\right] = \frac{\widehat{\sigma}^2}{N}\sum_i x^c_{it1}\left(x_i' M_{\varphi_i} x_i\right)^{-1} x^{c\prime}_{it2} \tag{S.8}$$

where $\widehat{\sigma^2}$ is the unbiased OLS estimator of σ^2 given by equation (**??**). Finally, a consistent estimator of the covariance is given by:

$$\widehat{cov}\left(M_{t1}, M_{t2}\right) = \widehat{cov}\left(\widehat{M}_{t1}, \widehat{M}_{t2}\right) - \widehat{E}\left[cov\left(\widehat{M}_{t1}, \widehat{M}_{t2} \left| M_t, M_{t1}, M_{t2} \right.\right)\right] \tag{S.8}$$

where we replace factors φ_t with their estimators.

Reintroducing the conditioning $S_t = e$, sum should involve only observations at date t when individuals are in employment and N should be replaced by N_t (the number of individuals in employment at date t).

E.2.2 Variance of the estimator

We now compute the variance of the covariance estimator that can be rewritten as:

$$\widehat{cov}\left(M_{t1}, M_{t2}\right) = \frac{1}{N-1}\sum_{i}\widehat{m}_{it1}\widehat{m}_{it2} - \frac{1}{\left(N-1\right)N}\sum_{i,j}\widehat{m}_{it1}\widehat{m}_{jt2} - \frac{\widehat{\sigma}^{2}}{N}\sum_{i}\frac{1}{T}x_{it1}^{c}\Psi_{i}x_{it2}^{c\prime}$$

where $\Psi_i = \left(\frac{x_i' M_{\varphi_i} x_i}{T}\right)^{-1}$. As shown by Fernandez-Val and Weidner (2018), the debiasing term can be ignored in the computation, and we are thus interested in the variance of:

$$\widetilde{cov}\left(M_{t1}, M_{t2}\right) = \frac{1}{N-1}\sum_{i}\widehat{m}_{it1}\widehat{m}_{it2} - \frac{1}{\left(N-1\right)N}\sum_{i,j}\widehat{m}_{it1}\widehat{m}_{jt2}$$

The expectation of this covariance is given by:

$$\begin{aligned} E\left[\widetilde{cov}\left(M_{t1}, M_{t2}\right)\right] &= \frac{1}{N-1}\sum_{i}E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right) - \frac{1}{\left(N-1\right)N}\sum_{i,j}E\left(\widehat{m}_{it1}\widehat{m}_{jt2}\right) \\ &= \frac{1}{N}\sum_{i}E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right) - \frac{1}{\left(N-1\right)N}\sum_{i,j}E\left(\widehat{m}_{it1}\widehat{m}_{jt2}\right) \end{aligned}$$

We consider that individual observed and unobserved characteristics are identically and independently distributed across individuals (except x^c in the case where it is deterministic or partly deterministic, with deterministic parts taking the same values for all individuals). Using also that fact that $E\left(\widehat{m}_{itj}\right)$, we get:

$$E\left[\widetilde{cov}\left(M_{t1}, M_{t2}\right)\right] = E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)$$

and thus:

$$E\left[\widetilde{cov}\left(M_{t1}, M_{t2}\right)\right]^{2} = \left[E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)\right]^{2} \tag{S.9}$$

We also have:

$$\begin{aligned} E\left[\widetilde{cov}\left(M_{t1}, M_{t2}\right)^{2}\right] &= E\left(\frac{1}{N_t-1}\sum_{i}E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right) - \frac{1}{\left(N-1\right)N}\sum_{i,j}E\left(\widehat{m}_{it1}\widehat{m}_{jt2}\right)\right)^{2} \\ &= E\left(\frac{1}{\left(N-1\right)^{2}}\sum_{i,j}\widehat{m}_{it1}\widehat{m}_{it2}\widehat{m}_{jt1}\widehat{m}_{jt2}\right) \\ &\quad +E\left(\frac{1}{\left(N-1\right)^{2}N^{2}}\sum_{i,j,k,l}\widehat{m}_{it1}\widehat{m}_{jt2}\widehat{m}_{kt1}\widehat{m}_{lt2}\right) \\ &\quad -\frac{2}{\left(N-1\right)^{2}N}E\left(\sum_{i,k,l}\widehat{m}_{it1}\widehat{m}_{it2}\widehat{m}_{kt1}\widehat{m}_{lt2}\right) \end{aligned} \tag{S.10}$$

We are going to compute the three right-hand side terms. We have for the first one:

$$\begin{aligned}
& E\left(\sum_{i,j}\widehat{m}_{it1}\widehat{m}_{it2}\widehat{m}_{jt1}\widehat{m}_{jt2}\right) \\
= & \sum_{i} E\left(\widehat{m}_{it1}^{2}\widehat{m}_{it2}^{2}\right)+\sum_{i,j|i\neq j} E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)E\left(\widehat{m}_{jt1}\widehat{m}_{jt2}\right) \\
= & NE\left(\widehat{m}_{it1}^{2}\widehat{m}_{it2}^{2}\right)+N\left(N-1\right)E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)^{2} \qquad \text{(S.11)}
\end{aligned}$$

We have for the second term of expression (S.10):

$$E\left(\sum_{i,j,k,l}\widehat{m}_{it1}\widehat{m}_{jt2}\widehat{m}_{kt1}\widehat{m}_{lt2}\right)=E\left(\sum_{i}\widehat{m}_{it1}^{2}\widehat{m}_{it2}^{2}\right)+2E\left(\sum_{i,j|i\neq j}\widehat{m}_{it1}\widehat{m}_{it2}\widehat{m}_{jt1}\widehat{m}_{jt2}\right)$$

Developping this expression leads to:

$$E\left(\sum_{i,j,k,l}\widehat{m}_{it1}\widehat{m}_{jt2}\widehat{m}_{kt1}\widehat{m}_{lt2}\right)=NE\left(\widehat{m}_{it1}^{2}\widehat{m}_{it2}^{2}\right)+2N\left(N-1\right)E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)^{2} \qquad \text{(S.12)}$$

Finally, we have for the third term of expression (S.10):

$$E\left(\sum_{i,j,k}\widehat{m}_{it1}\widehat{m}_{it2}\widehat{m}_{jt1}\widehat{m}_{kt2}\,|S_t=e\right)=E\left(\sum_{i}\widehat{m}_{it1}^{2}\widehat{m}_{it2}^{2}\right)+E\left(\sum_{i,j|i\neq j}\widehat{m}_{it1}\widehat{m}_{it2}\widehat{m}_{jt1}\widehat{m}_{jt2}\right)$$

Developping this expression leads to:

$$E\left(\sum_{i,j,k\in\Phi_t}\widehat{m}_{it1}\widehat{m}_{it2}\widehat{m}_{jt1}\widehat{m}_{kt2}\right)=NE\left(\widehat{m}_{it1}^{2}\widehat{m}_{it2}^{2}\right)+N\left(N-1\right)E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)^{2} \qquad \text{(S.13)}$$

Finally, equations (S.11), (**??**) and (**??**) can be used to compute $E\left[\widetilde{cov}\left(X_{t1}^{c}\mathsf{H},X_{t2}^{c}\mathsf{H}\right)^{2}\right]$ on the left-hand side of equation (S.10) after replacing expectations by their empirical counterparts:

$$\begin{aligned}
E\left[\widetilde{cov}\left(M_{t1},M_{t2}\right)^{2}\right] = & \frac{1}{\left(N-1\right)^{2}}\left[NE\left(\widehat{m}_{it1}^{2}\widehat{m}_{it2}^{2}\right)+N\left(N-1\right)E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)^{2}\right] \\
& +\frac{1}{\left(N-1\right)^{2}N^{2}}\left[NE\left(\widehat{m}_{it1}^{2}\widehat{m}_{it2}^{2}\right)+2N\left(N-1\right)E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)^{2}\right] \\
& -\frac{2}{\left(N-1\right)^{2}N}\left[NE\left(\widehat{m}_{it1}^{2}\widehat{m}_{it2}^{2}\right)+N\left(N-1\right)E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)^{2}\right] \\
= & \left[\frac{N}{\left(N-1\right)^{2}}+\frac{1}{\left(N-1\right)^{2}N}-\frac{2}{\left(N-1\right)^{2}}\right]E\left(\widehat{m}_{it1}^{2}\widehat{m}_{it2}^{2}\right) \\
& +\left[\frac{N}{N-1}+\frac{2}{\left(N-1\right)N}-\frac{2}{N-1}\right]E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)^{2} \\
& +\left[1-\frac{1}{N}+\frac{1}{N\left(N-1\right)}\right]E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)^{2} \qquad \text{(S.14)}
\end{aligned}$$

Using equations (??) and (S.14), we obtain:

$$\begin{aligned} V\left[\widetilde{cov}\left(M_{t1}, M_{t2}\right)\right] &= \frac{1}{N}\left[E\left(\widehat{m}_{it1}^2 \widehat{m}_{it2}^2\right) - E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)^2\right] + \frac{1}{N\left(N-1\right)} E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)^2 \\ &= \frac{1}{N}\left[E\left(\widehat{m}_{it1}^2 \widehat{m}_{it2}^2\right) - E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)^2\right] + O\left(\frac{1}{N^2}\right) \end{aligned}$$

This formula can be computed from the data using the empirical counterparts of $E\left(\widehat{m}_{it1}\widehat{m}_{it2}\right)$ and $E\left(\widehat{m}_{it1}^2 \widehat{m}_{it2}^2\right)$. When considering again only observations when individuals are employed, sum should involve only observations at date t when individuals are in employment, N should be replaced by N_t, x_i should involve only observations of individual i when she is employed, and T should be replaced by T_i, the number of observations that individual i is employed.

