WORKING PAPERS

N° 1041

October 2019

"Heterogeneity and Wage Inequalities over the Life Cycle"

Thierry Magnac and Sébastien Roux

Toulouse School of Economics

Heterogeneity and Wage Inequalities over the Life Cycle

Thierry Magnac*and Sébastien Roux†

This version: 30 September 2019

Abstract

Using panel data from a single cohort of French male wage earners observed over a long span of 30 years starting at their entry in the labor market, we estimate parameters of a human capital investment model by random and fixed effect methods. Individual wage profiles are described by their individual-specific level, slope and curvature. This allows a fine decomposition of the variance of (log-)wages at different times of the life-cycle and in the long run. Among salient results, short run time-varying inequalities are shown to be larger that long run inequality by a factor of 20% to 80%. Individual permanent heterogeneity explain between 60 to 90% of the variance of wages. Single dimensional heterogeneity explains well those variances at a point in time but not over the whole period or in the long run. Multidimensional heterogeneity is needed and in particular under the form of a horizon individual effect.

JEL Codes: C33, D91, I24, J24, J31

Keywords: human capital investment, inequality, wage dynamics, post-schooling wage growth, random and fixed effects

*Toulouse School of Economics, Université de Toulouse Capitole, c*orresponding author*, thierry.magnac@tse-fr.eu

†Insee-Crest-Ined, Paris, sebastien.roux@insee.fr

1 Introduction[1]

Income inequalities have become again the focus of many debates in economics as revived by the work of Piketty (2013) and coauthors on the brutal increase of top incomes in the last 50 years in most countries. More contrasted is the evidence on earnings inequalities at the country level, across developed countries (Atkinson and Morelli, 2014) or at the world level (Milanovic, 2016). Fewer studies look at the building up of inequalities over the life-cycle (Lagakos, Moll, Porzio, Qian and Shoellman, 2018) although this is an important component of wage inequalities.

Many competing theories account for the shape of wage profiles over the life-cycle. Investments in human capital, learning by doing or job search are the most popular (Rubinstein and Weiss, 2006). It is difficult, however, to test them apart, in particular when empirical researchers allow parameters governing these models to be individual specific. These theories are helpful in disciplining the specification of wage profiles and the interpretation of empirical results, when estimating models of wage profiles and by consequence, life-cycle inequalities.

In this paper, we use a human capital investment setup à la Ben Porath (1967), as developed in Magnac, Pistolesi and Roux (2018), to specify wage profiles. Analyzing wages, and not earnings, allows to abstract from labor supply issues and to remain closer to the skill-building view of human capital. In the empirical application proposed here, we use French administrative data on wage profiles of a single large cohort of around 7500 male workers, entering the labour market in 1977, and followed until 2007. Focusing on a single cohort allows to concentrate on life cycle issues, since France is one of the countries in which earnings inequality was stable over these years (Atkinson and Morelli, 2014) – at the population level at least, in contrast with top incomes. We show that the cohort under study has no specificities and we control for aggregate effects using a flat spot approach (Heckman, Lochner and Taber, 1998, Bowlus and Robinson, 2012).

Our theoretical set-up leads to an empirical factor model in which an individual wage profile are described by three individual-specific parameters, a level, a slope and a curvature. The slope is expected to be positive, and the curvature negative, since profiles are generally increasing and concave. Pervasive heterogeneity is simpler to deal with, in this linear framework, than in non-linear ones (e.g. Browning, Ejrnaes and Alvarez, 2012, Polachek, Das and Thamma-Apiroam,

[1]We thank Christian Belzil, Richard Blundell, Laurent Gobillon, Jim Heckman, Nicolas Pistolesi and Bernard Salanié for helpful comments as well as participants in numerous seminars where we presented earlier versions of this research This research has received financial support from the European Research Council under the European Community's Seventh Framework Program FP7/2007-2013 grant agreement N°295298 and funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program). All errors remain ours.