E.3 Bias-corrected couterfactual inter-centile gaps

E.3.1 Estimator

We can get a formula for bias-corrected quantile differences in the unfeasible case in which factors φ_t are known using the approach of Jochmans and Weidner (2019). Our objects of interest are of the form:

$$\Delta q_t\left(\tau_1, \tau_2\right) = q_t\left(\tau_2\right) - q_t\left(\tau_1\right)$$

with $q_t\left(\tau\right) = \inf_q \left\{q \left| F_t\left(q\right) \geqslant \tau\right.\right\}$ where $F_t\left(\bullet\right)$ is the cumulative of $X_t^c \mathsf{H}$ (the individual-specific effect of counterfactual explanatory variables). Realizations of $X_t^c \mathsf{H}$ are denoted $x_{it}^c \eta_i$, but η_i are not observed and are replaced by their estimators $\widehat{\eta}_i$ given by equation (S.2). As previously, denote $M_{it} = X_t^c \mathsf{H}$, $m_{it} = x_{it}^c \eta_i$ and $\widehat{m}_{it} = x_{it}^c \widehat{\eta}_i$. For a given value of potential experience t, $\widehat{m}_{it}$ is of the form studied by Jochmans and Weidner (2019):

$$\widehat{m}_{it} = m_{it} + \frac{1}{\sqrt{T}} \sigma_{it} \widetilde{\varepsilon}_{it} \tag{S.15}$$

in which $\sigma_{it}^2 = \frac{T}{T_l} \sigma^2 x_{it}^c \left(\frac{x_i' M_{\varphi_i} x_i}{T_i}\right)^{-1} x_{it}^{c\prime}$ and $\widetilde{\varepsilon}_{it}$ is a centered disturbance with unit variance.

Consider a given t and suppose that individuals are sorted in increasing order of values $\widehat{m}_{it}$. The plug-in estimator for quantiles of the distribution of $\widehat{m}_{it}$ is given by:

$$\widehat{q}_t\left(\tau\right) = \widehat{m}_{\lfloor \tau N_t^c \rfloor t}$$

where $\lfloor \bullet \rfloor$ denotes the integer function and N_t^c is the number of individuals who are employed at period t in the counterfactual scenario. According to Jochmans and Weidner (2019), we have:

$$\sqrt{N_t^c}\left(\widehat{q}_t\left(\tau\right) - q_t\left(\tau\right) - \frac{b_t\left(\tau\right)}{T}\right) \longrightarrow N\left(0, \sigma_t^2\left(\tau\right)\right) \tag{S.16}$$

where:

$$b_t(\tau) = -\frac{b_t^F(q_t(\tau))}{f_t(q_t(\tau))} \tag{S.17}$$

$$\sigma_t^2(\tau) = \frac{\tau(1-\tau)}{f_t(q_t(\tau))^2} \tag{S.18}$$

with $f_t(\bullet)$ the density fonction of M_{it} and

$$b_t^F(q) = \frac{\partial g_t(q)}{\partial q} \text{ with } g_t(q) = E\left(\sigma_{it}^2 | m_{it} = q, \varphi\right) f_t(q)/2$$

In fact, Jochmans and Weidner (2019) argue that a direct analytical correction of the bias is not very suitable because it has the unattractive property that it requires a non-parametric estimator of the density $f_t(\bullet)$, which further shows up in the denominators of $b_t(\tau)$. Consequently, they rather propose the following correction:

$$\overset{\vee}{q}_t(\tau) = \widehat{m}_{\lfloor \widehat{\tau}_t^*(\tau) N_t^c \rfloor t} \tag{S.19}$$

with:

$$\widehat{\tau}_t^*(\tau) = \tau + \frac{1}{T}\widehat{b}_t^F(\widehat{q}_t(\tau))$$

where:

$$\widehat{b}_t^F(q) = -\frac{1}{2N_t^c h^2} \sum_{i | s_{it}^c = e} \widehat{\sigma}_{it}^2 \kappa' \left(\frac{\widehat{m}_{it} - q}{h}\right)$$

with $s_{it}^c \in \{e, n\}$ the employment status of individual i at period t in the counterfactual scenario, κ' is the derivative of a Gaussian kernel and h is the bandwidth, and $\widehat{\sigma}_{it}^2$ is an estimator of σ_{it}^2. To construct this estimator, we recover as before the OLS estimator of σ^2 denoted $\widehat{\sigma}^2$ and use the formula:

$$\widehat{\sigma}_{it}^2 = \frac{T}{T_i}\widehat{\sigma}^2 x_{it}^c \left(\frac{x_i' M_i x_i}{T_i}\right)^{-1} x_{it}^{c\prime}$$

The idea of the bias correction given by equation (**??**) is to correct the ranks at which quantiles are computed because the sampling error inflates quantiles. Jochmans and Weidner (2019) show in fact that:

$$\sqrt{N}\left(\overset{\vee}{q}_t(\tau) - q_t(\tau)\right) \longrightarrow N\left(0, \sigma_t^2(\tau)\right) \tag{S.20}$$

A bias-corrected estimator of any inter-quantile range is then given by:

$$\widehat{\Delta q}_t(\tau_1, \tau_2) = \overset{\vee}{q}_t(\tau_2) - \overset{\vee}{q}_t(\tau_1)$$

E.3.2 Asymptotic variance of the estimator

We now explain how to compute the asymptotic variance of the bias-corrected estimator. Jochmans and Weidner (2019) establish the asymptotic distribution of only one quantile $\overset{\vee}{q}_t(\tau)$. But consistently with the literature on sample quantiles (see for instance Moore, 1969), it can be generalized such that we have for $0 < \tau_1 < \tau_2 < 1$:

$$\sqrt{N_t^c}\left(\left(\begin{array}{c} \overset{\vee}{q}_t(\tau_1) \\ \overset{\vee}{q}_t(\tau_2) \end{array}\right) - \left(\begin{array}{c} q_t(\tau_1) \\ q_t(\tau_2) \end{array}\right)\right) \longrightarrow N\left(0, \left(\begin{array}{cc} \sigma_t^2(\tau_1) & c_t(\tau_1,\tau_2) \\ c_t(\tau_1,\tau_2) & \sigma_t^2(\tau_2) \end{array}\right)\right) \tag{S.21}$$

with:

$$c_t(\tau_1,\tau_2) = \frac{\tau_1(1-\tau_2)}{f_t(q_t(\tau_1))\, f_t(q_t(\tau_2))}$$

It is easy to show from (**??**) that $\widehat{\Delta q}_t(\tau_1,\tau_2)$ is asymptotically normal with variance:

$$V\left(\widehat{\Delta q}_t(\tau_1,\tau_2)\right) = \sigma_t^2(\tau_1) + \sigma_t^2(\tau_2) - 2c_t(\tau_1,\tau_2)$$

E.4 Bias-corrected rank correlation

Consider two variables $M_1 = X_1^c\mathsf{H}$ and $M_2 = X_2^c\mathsf{H}$ such that we do not observe their values for every individual and date, m_{it1} and m_{it1}, but rather some estimators $\widehat{m}_{it1}$ and $\widehat{m}_{it2}$ that are realizations of $\widehat{M}_1$ and $\widehat{M}_2$. We are interested in the Spearman correlation which is the rank correlation given by:

$$r_s = \frac{cov(\tau_{M_1}, \tau_{M_2})}{\sigma_{\tau_{M_1}}\sigma_{\tau_{M_2}}} \tag{S.22}$$

where $\tau_{M_1} = F_{M_1}(M_1)$ and $\tau_{M_2} = F_{M_2}(M_2)$. Note that, in contrast with previous section, we are pooling all the years when constructing an empirical counterpart of this rank correlation. We want to compute this correlation using empirical counterparts of the covariance and standard deviations, but we do not observe the empirical ranks $\widehat{F}_{M_1}(m_{it1})$ and $\widehat{F}_{M_2}(m_{it2})$. We rather observe $\widehat{\tau}_{it}^1 \equiv \widehat{F}_{\widehat{M}_1}(\widehat{m}_{it1})$ and $\widehat{\tau}_{it}^2 \equiv \widehat{F}_{\widehat{M}_2}(\widehat{m}_{it2})$. There are two problems here. The distributions of $\widehat{M}_1$ and $\widehat{M}_2$ are distorted compared to those of M_1 and M_2 due to sampling error, and this creates biases on rank estimators. Also, we resort to estimators $\widehat{m}_{it1}$ and $\widehat{m}_{it2}$ of quantities m_{it1} and m_{it2}.

First note that:

$$\sqrt{T}\left(\begin{array}{c} \widehat{m}_{it1} - m_{it1} \\ \widehat{m}_{it2} - m_{it2} \end{array}\right) \longrightarrow N\left(0, \left(\begin{array}{cc} \sigma_{1it}^2 & \sigma_{12it} \\ \sigma_{12it} & \sigma_{2it}^2 \end{array}\right)\right) \tag{S.23}$$

where $\sigma^2_{jit} = \sigma^2 p\lim\left[\frac{T}{T_i}x^{cj}_{it}\left(\frac{x_i'M_ix_i}{T_i}\right)^{-1}x^{cj\prime}_{it}\right]$ and $\sigma^2_{12it} = \sigma^2 p\lim\left[\frac{T}{T_i}x^{c1}_{it}\left(\frac{x_i'M_ix_i}{T_i}\right)^{-1}x^{c2\prime}_{it}\right]$. Hence, we can write that:

$$\widehat{m}_{itj} - m_{itj} = \frac{\sigma^2_{jit}}{\sqrt{T}}\omega^j_{it} + o_P\left(1/\sqrt{T}\right) \tag{S.24}$$

where ω^j_{it} follows a standard normal distribution.