2015). For the remaining individual-and-time shocks, we fit a general ARMA specification and do not take a stand on how this would be decomposed into persistent and transitory components since consumption data are not available and the decomposition cannot be identified (Ejrnæs & Browning, 2014).

Our first contribution is to decompose wage profiles in aggregate effects due to groups of skills (age at entry and initial skill occupation) and unobserved individual effects due to specific human capital investments. In contrast with the US (Heckman, Lochner and Taber, 1998), we find that between and within group covariances are very similar.

Our second contribution is to propose a sequential estimation method by random effects, first, and fixed effects, second. There are good econometric reasons for estimating by random effect methods the global characteristics of wage profiles (Alvarez & Arellano, 2004) and this might explain why estimates using covariance matrices of log wages and minimum distance might yield different results (Hryshko, 2012). A random effect method deals with *(i)* issues of initial conditions, quite out of the stationary path in the case of wages, delivers *(ii)* an estimate of serial correlation at the population level and provides *(iii)* an estimate of the covariance of individual effects. As a consequence of these three arguments, random effects discipline the estimation of fixed effects. Random effects however do not deliver other moments than the variance of the distribution of individual effects, while fixed effects do at the price of a $1/T$ bias which can be bias-corrected (Arellano and Bonhomme, 2012).

Our third set of contributions is empirical. First, we find that the variance of the long-run value of a wage profile, as accounted by discounted sums of log-wages over the (observed) life-cycle, is about the same as the cross-section variance of log wages after 5 years of the life-cycle but only 60%, after 30 years. This is in line with estimates of Bonke, Cormeo and Lutken (2014) using German data in which they find that the former is about 2/3 of the latter. Second, we find that heterogeneity explains 70% of the variance at the beginning of the life-cycle and 90% after 30 years which is in line with Keane and Wolpin (1997) but larger than what is obtained by Huggett, Ventura and Yaron (2011) or Bagger, Fontaine, Postel-Vinay and Robin (2014) in admittedly less heterogeneous set-ups. Third, we find that the relative explanatory power of observed skills and unobserved individual heterogeneity changes over the life-cycle, from half and half after five years in the labor market, to a 30-70% decomposition after 30 years. The persistence of observed heterogeneity is more noticeable though in the long run since observed heterogeneity account for more than 50 percent of the long run inequality. Fourth, a single dimension heterogeneity term does not describe well the variance of log wages in cross sections.

The fixed individual effect in level explains well the variance at the beginning of the life cycle but not at the end. In contrast, the individual slope and curvature of profiles explain well the variance at the end of the life cycle but not at the beginning.

This is why inequality in the long run cannot be explained by a single heterogeneity component but the three (levels, slope and curvature) are needed and in particular, for the high skill group. We also show that returns are correlated negatively with the initial wage level, purged from transient effects, as in Gladden and Taber (2009) or Sorensen and Vejlin (2014) but this result is not uniform over the life-cycle. This is only the case over the 20 first years of the life-cycle, this correlation turning positive afterwards, because of the heterogeneous curvature of the wage profiles. Allowing for an horizon effect, that the curvature of the profile picks out, is an innovation in this literature and seems to be key in these impacts.

Literature review It is useful to start with a brief comparison with the extensive empirical literature on earnings dynamics (see Meghir and Pistaferri, 2010, 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 like the one described as heterogeneous income profiles (HIP) or restricted income profiles (RIP). Up to now, there is no consensus in the literature about which specification fits the data best (see e.g. Baker, 1997, Guvenen, 2007, Hryshko, 2012 and Hoffmann, 2019). Our linear factor structure embeds both models since the permanent component includes individual specific levels and growth rates of earnings as HIP does and the stochastic component can be any mixture of permanent and transitory shocks like in RIP. Nonetheless, our three factor structure invalidates the key identifying assumption about the correlations between first differences of within shocks (for instance Blundell, 2014) because of the presence of the curvature term.

Our paper also touches the estimation of the traditional homogeneous wage equation (Mincer, 1974). The state-of-the-art study is Lagakos et al. (2018) which studies an impressive set of countries and shows that experience-wage profiles are twice as steep in rich countries as in poor countries. Furthermore, more educated workers have steeper profiles. What we observe in our administrative data is similar although other studies (Engbom, 2017), using survey data (EHCP and SILC), finds that wage growth in France is relatively small among 12 OECD countries (same as in Germany but less than in the US and the UK).

There has been recently some non-linear alternative proposals to the previous literature such as Browning et al. (2012), Hospido (2012), Song, Price, Guvenen, Bloom, and Von Wachter (2018) or Bonhomme and Robin (2009) and Pora and Wilner (2017) using French data. There are also semiparametric analyses such as Lochner and Shin (2014) and Arellano, Blundell and Bonhomme (2018) using US data which have different characteristics from the data we use since the returns to observable components increased dramatically in the US (Autor, Katz and Kearney, 2008). It is generally difficult to compare these non linear estimations with ours because our linear model is designed to capture means and covariances, while using pervasive heterogeneity. The bridge between those methods could be the generalization of our procedures to the estimation of quantiles.

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) takes stock of the literature before the 2000, and distinguishes job search and human capital theories by some of their predictions. Job search models predict a negative correlation between wage and subsequent wage growth over the life-cycle, while human capital models predicts that it is negative at the beginning of the life cycle but turns positive afterwards. The latter is what we find in our empirical analysis. Some recent literature models explicitly job search, in contrast with this paper in which we treat job search as a transient residual cause. Bowlus & Liu (2013) decomposes earnings growth into human capital (50%), job search (20%), the rest being their interaction. In contrast, Bagger et al. (2014) finds that job search, or "job-shopping", significantly contributes to wage growth but seems to be mostly occurring over the first ten years of the working life. Furthermore, Burdett, Carrillo-Tudela and Coles (2016) finds that most of the effect of experience on wages is due to passive learning-by-doing.

Contratsing human capital investments and learning by doing is the objective of fewer papers. Heckman, Lochner and Cossa (2003) show that distinguishing job training and learning by doing might use that wage subsidies, such as EITC, provide additional incentives to work, enhance learning by doing and decrease investments in human capital. Belley (2017) contends that learning by doing does not seem prevalent since it does not predict the trade-off between current and future earnings, observed in the data and a prediction of human capital models. Blandin (2018) also points out that learning by doing does not predict a decrease in investments at the end of the working life.

Finally, within the human capital paradigm, Sanders and Taber (2011) reviews models with multidimensional human capital. Sorensen and Vejlin (2014) estimates the correlation between initial wages and later wage growth as well as its non parametric equivalent using Danish data over 20 years. They, as well as Gladden and Taber (2009), find that this correlation is negative. We find the same negative correlation over the first 20 years but it turns positive afterwards. Our approach differs from theirs in two aspects. First, we have a 3-factor linear model in which the horizon effect is key. Second, these authors use the observed initial wage, which is transient at this age, while we try to filter out these transient initial conditions. We use the reconstructed non-transient initial log-wage by using a combination of random and fixed effects.

Section 2 briefly describes the evolution of earnings inequality in France and the data we use. Section 3 details our empirical strategy and Section 4 the econometric methods. Section 5 reports estimation results and Section 6 gathers the results of various decompositions of life-cycle inequalities.

2 A Brief Description of the Data

We briefly summarize the evolution of wage inequality in France over the last 40 years and present stylized facts about means, variances and autocorrelations of log wages in our sample, after reporting how we constructed this sample from administrative sources.