To estime the covariance term entering the rank correlation (**??**), we consider the consistent estimors of cumulatives proposed by Jochmans and Weidner (2019):

$$\breve{F}^j(m) = \widehat{F}^j\left(m\right) - \frac{b^{\widehat{F}^j}\left(m\right)}{T} \tag{S.25}$$

with:

$$\widehat{b}^{F^j}\left(m\right) = -\frac{1}{2N^c_{\cdot}h^2}\sum_{t=1}^{T}\sum_{i\left|s^c_{it}=e\right.}\widehat{\sigma}^2_{it}\kappa'\left(\frac{\widehat{m}_{itj}-m}{h}\right)$$

where $N^c_{\cdot} = \sum_{t=1}^{T} N^c_t$. We have:

$$\begin{aligned}
& Cov\left(\breve{F}^1(\widehat{m}_{it1}), \breve{F}^2(\widehat{m}_{it2})\right) \\
= \; & Cov\left(\breve{F}^1(\widehat{m}_{it1}) - F^1(\widehat{m}_{it1}) + F^1(\widehat{m}_{it1}), \breve{F}^2(\widehat{m}_{it2}) - F^2(\widehat{m}_{it2}) + F^2(\widehat{m}_{it2})\right) \\
= \; & Cov\left(F^1(\widehat{m}_{it1}), F^2(\widehat{m}_{it2})\right) + cov\left(\breve{F}^1(\widehat{m}_{it1}) - F^1(\widehat{m}_{it1}), \breve{F}^2(\widehat{m}_{it2}) - F^2(\widehat{m}_{it2})\right) \\
& + cov\left(F^1(\widehat{m}_{it1}), \breve{F}^2(\widehat{m}_{it2}) - F^2(\widehat{m}_{it2})\right) + cov\left(F^2(\widehat{m}_{it2}), \breve{F}^1(\widehat{m}_{it1}) - F^1(\widehat{m}_{it1})\right)
\end{aligned} \tag{S.26}$$

We are first going to compute the first right-hand side term and then show that the three additional right-hand side terms are negligible. We have:

$$Cov\left(F^1(\widehat{m}_{it1}), F^2(\widehat{m}_{it2})\right) = E\left[F^1(\widehat{m}_{it1})F^2(\widehat{m}_{it2})\right] - E\left[F^1(\widehat{m}_{it1})\right]E\left[F^2(\widehat{m}_{it2})\right]$$

As F is 3 times differentiable, we get:

$$F^j(\widehat{m}_{itj}) = F^j(m_{itj}) + f^j(m_{itj})(\widehat{m}_{itj} - m_{itj}) + f^{j\prime}(m_{itj})(\widehat{m}_{itj} - m_{itj})^2/2 + o_P((\widehat{m}_{itj} - m_{itj})^2)$$

Hence, we have:

$$\begin{aligned}
& E\left[F^1(\widehat{m}_{it1})F^2(\widehat{m}_{it2})\right] \\
= \; & E\left[F^1(m_{it1})F^2(m_{it2})\right] \\
& + E\left[F^1(m_{it1})\left[f^2(m_{it2})(\widehat{m}_{it2} - m_{it2}) + f^{2\prime}(m_{it2})(\widehat{m}_{it2} - m_{it2})^2/2 + o_P((\widehat{m}_{it2} - m_{it2})^2)\right]\right] \\
& + E\left[F^2(m_{it2})\left[f^1(m_{it1})(\widehat{m}_{it1} - m_{it1}) + f^{1\prime}(m_{it1})(\widehat{m}_{it1} - m_{it1})^2/2 + o_P((\widehat{m}_{it1} - m_{it1})^2)\right]\right] \\
& + E\left[\begin{array}{c} [f^2(m_{it2})(\widehat{m}_{it2} - m_{it2}) + f^{2\prime}(m_{it2})(\widehat{m}_{it2} - m_{it2})^2/2 + o_P((\widehat{m}_{it2} - m_{it2})^2)] \\ \times [f^1(m_{it1})(\widehat{m}_{it1} - m_{it1}) + f^{1\prime}(m_{it1})(\widehat{m}_{it1} - m_{it1})^2/2 + o_P((\widehat{m}_{it1} - m_{it1})^2)] \end{array}\right]
\end{aligned} \tag{S.27}$$

For a centered bivariate normal (X,Y), we have $E\left(X^2Y\right)=0$. We can thus use the following equalities obtained from (??):

$$\begin{aligned}
E\left[(\widehat{m}_{it1}-m_{it1})(\widehat{m}_{it2}-m_{it2})\,|m_{it1},m_{it2}\right] &= \sigma_{12it}/T+o_P\left(1/T\right)\\
E\left[(\widehat{m}_{it1}-m_{it1})(\widehat{m}_{2}-m_{it2})^2\,|m_{it1},m_{it2}\right] &= o_P\left(1/T^{3/2}\right)\\
E\left[(\widehat{m}_{it2}-m_{it2})(\widehat{m}_{it1}-m_{it1})^2\,|m_{it1},m_{it2}\right] &= o_P\left(1/T^{3/2}\right)
\end{aligned}$$

Equation (S.27) can then be rewritten as:

$$\begin{aligned}
E\left[F^1(\widehat{m}_{it1})F^2(\widehat{m}_{it2})\right] &= E\left[F^1(m_{it1})F^2(m_{it2})\right]\\
&\quad+\frac{1}{T}E\left[F^1(m_{it1})f^{2\prime}(m_{it2})\sigma_{2it}^2\right]\\
&\quad+\frac{1}{T}E\left[F^2(m_{it2})f^{1\prime}(m_{it1})\sigma_{1it}^2\right]\\
&\quad+\frac{1}{T}E\left[f^1(m_{it1})f^2(m_{it2})\sigma_{12it}^2\right]+o_P\left(1/T\right) \qquad \text{(S.28)}
\end{aligned}$$

We also have:

$$E\left[F(\widehat{m}_{itj})\right]=E\left[F(m_{itj})\right]+\frac{1}{T}E\left[f'(m_{itj})\sigma_{jit}^2\right]+o_P(1/T)$$

Hence:

$$\begin{aligned}
E\left[F^1(\widehat{m}_{it1})\right]E\left[F^2(\widehat{m}_{it2})\right] &= E\left[F^1(\widehat{m}_{it1})\right]E\left[F^2(\widehat{m}_{it2})\right]\\
&\quad+\frac{1}{T}E\left[F^1(m_{it1})\right]E\left[f^{2\prime}(m_{it2})\sigma_{2it}^2\right]\\
&\quad+\frac{1}{T}E\left[F^2(m_{it2})\right]E\left[f^{1\prime}(m_{it1})\sigma_{1it}^2\right]+o_P(1/T) \qquad \text{(S.29)}
\end{aligned}$$

Putting together (S.28) and (S.29), we get:

$$\begin{aligned}
Cov\left(F^1(\widehat{m}_{it1}),F^2(\widehat{m}_{it2})\right) &= Cov\left(F^1(m_{it1}),F^2(m_{it2})\right)\\
&\quad+\frac{1}{T}Cov\left(F^1(m_{it1}),f^{2\prime}(m_{it2})\sigma_{2it}^2\right)\\
&\quad+\frac{1}{T}Cov\left(F^2(m_{it2}),f^{1\prime}(m_{it1})\sigma_{1it}^2\right)\\
&\quad+\frac{1}{T}E\left[f^1(m_{it1})f^2(m_{it2})\sigma_{12it}\right]+o_P\left(1/T\right) \qquad \text{(S.30)}
\end{aligned}$$

We now show that the three other terms in equation (S.26) are negligible. Applying Cauchy-

Schwartz inequality, we get:

$$\begin{aligned}
& cov\left(F^1(\widehat{m}_{it1}), \breve{F}^2(\widehat{m}_{it2}) - F^2(\widehat{m}_{it2})\right) \\
\leqslant\; & V\left(F^1(\widehat{m}_{it1})\right)^{1/2} V\left(\breve{F}^2(\widehat{m}_{it2}) - F^2(\widehat{m}_{it2})\right)^{1/2} \qquad \text{(S.31)} \\
& cov\left(F^2(\widehat{m}_{it2}), \breve{F}^1(\widehat{m}_{it1}) - F^1(\widehat{m}_{it1})\right) \\
\leqslant\; & V\left(F^2(\widehat{m}_{it2})\right)^{1/2} V\left(\breve{F}^1(\widehat{m}_{it1}) - F^1(\widehat{m}_{it1})\right)^{1/2} \qquad \text{(S.32)} \\
& cov\left(\breve{F}^1(\widehat{m}_{it1}) - F^1(\widehat{m}_{it1}), \breve{F}^2(\widehat{m}_{it2}) - F^2(\widehat{m}_{it2})\right) \\
\leqslant\; & V\left(\breve{F}^1(\widehat{m}_{it1}) - F^1(\widehat{m}_{it1})\right)^{1/2} V\left(\breve{F}^2(\widehat{m}_{it2}) - F^2(\widehat{m}_{it2})\right)^{1/2} \qquad \text{(S.33)}
\end{aligned}$$

where terms $V\left(F^j(\widehat{m}_{itj})\right)$ are bounded since $F^j(\cdot)$ is bounded by 1. We now show that terms $V\left(\breve{F}^j(\widehat{m}_{itj}) - F^j(\widehat{m}_{itj})\right)$ are negligible. Using (**??**), (**??**) and (S.18), we then get (see Proposition 1 in Jochmans and Weidner, 2019):

$$\left[\breve{F}^j(m) - F^j(m)\right]^2 = \frac{F^j(m)\left[1 - F^j(m)\right]}{N_{\cdot}^c} + o_P\left(\frac{1}{NT}\right) \qquad \text{(S.34)}$$

since that the total number of observations in the counterfactual sample $N_{\cdot}^c$ grows at the same speed as NT. As a consequence, we have:

$$\begin{aligned}
E\left(\left[\breve{F}^j(\widehat{m}_{itj}) - F^j(\widehat{m}_{itj})\right]^2\right) &= E\left[E\left(\left[\breve{F}^j(\widehat{m}_{itj}) - F^j(\widehat{m}_{itj})\right]^2 \mid m = \widehat{m}_{itj}\right)\right] \\
&= E\left[\frac{F^j(\widehat{m}_{itj})\left[1 - F^j(\widehat{m}_{itj})\right]}{N_{\cdot}^c}\right] + o_P\left(\frac{1}{NT}\right) \\
&\leq M_\sigma / N_{\cdot}^c + o_P\left(\frac{1}{NT}\right) \qquad \text{(S.35)}
\end{aligned}$$

in which $E\left[\frac{F^j(\widehat{m}_{itj})\left[1-F^j(\widehat{m}_{itj})\right]}{N_{\cdot}^c}\right]$ is bounded, say be M_σ, since $F^j(\widehat{m}_{itj})$ is bounded by one. The bound also applies to the variance when assuming that $N/T^4 \to 0$ since, according to Proposition 1 in Jochmans and Weidner (2019), we have:

$$E\left(\left[\breve{F}^j(\widehat{m}_{itj}) - F^j(\widehat{m}_{itj})\right]\right) = E\left[E\left(\left[\breve{F}^j(\widehat{m}_{itj}) - F^j(\widehat{m}_{itj})\right] \mid \widehat{m}_{itj} = m\right)\right] = o_P(1/T^2) \qquad \text{(S.36)}$$

Finally, we get under the assumption $N/T^2 \to c > 0$ that :

$$\begin{aligned}
Cov\left(F^1(m_{it1}), F^2(m_{it2})\right) &= Cov\left(\breve{F}^1(\widehat{m}_{it1}), \breve{F}^2(\widehat{m}_{it2})\right) \\
&\quad -\frac{1}{T} Cov\left(F^1(m_{it1}), f^{2\prime}(m_{it2})\sigma_{2it}^2\right) \\
&\quad -\frac{1}{T} Cov\left(F^2(m_{it2}), f^{1\prime}(m_{it1})\sigma_{1it}^2\right) \\
&\quad -\frac{1}{T} E\left[f^1(m_{it1}) f^2(m_{it2}) \sigma_{12it}\right] \\
&\quad + o_P(1/T)
\end{aligned}$$

A consistent estimator of the covariance of ranks can then be computed from the empirical counterparts of the first four right-hand side terms after replacing $F^j(m_{itj})$, $f^j(m_{itj})$ and $f^{j\prime}(m_{itj})$ with their empirical counterparts $\breve{F}^j(\widehat{m}_{itj})$, $f^j(\widehat{m}_{itj})$ and $f^{j\prime}(\widehat{m}_{itj})$. We can also recover the variances from this expression using $F^j(m_{itj})$ for the two arguments of the covariance. Finally, we are able to derive an estimate of the rank correlation (**??**).

F Monte-Carlo simulations

We carried out Monte-Carlo simulations to investigate biases, bias correction and the statistics particularly affected by bias in a setting which is inspired by our working sample and the model we use. We describe the empirical setting and present results under correct specification. We also explore results under incorrect specification mimicking our empirical models whereby we omit factors only – and thus neglect selection and endogeneity of experience – or we omit factors as well as variables describing interruptions in participation (x_{i3}, x_{i4}).

F.1 Experimental setting

We consider a single cohort and the time horizon is the one of the cohort that has the longest horizon in the data ($T = 27$). The wage specification is:

$$\ln y_{it} = x_{it}\eta_i + \varphi_t\theta_i + \sigma\varepsilon_{it}$$

in which the idiosyncratic errors, ε_{it}, are normal, homoskedastic and independent over time and in which (η_i, θ_i) is normally distributed. The mean of individual specific parameters $(E(\eta_i), E(\theta_i))$ is fixed at the estimated value obtained in our preferred estimations when using two factors. Values of factors φ_t are as well taken to be equal to their estimates.

If the specification is correct, means are unbiased asymptotically. In contrast, the estimated covariance matrix, $V(\eta_i, \theta_i)$, is biased upwards. In order to reduce this bias, we use the estimated value of the variance in our preferred experiment and extract its eigenvalues, say $\hat{\lambda}$, ordered from $\hat{\lambda}_{\max}$ to $\hat{\lambda}_{\min}$. We then replace these eigenvalues with $\lambda^* = \hat{\lambda}_{\min}(\hat{\lambda}/\hat{\lambda}_{\min})^{.8}$ because it reduces their range while holding fixed the minimum value. We also experiment with a less truncated experiment by replacing .8 with .9. We then reconstruct the covariance matrix using the estimated eigenvectors. The value of the standard deviation σ is derived from the estimated residual variance.

The construction of wages is sequential over time since wages depend on real experience.

Participation is modelled such that a worker is in the private sector in year t if and only if:

$$E(\ln y_{it} \mid x_{it}, \eta_i, \varphi_t, \theta_i) > \ln y_{it}^*$$

in which:

$$\ln y_{it}^* = a_1 + a_2 b_i \left(x_{it}\eta_i + \varphi_t\theta_i\right) + \sigma_\zeta \zeta_{it}$$

where b_i is drawn in a uniform distribution $[0, 1]$ and ζ_{it} is drawn in a normal distribution, $N(0, 1)$, both independently of any other random variables. Parameters (a_1, a_2, σ_ζ) are calibrated by minimum distance so that generated data are in line with the estimates. Note that $\ln y_{it}^*$ is correlated with $E(\ln y_{it} \mid x_{it}, \eta_i, \varphi_t, \theta_i)$ and selection exists if we do not condition on individual-specific parameters and factors while selection is absent if we do condition on them.

The experiment involves $8,000$ individuals although the sample is truncated by imposing that the number of years of interruptions is less than 12 in each Monte Carlo experiment. It does not affect much the number of individuals – a few units at most in our replications. Bias is corrected under the assumption of homoskedasticity.

The implementation is as follows:

- The number of replications is 1000.
- For each replication, parameters η_i, θ_i and φ_t are estimated using the procedure proposed in the paper.
- We compute empirical counterparts of functions of estimated parameters $\widehat{\eta}_i$ – such as means, medians, variances and quantiles ($q10$, $q25$, $q75$, $q90$) – and of estimated potential wages, $\widehat{\eta}_{i0} + x_{it1}\widehat{\eta}_{i1} + x_{it2}\widehat{\eta}_{i2}$, predicted wage $x_{it}\widehat{\eta}_i$ and factor effects $\widehat{\varphi}_t\widehat{\theta}_i$ for different values of potential experience at selected periods $t = 1, 9, 17, 25$.
- Functions of estimated parameters $\widehat{\eta}_{i3}$ and $\widehat{\eta}_{i4}$ are computed using only observations such that η_{i4} is identifiable. Nonetheless, we chose not to report statistics related to the experience parameters, $\widehat{\eta}_{i3}$ and $\widehat{\eta}_{i4}$, but only to their contribution to predicted wages, $x_{it3}\widehat{\eta}_{i3} + x_{it4}\widehat{\eta}_{i4}$, since the former estimates are more biased than the latter. Statistics relative to experience parameters magnify the bias results that we obtain, and cannot but reinforce our conclusions. Results are available upon request.

- Functions of estimated potential wage, contribution of interruptions and predicted wage are computed for a given value of potential experience t using observations alternatively for all workers or for those who are not in the private sector at period t.

- We are interested in differences between these empirical means, and those constructed from random draws of true parameters η_i across Monte Carlo experiments. We do not look at differences between these draws and their expectations since we are not interested in conducting a standard analysis of sampling variation.

We assess the quality of the estimation method for a specific statistic from: 1/ Its bias (*Bias*), 2/ The mean absolute deviation (*MAD*) 3/ The root mean square error (*RMSE*) 4/ The standard deviation (*SD*) 5/ The coverage probabilities (*CP*) of Monte Carlo confidence intervals at a 95% level constructed using the aggregate standard error, SD. The average of true values (*True*) is also reported in Tables S.4 to S.25.

F.2 Results under correct specification

We assess the quality of the estimation method either by computing the ratio of RMSE and SD – a ratio of 1 meaning that the bias is absent, and a ratio of 10 meaning that the bias is very important – or by using the coverage probabilities (CP) – a value of .95 meaning that the estimation is of high quality, and a value between .50 and .80 meaning that the estimation is slightly biased. CP values which are less than .5 indicate an important bias. As standard errors are not biased (see Jochmans and Weidner, 2019, and all our Monte Carlo experiments below), the coverage probabilities are almost never greater than .95.

Here is a summary of Monte-Carlo results for parameters:

- Biases in means and medians for parameters η_{i0}, η_{i1} and η_{i2} are negligible (Table S.4). There are some significant biases in means and medians for parameters η_{i3} and η_{i4} (results not reported here). There are only very small biases on the mean and median for the first factor loading θ_{i1}, and the biases are slightly larger for the second factor θ_{i2} (results not reported here).

- Biases are large in variances for parameters η_{i0}, η_{i1} and η_{i2}. The coverage probabilities are equal to 0 (Table S.5). The correction of biases is partly successful only since the bias is still at the same level for (η_{i1}, η_{i2}) or 2 times greater than the standard deviation for η_{i0}, and the confidence intervals remain uncentered. In consequence, the coverage probabilities are between .43 to .83. The biases are huge for parameters η_{i3} and η_{i4} and are sizable for the two factors loadings θ_i, especially the second one (results not reported here).

- Table S.6 show large biases in estimated quantiles of parameters, $\eta_{i0}, \eta_{i1}, \eta_{i2}$ at percentiles 10, 25, 75 and 90 (but not for median). These biases are nonetheless less sizeable than the ones affecting variances since the coverage probabilities are in the range between 0 and 0.6. Parameter η_{i0} is particularly biased and the coverage probability is small (between 0 and 0.02 across percentiles). The parameters associated to slope and curvature are less biased since the coverage probabilities are between 0.3 and 0.6 for slope parameters, and 0.14 to 0.48 for curvature ones. It is thus not a surprise to find that bias correction is particularly successful for the coefficients associated to slope and curvature. The bias-corrected coverage probabilities are greater than .92. Bias correction is less successful for the level parameter, η_{i0} but the coverage probability is above .75 for the 90% quantile to be compared with .43 for the variance.

Here is a summary of Monte-Carlo results for potential wages, contribution of interruptions, predicted wages and factor effects:

- For potential wages, there are negligible biases in means all along the period for the full sample of males as well as for the subsample of non working males only (Table S.7).

- Table S.8 reports strong biases for variances in line with results for parameters η_{i0}, η_{i1} and η_{i2}. Bias correction is more successful in the first period of observation (the coverage probability is equal to .92) since there is no interruption but it becomes much worse afterwards (the coverage probability is between .14 and .35).

- Table S.9 reports the behavior of estimated quantiles. The raw estimated quantiles are slightly biased (except in period 1 and the medians), and the coverage probabilities can become as low as .05. Nonetheless, bias correction is achieving a much better job since the coverage probabilities are now no less than .8 and vary up to .92 (at $t = 20$ for 25% quantile).

- For the contribution of interruptions, $\eta_{i3}x_{i3}+\eta_{i4}x_{i4}$, results are similar to those for potential wages. In particular, biases in means (Table S.10) and medians (Table S.12, middle panel) are absent although this variable, $\eta_{i3}x_{i3} + \eta_{i4}x_{i4}$, has an accumulation point at 0 for those who have no interruptions up until period t. Variances are as strongly biased as the ones for potential wages (Table S.11), while the analysis of quantiles is complicated by the bunching at zero. In particular, bias correction is failing when the coverage probabilities of the raw quantiles are zero and the underlying conditions for bias correction do not seem to be satisfied because of the bunching.

- For predicted wages, there are negligible biases for means and medians (Tables S.13 and S.15). Variances are biased but bias correction works much better than for potential wages and for the predicted wage effect of interruptions (Table S.14). Coverage probabilities are in the range .73 to .95 (Period 25) and **are** U-shaped with potential experience. This is consistent with a compensation of biases affecting potential wage and the wage effect of interruptions. Interestingly, lower biases are confirmed by the analysis of estimated quantiles (Table S.15). Biases on estimated raw quantiles are relatively small (coverage probabilities above .82) and in particular absent in Period 25. Unsurprisingly, bias correction virtually solves the bias issue. Bias-corrected coverage probabilities are all between .93 and .96.