2.1 Earnings and Wage Inequality in France

The sharp increase in earnings inequality in the UK and in the US over the last thirty years is a well known empirical fact (see for example Autor, Katz and Kearney, 2008, or Moffitt and Gottschalk, 2011, for the US and Blundell & Etheridge, 2010, for the UK). Yet, the picture is more balanced in other OECD countries and while some European countries have experienced an increasing dispersion in earnings, others have not been affected by this trend and have had stable or decreasing dispersion. Atkinson and Morelli (2014) compute international earnings inequality comparisons over the second half, or so, of the twentieth century for 25 countries. As regards European economies, they conclude that earnings inequality has increased in Germany, Italy, Portugal, Sweden, Switzerland while in Finland, France, the Netherlands, Norway, and Spain earnings dispersion has stayed constant or decreased over this period.

In France, earnings inequality in 2010 is broadly comparable to its level in the sixties and if anything has decreased. Atkinson and Morelli (2014) report an unchanged Gini coefficient for earnings over the period. Using Labour Force Surveys (LFS), they also compute yearly measures of inequality and show a very stable inequality level. Using two different datasets, the DADS, used here, and the French LFS, Verdugo (2014) concludes that the two data sets provide strikingly similar figures of constant or decreasing earnings dispersion between 1964 and 2005. Verdugo (2014) decomposes the total earnings dispersion into upper and lower-tail earnings inequality. The dispersion at the top of the distribution remained constant, since the P90/P50 index in earnings fluctuates around 2, while the dispersion at the bottom measured by the P50/P10 index decreased from 1.9 to 1.5. Charnoz, Coudin and Gaini (2011) also use the DADS data to reach the same conclusion that earnings inequality in France has been rather stable from 1976 to 1992 and has been slightly decreasing from 1995 to 2004. This stability has been attributed, at least partly, to a strong policy driven increase in education at the end of the 1980s and labor market policy regulations at the end of the 1990s (Charnoz, Coudin and Gaini, 2014).

A note of caution is in order. While these studies consider changes in the cross-sectional earnings distributions, changes in the structure of the population that play an important role in the previous studies, are neutralized here. We follow a single cohort of individuals entering the labor market in 1977 to focus on inequalities unraveling over the life-cycle.

2.2 Our Working Sample

Our panel dataset on wages is extracted from a French administrative source named *Déclarations Annuelles de Données Sociales* (DADS) which has been used for employee-employer studies as Abowd, Kramarz and Margolis (1999).[2] DADS data is collected through a mandatory data requirement for social security and tax verification purposes. All employers must send to the social security and tax administrations the list of all persons who have been employed in their establishments during the year. Firms report the full earnings they have paid to every employee and payments exclude other labor costs borne by the firm. Each person is identified by a unique individual social security number which facilitates the follow-up of individuals through time although we cannot reconstruct taxes they pay. The tax system is household-based in France and the linking of this dataset with fiscal records is not authorized yet.

The French National Statistical Institute (INSEE) has been drawing, since 1976, a sample from this dataset at a sampling rate of around 4% by retaining all individuals who were born in October of even years. Using administrative data is an important advantage since these data are less subject to attrition or measurement errors. Unlike survey data, the collection of information

[2] These data are accessible through securitized access (CASD). Other contributions in the earnings literature that use administrative data is Hoffmann (2019) and Daly, Hryshko and Manovskii (2016) (German and Danish data).

does not rely on individual response behavior and individuals are better followed over time. Moreover, the large sample size enables us to use a single large cohort of individuals who entered the labor market in the same year.

This dataset is restricted to individuals employed in the private sector or in publicly-owned companies and we consider only males to lessen selection issues. Observations can yet be missing for different reasons. Data were not collected in three years (1981, 1983 and 1990) for reasons specific to INSEE. It is also quite frequent that employees exit the panel and a significant fraction of those reenter it after a few years (see Table S.i in the Supplementary Appendix). Those absences might stand for spells in the public sector, as self-employed, or out-of-employment. We also code as missing part-time employment in any given year.