- For factor effects, there are negligible biases for means and medians, sizable biases for non-corrected variances and quantiles, but rather small biases for corrected variances (results not reported here).

F.3 Results under incorrect specification

Using the same 1000 replication samples, we also estimate parameters in two misspecified cases:

1. The basic model: regressors are reduced to $(1, t, \beta^{-t})$ and interruption variables are omitted as well as factors. Interruptions are supposed to have no impact and selection and endogeneity issues are neglected.

2. The no factor model: regressors are reduced to $(1, t, \beta^{-t})$ and $(x_t^{(3)}, x_t^{(4)})$ but factors are omitted. Selection and endogeneity issues are neglected.

Generally speaking, estimates are biased for the two models, including means and medians, and the biases do not disappear when $T \to \infty$ in contrast with the previous case. More surprisingly, the biases for the basic model are equal or smaller than the biases for the no factor one. This means that biases due to selection and endogeneity on the one hand and the omission of interruption variables are going in opposite directions.

More precisely we have the following results:

- Biases in means and medians of estimates of parameters $\eta_{0,1,2}$ are similar for the two models (Tables S.16 and S.21). Bias corrections in the medians are ineffective (Tables S.18 and S.23, middle panel) and we expect the same for means. On the other hand, variances are highly biased. The biases are larger for the no factor model than for the basic one, by a factor varying between 3 and 10 (Tables S.17 and S.22). Bias correction reduces somewhat the magnitude of biases but this is far from enough and the bias-corrected variances are still much more biased for the no factor model than for the basic one. This broad set of results is also true for quantiles as shown in Tables S.18 and S.23.

- Results are also similar for the predicted potential wages. Means and medians are strongly biased and the more so for the no factor model except when the missing data subsample (which consists in the observations out of the private sector) is used, in which case biases are the same for the basic and no factor models (Tables S.19 and S.24). In contrast, variances are always less biased for the basic model than for the no factor one (Tables S.20 and S.25). This is also true for quantiles (results not reported here).

REFERENCES

Fernández-Val, I., & M. Weidner, 2018, "Fixed effects estimation of large-T panel data models", *Annual Review of Economics*, 10, 109-138.

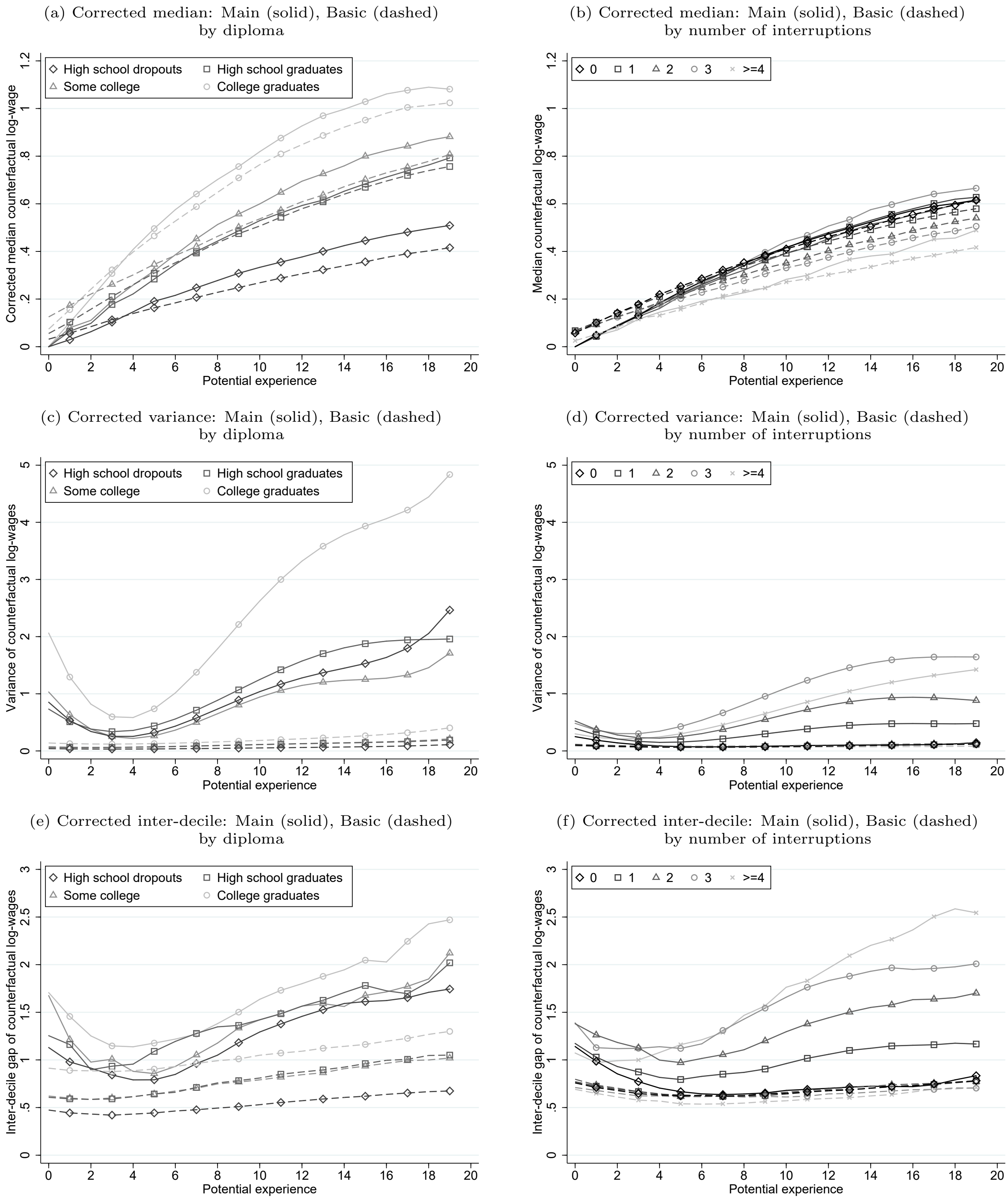
Jochmans, K., & M., Weidner, 2019, "Inference on a distribution from noisy draws", CEMMAP Working Paper 44/19.

Moore, D., 1969, "Limiting distributions for sample quantiles", *The American Mathematical Monthly*, 76(8), 927-929.

Song, M., 2013, "Asymptotic Theory for Dynamic Heterogeneous Panels with Cross-Sectional Dependence and its Applications", unpublished manuscript.

Su, L., & **G.**, **Gaosheng**, 2018, "Identifying latent grouped patterns in panel data models with interactive fixed effects", *Journal of Econometrics*, 206(2), 554-573.

Figure S.1: Median, variance and inter-decile range of counterfactual log-wages as a function of potential experience, main and basic specifications, by education and number of interruptions



Note: "Main": main specification that includes variables x_{i1}, x_{i2}, x_{i3} and x_{i4} as well as the additive individual effect and two interactive factors; "Basic": specification that includes only variables x_{i1} and x_{i2}, and the additive individual effect. In panels (a) and (b), for each diploma and number of interruptions, the levels of corrected median counterfactual log-wages are normalized for the two specifications using the value at period zero of the benchmark specification. "Corrected" statistics are obtained after bias correction as described in the Online Appendix.

Table S.1: Descriptive statistics on distributions of uncorrected and corrected parameters, individuals in employment 15-19 years

	η_{i0}	η_{i1}	η_{i2}	η_{i3}	η_{i4}	δ_1	δ_2
Mean	-0.56	0.08	-0.35	-0.09	0.54	-0.02	-0.01
Variance	3.45	0.24	13.59	2.56	258.52	0.22	0.02
Corrected Variance	2.44	0.16	9.19	1.99	206.04	0.15	0.01
Q5	-2.97	-0.65	-5.82	-2.06	-14.58	-0.70	-0.20
Q10	-2.10	-0.40	-3.83	-1.14	-8.36	-0.45	-0.14
Q25	-1.28	-0.10	-1.71	-0.39	-2.14	-0.19	-0.07
Median	-0.59	0.07	-0.29	-0.00	0.00	-0.03	-0.01
Q75	0.07	0.25	1.04	0.32	2.27	0.15	0.05
Q90	0.94	0.54	3.34	0.99	9.44	0.45	0.12
Q95	1.85	0.81	5.13	1.65	17.31	0.71	0.18
Corrected Q5	-2.53	-0.56	-4.70	-1.66	-11.47	-0.59	-0.17
Corrected Q10	-1.81	-0.29	-2.95	-0.92	-6.16	-0.37	-0.11
Corrected Q25	-1.16	-0.07	-1.45	-0.34	-1.89	-0.17	-0.06
Corrected Median	-0.56	0.07	-0.31	-0.00	0.00	-0.03	-0.01
Corrected Q75	-0.06	0.22	0.79	0.27	1.72	0.12	0.04
Corrected Q90	0.56	0.43	2.62	0.74	7.04	0.35	0.09
Corrected Q95	1.40	0.65	4.18	1.30	15.49	0.65	0.14
N	2256	2256	2256	2256	2256	2256	2256

Note: For individuals with fewer than 2 interruptions, parameters η_{i3} and η_{i4} are normalized as they are not identified, and this normalization contaminates descriptive statistics. "Corrected" statistics are obtained after bias correction as described in the Online Appendix.

Table S.2: Descriptive statistics on distributions of uncorrected and corrected parameters, individuals in employment 20 years and more

	η_{i0}	η_{i1}	η_{i2}	η_{i3}	η_{i4}	δ_1	δ_2
Mean	-0.48	0.07	-0.36	-0.03	0.23	-0.03	-0.01
Variance	0.79	0.05	2.72	1.04	122.86	0.05	0.00
Corrected Variance	0.60	0.03	1.89	0.74	88.20	0.03	0.00
Q5	-1.62	-0.23	-2.74	-1.06	-8.65	-0.35	-0.10
Q10	-1.26	-0.12	-1.91	-0.53	-3.11	-0.23	-0.07
Q25	-0.86	-0.01	-0.91	-0.11	0.00	-0.10	-0.04
Median	-0.49	0.06	-0.29	0.00	0.00	-0.03	-0.01
Q75	-0.13	0.15	0.24	0.10	0.00	0.04	0.02
Q90	0.31	0.28	1.09	0.47	2.70	0.17	0.05
Q95	0.68	0.40	1.91	0.97	8.88	0.29	0.08
Corrected Q5	-1.44	-0.16	-2.37	-0.81	-6.34	-0.29	-0.09
Corrected Q10	-1.16	-0.07	-1.62	-0.45	-2.76	-0.19	-0.06
Corrected Q25	-0.82	0.00	-0.82	-0.09	0.00	-0.09	-0.03
Corrected Median	-0.49	0.06	-0.29	0.00	0.00	-0.03	-0.01
Corrected Q75	-0.17	0.13	0.16	0.10	0.00	0.03	0.01
Corrected Q90	0.18	0.25	0.78	0.38	2.67	0.13	0.04
Corrected Q95	0.52	0.35	1.46	0.63	6.79	0.24	0.07
N	4748	4748	4748	4748	4748	4748	4748

Note: For individuals with fewer than 2 interruptions, parameters η_{i3} and η_{i4} are normalized as they are not identified, and this normalization contaminates descriptive statistics. "Corrected" statistics are obtained after bias correction as described in the Online Appendix.

Table S.3: Descriptive statistics on distributions of uncorrected and corrected parameters, individuals in employment 10-14 years

	η_{i0}	η_{i1}	η_{i2}	η_{i3}	η_{i4}	δ_1	δ_2
Mean	-0.67	0.08	-0.37	0.02	-0.56	-0.02	-0.01
Variance	45.15	1.88	132.79	7.86	929.70	1.08	0.11
Corrected Variance	29.11	1.30	90.75	5.87	716.40	0.66	0.07
Q5	-8.19	-1.34	-13.81	-2.22	-21.09	-1.27	-0.42
Q10	-5.08	-0.84	-9.23	-1.33	-10.34	-0.78	-0.26
Q25	-2.20	-0.28	-3.36	-0.43	-2.50	-0.31	-0.12
Median	-0.57	0.06	-0.26	0.00	0.00	-0.04	-0.01
Q75	0.93	0.43	2.68	0.39	2.21	0.23	0.09
Q90	3.93	1.09	7.70	1.28	9.85	0.74	0.23
Q95	7.42	1.76	11.94	2.40	20.01	1.26	0.41
Corrected Q5	-6.96	-1.02	-11.50	-1.66	-14.78	-0.83	-0.34
Corrected Q10	-4.09	-0.74	-6.32	-0.93	-7.83	-0.58	-0.20
Corrected Q25	-1.89	-0.21	-2.77	-0.36	-2.25	-0.27	-0.09
Corrected Median	-0.55	0.07	-0.35	0.00	0.00	-0.05	-0.01
Corrected Q75	0.71	0.37	1.99	0.33	1.86	0.18	0.07
Corrected Q90	2.37	0.79	6.26	0.91	6.86	0.53	0.15
Corrected Q95	6.36	1.50	9.11	1.76	14.20	0.87	0.32
N	1896	1896	1896	1896	1896	1896	1896

Note: For individuals with fewer than 2 interruptions, parameters η_{i3} and η_{i4} are normalized as they are not identified, and this normalization contaminates descriptive statistics. "Corrected" statistics are obtained after bias correction as described in the Online Appendix.

Table S.4: Means and medians of η_j estimates under correct specification

Statistics	Means			Medians		
Parameter	η_0	η_1	η_2	η_0	η_1	η_2
True	-0.23	0.075	-0.43	-0.23	0.075	-0.43
Bias	9.7e-05	5.4e-06	-7.4e-05	0.00091	-4.8e-06	0.00054
MAD	0.0019	0.00025	0.0021	0.0041	7e-04	0.0056
RMSE	0.0024	0.00031	0.0026	0.0051	0.00089	0.0071
SD	0.0024	0.00031	0.0026	0.0051	0.00089	0.0071
CP	0.94	0.94	0.95	0.94	0.95	0.95

Note: 8000 individuals, 27 time periods, 1000 replications. "Correct specification": the estimated model coincides with the Data Generating Process. "True": Average true value of parameters; "Bias": Bias; "MAD": Mean absolute deviation; "RMSE": Root mean square error; "SD": standard deviation; "CP": Monte Carlo confidence intervals at a 95% level constructed using the aggregate standard deviation, SD.

Table S.5: Variances of η_j estimates under correct specification

	Raw variance			Bias-corrected		
Parameter	η_0	η_1	η_2	η_0	η_1	η_2
True	0.31	0.019	1	0.31	0.019	1
Bias	0.046	0.00078	0.055	0.0068	9.8e-05	0.0061
MAD	0.046	0.00078	0.055	0.0069	0.00011	0.0069
RMSE	0.046	0.00079	0.055	0.0075	0.00014	0.0084
SD	0.0032	9.4e-05	0.0058	0.0031	9.4e-05	0.0058
CP	0	0	0	0.43	0.83	0.83

Note: See Table S.4.

Table S.6: Quantiles of η_j estimates under correct specification

	Raw quantile			Bias-corrected		
Parameter	η_0	η_1	η_2	η_0	η_1	η_2
10%-Quantile						
True	-0.94	-0.1	-1.7	-0.94	-0.1	-1.7
Bias	-0.048	-0.0035	-0.033	-0.012	-0.00064	-0.006
MAD	0.048	0.0035	0.033	0.013	0.0013	0.011
RMSE	0.048	0.0037	0.035	0.015	0.0017	0.013
SD	0.008	0.0014	0.011	0.0095	0.0015	0.012
CP	0	0.32	0.14	0.77	0.93	0.92
25%-Quantile						
True	-0.6	-0.019	-1.1	-0.6	-0.019	-1.1
Bias	-0.023	-0.0018	-0.017	-0.0059	-0.00044	-0.0033
MAD	0.023	0.0019	0.017	0.0074	0.00096	0.0074
RMSE	0.024	0.0021	0.019	0.0091	0.0012	0.0093
SD	0.0058	0.0011	0.0081	0.0069	0.0011	0.0087
CP	0.021	0.6	0.48	0.88	0.93	0.93
Median						
True	-0.23	0.075	-0.43	-0.23	0.075	-0.43
Bias	0.00091	-4.8e-06	0.00054	2e-04	-4.8e-06	0.00042
MAD	0.0041	7e-04	0.0056	0.0048	0.00073	0.006
RMSE	0.0051	0.00089	0.0071	0.006	0.00093	0.0076
SD	0.0051	0.00089	0.0071	0.006	0.00093	0.0076
CP	0.94	0.95	0.95	0.95	0.95	0.95
75%-Quantile						
True	0.15	0.17	0.25	0.15	0.17	0.25
Bias	0.025	0.0018	0.017	0.0066	0.00039	0.0035
MAD	0.025	0.0018	0.017	0.0079	0.00095	0.0077
RMSE	0.025	0.0021	0.019	0.0095	0.0012	0.0096
SD	0.0058	0.001	0.0084	0.0069	0.0011	0.009
CP	0.01	0.61	0.46	0.83	0.94	0.92
90%-Quantile						
True	0.48	0.25	0.87	0.48	0.25	0.87
Bias	0.047	0.0034	0.033	0.012	0.00064	0.006
MAD	0.047	0.0034	0.033	0.013	0.0013	0.011
RMSE	0.048	0.0037	0.035	0.015	0.0016	0.014
SD	0.0078	0.0015	0.011	0.0093	0.0015	0.012
CP	0	0.35	0.17	0.75	0.93	0.92

Note: See Table S.4.

Table S.7: Means of potential wages under correct specification

Means	All observations				Missing obs.		
Parameter	$t = 1$	$t = 9$	$t = 17$	$t = 25$	$t = 9$	$t = 17$	$t = 25$
True	-0.61	-0.24	0.011	0.086	-0.29	-0.1	-0.13
Bias	2.4e-05	2.8e-05	1.2e-05	-3.5e-05	3.9e-05	-0.00067	-0.00015
MAD	0.00062	0.001	0.0015	0.0018	0.0053	0.0073	0.0083
RMSE	0.00076	0.0012	0.0019	0.0023	0.0066	0.0091	0.011
SD	0.00076	0.0012	0.0019	0.0023	0.0066	0.0091	0.011
CP	0.95	0.94	0.95	0.96	0.95	0.95	0.95

Note: See Table S.4.

Table S.8: Variances of potential wages under correct specification

	Raw variance				Bias-corrected			
Parameter	$t = 1$	$t = 9$	$t = 17$	$t = 25$	$t = 1$	$t = 9$	$t = 17$	$t = 25$
True	0.48	0.18	0.42	0.74	0.48	0.18	0.42	0.74
Bias	0.0043	0.012	0.027	0.039	0.00058	0.0038	0.0073	0.0099
MAD	0.0043	0.012	0.027	0.039	0.00095	0.0038	0.0073	0.01
RMSE	0.0044	0.012	0.028	0.039	0.0012	0.004	0.0077	0.011
SD	0.001	0.0012	0.0027	0.0042	0.001	0.0012	0.0027	0.0042
CP	0.011	0	0	0	0.92	0.14	0.25	0.35

Note: See Table S.4.

Table S.9: Quantiles of potential wages under correct specification

	Raw quantile				Bias-corrected			
Parameter	$t=1$	$t=9$	$t=17$	$t=25$	$t=1$	$t=9$	$t=17$	$t=25$
10%-Quantile								
True	-1.5	-0.79	-0.82	-1	-1.5	-0.79	-0.82	-1
Bias	-0.0057	-0.024	-0.036	-0.041	-0.00048	-0.006	-0.0087	-0.009
MAD	0.007	0.024	0.036	0.041	0.0041	0.0067	0.0097	0.011
RMSE	0.0086	0.025	0.037	0.043	0.0051	0.008	0.012	0.014
SD	0.0065	0.0066	0.0096	0.013	0.0051	0.0053	0.0078	0.01
CP	0.88	0.087	0.055	0.15	0.96	0.8	0.80	0.85
25%-Quantile								
True	-1.1	-0.53	-0.43	-0.49	-1.1	-0.53	-0.43	-0.49
Bias	-0.0019	-0.007	-0.011	-0.012	-0.00031	-0.0026	-0.0037	-0.0035
MAD	0.0033	0.007	0.011	0.012	0.003	0.0036	0.0055	0.0063
RMSE	0.0042	0.0077	0.013	0.014	0.0038	0.0045	0.0068	0.0079
SD	0.0037	0.0033	0.0053	0.0067	0.0038	0.0036	0.0058	0.0071
CP	0.91	0.44	0.43	0.57	0.94	0.88	0.91	0.92
Median								
True	-0.61	-0.24	0.011	0.086	-0.61	-0.24	0.011	0.086
Bias	0.00026	0.0014	0.0015	0.0029	0.00014	0.00075	0.00084	0.0015
MAD	0.0026	0.0025	0.0039	0.0053	0.0026	0.0026	0.0041	0.0052
RMSE	0.0033	0.0032	0.0049	0.0067	0.0033	0.0033	0.0052	0.0066
SD	0.0033	0.0029	0.0047	0.006	0.0033	0.0032	0.0051	0.0064
CP	0.94	0.92	0.94	0.92	0.94	0.95	0.95	0.94
75%-Quantile								
True	-0.14	0.049	0.45	0.67	-0.14	0.049	0.45	0.67
Bias	0.0021	0.0087	0.013	0.015	0.00044	0.0034	0.0048	0.005
MAD	0.0034	0.0087	0.013	0.015	0.0029	0.0041	0.006	0.0069
RMSE	0.0042	0.0093	0.014	0.017	0.0037	0.005	0.0072	0.0085
SD	0.0036	0.0033	0.005	0.0066	0.0036	0.0036	0.0054	0.0069
CP	0.91	0.25	0.22	0.36	0.95	0.84	0.87	0.89
90%-Quantile								
True	0.28	0.31	0.85	1.2	0.28	0.31	0.85	1.2
Bias	0.0038	0.015	0.024	0.025	0.00063	0.0055	0.0074	0.0072
MAD	0.0051	0.015	0.024	0.025	0.004	0.0061	0.0085	0.0096
RMSE	0.0063	0.016	0.025	0.027	0.0051	0.0073	0.01	0.012
SD	0.0051	0.0044	0.0065	0.0091	0.0051	0.0048	0.0073	0.0094
CP	0.88	0.07	0.04	0.21	0.94	0.8	0.83	0.88

Note: See Table S.4.

Table S.10: Means of $\eta_3 x_3 + \eta_4 x_4$ under correct specification

Means		All observations			Missing obs.		
Parameter	$t=1$	$t=9$	$t=17$	$t=25$	$t=9$	$t=17$	$t=25$
True		-0.015	-0.023	-0.028	-0.024	-0.044	-0.091
Bias	0	-5.1e-05	1e-05	1.6e-05	-0.00019	0.00066	0.00014
MAD	0	0.001	0.0015	0.0018	0.0043	0.0066	0.0077
RMSE	0	0.0013	0.0019	0.0022	0.0054	0.0082	0.0097
SD	0	0.0013	0.0019	0.0022	0.0054	0.0082	0.0097
CP	0	0.95	0.95	0.95	0.95	0.95	0.95

Note: See Table S.4.

Table S.11: Variances of $\eta_3 x_3 + \eta_4 x_4$ under correct specification

	Raw variance				Bias-corrected			
Parameter	$t=1$	$t=9$	$t=17$	$t=25$	$t=9$	$t=17$	$t=25$	
True		0.036	0.09	0.2		0.036	0.09	0.2
Bias	0	0.013	0.026	0.037	0	0.0039	0.0065	0.0091
MAD	0	0.013	0.026	0.037	0	0.0039	0.0065	0.0091
RMSE	0	0.014	0.027	0.038	0	0.0042	0.0069	0.0098
SD	0	0.0015	0.0023	0.0038	0	0.0014	0.0023	0.0037
CP	0	0	0	0	0	0.22	0.19	0.32

Note: See Table S.4.

Table S.12: Quantiles of $\eta_3 x_3 + \eta_4 x_4$ under correct specification

	Raw quantile			Bias-corrected		
Parameter	$t = 9$	$t = 17$	$t = 25$	$t = 9$	$t = 17$	$t = 25$
10%-Quantile						
True	-0.2	-0.33	-0.49	-0.2	-0.33	-0.49
Bias	-0.029	-0.049	-0.05	-0.029	-0.049	-0.05
MAD	0.029	0.049	0.05	0.029	0.049	0.05
RMSE	0.03	0.05	0.051	0.03	0.05	0.051
SD	0.0059	0.0069	0.0095	0.0059	0.0069	0.0095
CP	0.001	0	0	0.85	0.65	0.79
25%-Quantile						
True	0	-0.098	-0.18	0	-0.098	-0.18
Bias	0	-0.013	-0.021	0	-0.0043	-0.0078
MAD	0	0.013	0.021	0	0.0058	0.0083
RMSE	0	0.014	0.022	0	0.0073	0.0098
SD	0	0.0045	0.0051	0	0.0059	0.006
CP	0	0.16	0.014	0	0.89	0.75
Median						
True	0	0	0	0	0	0
Bias	0	0	0	0	0	0
MAD	0	0	0	0	0	0
RMSE	0	0	0	0	0	0
SD	0	0	0	0	0	0
CP	0	0	0	0	0	0
75%-Quantile						
True	0	0.05	0.14	0	0.05	0.14
Bias	0	0.012	0.022	0	0.004	0.0084
MAD	0	0.012	0.022	0	0.0058	0.0089
RMSE	0	0.013	0.022	0	0.0073	0.01
SD	0	0.0047	0.0049	0	0.0061	0.0059
CP	0	0.24	0.01	0	0.9	0.71
90%-Quantile						
True	0.14	0.27	0.43	0.14	0.27	0.43
Bias	0.029	0.05	0.05	0.0092	0.016	0.014
MAD	0.029	0.05	0.05	0.011	0.016	0.015
RMSE	0.03	0.05	0.051	0.013	0.018	0.017
SD	0.0059	0.0066	0.0079	0.0094	0.0086	0.0096
CP	0.003	0	0	0.82	0.55	0.71

Note: See Table S.4.

Table S.13: Means of predicted wages under correct specification

Means	All observations				Missing obs.		
Parameter	$t = 1$	$t = 9$	$t = 17$	$t = 25$	$t = 9$	$t = 17$	$t = 25$
True	-0.61	-0.25	-0.011	0.059	-0.31	-0.15	-0.22
Bias	2.4e-05	-2.3e-05	2.2e-05	-1.9e-05	-0.00015	-5.1e-06	-9e-06
MAD	0.00062	4e-04	0.00038	0.00041	0.0028	0.0022	0.0031
RMSE	0.00076	0.00051	0.00047	0.00051	0.0035	0.0028	0.004
SD	0.00076	0.00051	0.00047	0.00051	0.0035	0.0028	0.004
CP	0.95	0.95	0.95	0.95	0.95	0.94	0.95

Note: See Table S.4.

Table S.14: Variances of predicted wages under correct specification

	Raw variance				Bias-corrected			
Parameter	$t = 1$	$t = 9$	$t = 17$	$t = 25$	$t = 1$	$t = 9$	$t = 17$	$t = 25$
True	0.48	0.19	0.42	0.8	0.48	0.19	0.42	0.8
Bias	0.0043	0.002	0.0017	0.0021	0.00058	0.00062	0.00037	0.00049
MAD	0.0043	0.002	0.0017	0.0021	0.00095	0.00066	0.00058	0.00094
RMSE	0.0044	0.0021	0.0018	0.0024	0.0012	0.00078	0.00071	0.0013
SD	0.001	0.00048	6e-04	0.0012	0.001	0.00048	6e-04	0.0012
CP	0.011	0.008	0.2	0.62	0.92	0.73	0.91	0.95

Note: See Table S.4.

Table S.15: Quantiles of predicted wages under correct specification

	Raw quantile				Bias-corrected			
Parameter	$t=1$	$t=9$	$t=17$	$t=25$	$t=1$	$t=9$	$t=17$	$t=25$
10%-Quantile								
True	-1.5	-0.81	-0.84	-1.1	-1.5	-0.81	-0.84	-1.1
Bias	-0.0037	-0.0032	-0.0016	-0.0017	-0.00048	-0.0012	-0.00039	-0.00059
MAD	0.0051	0.0038	0.0033	0.0042	0.0041	0.0029	0.0031	0.004
RMSE	0.0063	0.0046	0.0042	0.0052	0.0051	0.0036	0.0039	0.005
SD	0.005	0.0033	0.0038	0.0049	0.0051	0.0034	0.0038	0.0049
CP	0.88	0.84	0.93	0.94	0.96	0.93	0.95	0.94
25%-Quantile								
True	-1.1	-0.54	-0.44	-0.52	-1.1	-0.54	-0.44	-0.52
Bias	-0.0019	-0.0015	-0.00083	-0.00056	-0.00031	-0.00061	-2e-04	-3.7e-05
MAD	0.0033	0.0022	0.0024	0.0028	0.003	0.0019	0.0023	0.0028
RMSE	0.0042	0.0028	0.0029	0.0035	0.0038	0.0024	0.0028	0.0035
SD	0.0037	0.0023	0.0028	0.0034	0.0038	0.0023	0.0028	0.0035
CP	0.91	0.9	0.94	0.94	0.94	0.94	0.95	0.94
Median								
True	-0.61	-0.25	-0.008	0.072	-0.61	-0.25	-0.008	0.072
Bias	0.00026	8.7e-05	8.1e-05	0.00022	0.00014	2.5e-05	4.5e-05	0.00021
MAD	0.0026	0.0018	0.002	0.0023	0.0026	0.0018	0.002	0.0023
RMSE	0.0033	0.0022	0.0025	0.0029	0.0033	0.0022	0.0025	0.0029
SD	0.0033	0.0022	0.0025	0.0029	0.0033	0.0022	0.0025	0.0029
CP	0.94	0.95	0.95	0.96	0.94	0.95	0.95	0.96
75%-Quantile								
True	-0.14	0.041	0.42	0.66	-0.14	0.041	0.42	0.66
Bias	0.0021	0.0015	0.001	0.00075	0.00044	0.00048	0.00041	2e-04
MAD	0.0034	0.0022	0.0023	0.0027	0.0029	0.0019	0.0022	0.0027
RMSE	0.0042	0.0027	0.0029	0.0034	0.0037	0.0024	0.0027	0.0033
SD	0.0036	0.0023	0.0027	0.0033	0.0036	0.0023	0.0027	0.0033
CP	0.91	0.9	0.93	0.94	0.95	0.94	0.94	0.95
90%-Quantile								
True	0.28	0.3	0.81	1.2	0.28	0.3	0.81	1.2
Bias	0.0038	0.0029	0.0019	0.0013	0.00063	0.00098	0.00061	0.00025
MAD	0.0051	0.0034	0.0032	0.0038	0.004	0.0025	0.0029	0.0037
RMSE	0.0063	0.0042	0.004	0.0048	0.0051	0.0031	0.0036	0.0047
SD	0.0051	0.003	0.0035	0.0046	0.0051	0.003	0.0036	0.0047
CP	0.88	0.82	0.92	0.95	0.94	0.93	0.95	0.95

Note: See Table S.4.

Table S.16: Means and medians of η_j estimates when using the basic model

Statistics	Means			Medians		
Parameter	η_0	η_1	η_2	η_0	η_1	η_2
True	-0.23	0.075	-0.43	-0.23	0.075	-0.43
Bias	-0.013	-0.0026	0.014	-0.013	-0.0022	0.014
MAD	0.013	0.0026	0.014	0.013	0.0022	0.015
RMSE	0.014	0.0026	0.015	0.015	0.0025	0.017
SD	0.0047	0.00054	0.005	0.0071	0.0012	0.0097

Note: 8000 individuals, 27 time periods, 1000 replications. The basic model omits interruptions and factors; the estimated model thus does not coincide with the Data Generating Process. "True": Average true value of parameters; "Bias": Bias; "MAD": Mean absolute deviation; "RMSE": Root mean square error; "SD": standard deviation.

Table S.17: Variances of $\eta_{0,1,2}$ estimates when using the basic model

	Raw variance			Bias-corrected		
Parameter	η_0	η_1	η_2	η_0	η_1	η_2
True	0.31	0.019	1	0.31	0.019	1
Bias	0.11	0.00075	0.14	0.067	-0.00029	0.053
MAD	0.11	0.00075	0.14	0.067	0.00029	0.053
RMSE	0.11	0.00077	0.14	0.068	0.00033	0.054
SD	0.0074	0.00016	0.012	0.0074	0.00016	0.012

Note: See Table S.16.

Table S.18: Quantiles of $\eta_{0,1,2}$ estimates when using the basic model

	Raw quantile			Bias-corrected		
Parameter	η_0	η_1	η_2	η_0	η_1	η_2
10%-Quantile						
True	-0.94	-0.1	-1.7	-0.94	-0.1	-1.7
Bias	-0.11	-0.0058	-0.055	-0.062	-0.0014	-0.0083
MAD	0.11	0.0058	0.055	0.062	0.0019	0.015
RMSE	0.11	0.0061	0.057	0.064	0.0024	0.019
SD	0.011	0.0019	0.015	0.012	0.002	0.017
25%-Quantile						
True	-0.6	-0.019	-1.1	-0.6	-0.019	-1.1
Bias	-0.056	-0.0038	-0.018	-0.033	-0.0015	0.0054
MAD	0.056	0.0038	0.019	0.033	0.0017	0.011
RMSE	0.057	0.004	0.022	0.034	0.0021	0.013
SD	0.0078	0.0014	0.011	0.0084	0.0014	0.012
Median						
True	-0.23	0.075	-0.43	-0.23	0.075	-0.43
Bias	-0.013	-0.0022	0.014	-0.013	-0.0022	0.014
MAD	0.013	0.0022	0.015	0.013	0.0022	0.015
RMSE	0.015	0.0025	0.017	0.015	0.0025	0.018
SD	0.0071	0.0012	0.0097	0.0077	0.0013	0.01
75%-Quantile						
True	0.15	0.17	0.25	0.15	0.17	0.25
Bias	0.029	-0.00092	0.048	0.0062	-0.0032	0.024
MAD	0.029	0.0013	0.048	0.0084	0.0032	0.024
RMSE	0.03	0.0016	0.049	0.01	0.0035	0.027
SD	0.0077	0.0014	0.011	0.0084	0.0014	0.012
90%-Quantile						
True	0.48	0.25	0.87	0.48	0.25	0.87
Bias	0.078	3.9e-05	0.084	0.033	-0.0044	0.037
MAD	0.078	0.0014	0.084	0.033	0.0044	0.038
RMSE	0.078	0.0018	0.085	0.036	0.0048	0.041
SD	0.011	0.0018	0.015	0.012	0.0019	0.016

Note: See Table S.16.

Table S.19: Means of potential wage when using the basic model

Means	All observations				Missing obs.		
Parameter	$t = 1$	$t = 9$	$t = 17$	$t = 25$	$t = 9$	$t = 17$	$t = 25$
True	-0.61	-0.24	0.011	0.086	-0.29	-0.1	-0.13
Bias	-0.00085	-0.014	-0.023	-0.027	-0.036	-0.047	-0.092
MAD	0.0011	0.014	0.023	0.027	0.036	0.047	0.092
RMSE	0.0013	0.014	0.024	0.028	0.038	0.051	0.096
SD	0.001	0.0021	0.0035	0.0051	0.012	0.018	0.028

Note: See Table S.16.

Table S.20: Variances of potential wage when using the basic model

	Raw variance				Bias-corrected			
Parameter	$t = 1$	$t = 9$	$t = 17$	$t = 25$	$t = 1$	$t = 9$	$t = 17$	$t = 25$
True	0.48	0.18	0.42	0.74	0.48	0.18	0.42	0.74
Bias	0.015	0.0054	-0.003	0.059	0.0018	0.0024	-0.0069	0.052
MAD	0.015	0.0054	0.0045	0.059	0.002	0.0026	0.0072	0.052
RMSE	0.015	0.0058	0.0056	0.06	0.0023	0.0031	0.0084	0.053
SD	0.0015	0.0019	0.0047	0.01	0.0015	0.0019	0.0047	0.01

Note: See Table S.16.

Table S.21: Means and medians of $\eta_{0,1,2}$ estimates when using the no-factor model

Statistics	Means			Medians		
Parameter	η_0	η_1	η_2	η_0	η_1	η_2
True	-0.23	0.075	-0.43	-0.23	0.075	-0.43
Bias	-0.038	-0.0019	0.016	-0.021	-0.0016	0.019
MAD	0.038	0.0019	0.016	0.022	0.0018	0.019
RMSE	0.039	0.0021	0.018	0.023	0.0021	0.022
SD	0.0071	0.00091	0.0072	0.0085	0.0014	0.011

Note: 8000 individuals, 27 time periods, 1000 replications. The no-factor model omits factors; the estimated model thus does not coincide with the Data Generating Process. "True": Average true value of parameters; "Bias": Bias; "MAD": Mean absolute deviation; "RMSE": Root mean square error; "SD": standard deviation.

Table S.22: Variances of $\eta_{0,1,2}$ estimates when using the no-factor model

	Raw variance			Bias-corrected		
Parameter	η_0	η_1	η_2	η_0	η_1	η_2
True	0.31	0.019	1	0.31	0.019	1
Bias	0.33	0.007	0.42	0.2	0.0048	0.25
MAD	0.33	0.007	0.42	0.2	0.0048	0.25
RMSE	0.33	0.007	0.42	0.2	0.0049	0.25
SD	0.014	0.00036	0.023	0.014	0.00035	0.023

Note: See Table S.21.

Table S.23: Quantiles of $\eta_{0,1,2}$ estimates when using the no-factor model

	Raw quantile			Bias-corrected		
Parameter	η_0	η_1	η_2	η_0	η_1	η_2
10%-Quantile						
True	-0.94	-0.1	-1.7	-0.94	-0.1	-1.7
Bias	-0.28	-0.022	-0.14	-0.2	-0.014	-0.058
MAD	0.28	0.022	0.14	0.2	0.014	0.058
RMSE	0.28	0.022	0.14	0.2	0.014	0.062
SD	0.017	0.0024	0.02	0.02	0.0026	0.023
25%-Quantile						
True	-0.6	-0.019	-1.1	-0.6	-0.019	-1.1
Bias	-0.13	-0.0087	-0.035	-0.093	-0.0042	0.0092
MAD	0.13	0.0087	0.035	0.093	0.0042	0.014
RMSE	0.13	0.0088	0.037	0.093	0.0045	0.017
SD	0.0099	0.0016	0.013	0.012	0.0018	0.015
Median						
True	-0.23	0.075	-0.43	-0.23	0.075	-0.43
Bias	-0.021	-0.0016	0.019	-0.023	-0.0018	0.02
MAD	0.022	0.0018	0.019	0.023	0.002	0.02
RMSE	0.023	0.0021	0.022	0.025	0.0023	0.023
SD	0.0085	0.0014	0.011	0.01	0.0015	0.012
75%-Quantile						
True	0.15	0.17	0.25	0.15	0.17	0.25
Bias	0.078	0.0055	0.069	0.037	0.00085	0.026
MAD	0.078	0.0055	0.069	0.037	0.0017	0.026
RMSE	0.079	0.0057	0.07	0.039	0.002	0.029
SD	0.0094	0.0017	0.012	0.011	0.0019	0.014
90%-Quantile						
True	0.48	0.25	0.87	0.48	0.25	0.87
Bias	0.19	0.019	0.16	0.11	0.01	0.081
MAD	0.19	0.019	0.16	0.11	0.01	0.081
RMSE	0.19	0.019	0.17	0.11	0.011	0.084
SD	0.014	0.0027	0.018	0.017	0.0029	0.021

Note: See Table S.21.

Table S.24: Means of potential wage when using the no-factor model

Means	All observations				Missing obs.		
Parameter	$t = 1$	$t = 9$	$t = 17$	$t = 25$	$t = 9$	$t = 17$	$t = 25$
True	-0.61	-0.24	0.011	0.086	-0.29	-0.1	-0.13
Bias	-0.023	-0.03	-0.033	-0.028	0.0048	-0.041	-0.038
MAD	0.023	0.03	0.033	0.028	0.016	0.042	0.041
RMSE	0.024	0.03	0.033	0.029	0.02	0.048	0.048
SD	0.002	0.0042	0.0062	0.0066	0.019	0.026	0.03

Note: See Table S.21.

Table S.25: Variances of potential wage when using the no-factor model

	Raw variance				Bias-corrected			
Parameter	$t = 1$	$t = 9$	$t = 17$	$t = 25$	$t = 1$	$t = 9$	$t = 17$	$t = 25$
True	0.48	0.18	0.42	0.74	0.48	0.18	0.42	0.74
Bias	0.15	0.14	0.25	0.32	0.13	0.12	0.2	0.24
MAD	0.15	0.14	0.25	0.32	0.13	0.12	0.2	0.24
RMSE	0.15	0.14	0.25	0.32	0.13	0.12	0.2	0.25
SD	0.0046	0.0067	0.011	0.015	0.0045	0.0066	0.011	0.015

Note: See Table S.21.