We restrict our analysis to labor market entrants in 1977, a set that we call "cohort" in the following, even if their age is heterogenous. Entrants are defined as those who started working full time for more than 6 months in 1977, and are still employed the following year, possibly in a different firm. To make sure that these employees have a permanent attachment to the private sector and to mitigate the issue of missing years in 1981 and 1983, we keep only those who also work in 1982 and 1984 and who were aged between 16 and 30 in 1977. The distribution functions of unobserved factor loadings, or idiosyncratic components, that we estimate in the following refer to this subpopulation. Moreover, lacking a credible identification strategy to correct for selection, we shall assume that missingness is at random.

We define wages as the sum of all earnings during a year, divided by the number of days worked. This allows employees to have within-year periods out of the private sector. Shortcomings of administrative data are that other components of income, including other sources of earnings, are missing and that few observable characteristics are available, apart from age at labor market entry and the skill level of the first job.

First, it is likely that workers delaying entry have a higher education level than the ones who entered earlier. Second, initial skills are grouped into three categories based on a two-digit codification: high-skill (managers, professionals), medium-skill (blue-collar or white-collar skilled workers) and low-skill jobs. Our 20 resulting "education" groups are defined by the interaction between these two variables when groups are not too small (see e.g. Table 2 for the definition and size of each group). Since education is defined according to characteristics recorded at labor market entry, individuals are attached to the same education group during their whole working life.

Other details completing this description can be found in Data Appendix A.

2.3 Wage dynamics: a single cohort

Table 1 reports descriptive statistics on the composition of the sample over time. The sample size is 7446 observations in 1977 and 4670 in 2007. Human capital groups defined above are of unequal size, the groups with an early age of entry being the largest ones, and with a late age of entry, the smallest. Attrition follows a somewhat irregular pattern due in part to the original data and to our sampling design (see Supplementary Appendix, Table S.i). There are also more surprising features for instance in 1994 (or 2003 at a lesser degree) a year in which many observations are missing. This is due to the way INSEE reconstructed the data from the information in the original files.

We report in Figure 1, the evolution of average log-wages over the life cycle, in 2007-euros, for three age of entry groups (< 20, ≥ 20 and ≤ 24 and > 24) and these profiles display the familiar increasing and concave shape.

By taking deviations of (log) wages with respect to their means in the groups defined by age of entry, skills and years, we compute log-wage residuals. The left panel of Figure 2 plots the change in the cross-sectional variance of those (log) wage residuals for the full sample, while the right panel graphs the variance by age of entry groups. Choosing the variance as a description of the process is adapted to the random effect specification that we estimate. Using other inequality measures (Gini , Theil or Atkinson) does not change the qualitative features of our descriptions. The first few years witness a strong variability of wages. Until the sixth year of observation, 1982 (respectively the fourth, 1980), the variance of log wages drops for the low skill groups (resp. for the other groups) whereas it increases gradually over the rest of the sample period till around 1995. The variance profile is flat afterwards in contrast to the US where it continues to grow (e.g. Rubinstein and Weiss, 2006, using PSID). From the right panel one can notice that late entrants in the labor market experience higher levels and larger rates of growth for the variance of log-wages over the life-cycle as in most countries (Lagakos et al, 2018).

The covariance matrix of log-wage residuals is reported in the Supplementary Appendix Table S.ii although this is easier to use graphs to describe the main features of wage autocorrelations. Figure 3 displays the autocorrelation of residuals of log wages in year t with residuals in an early (resp. late) year, 1986 (resp. 2007). This Figure reveals an asymmetric pattern over time which is quite robust to the choice of these specific years (1986 and 2007). The correlation between wages in year t and in 1986 is swiftly increasing when t is before 1986 and this is also true for 2007 albeit at a lesser degree. In contrast, the correlation between wages in 1986 and in year t is only slowly decaying in t, if time t postdates 1986. Figure 4 takes a different view that confirms the previous diagnostic by plotting the autocorrelations of order 1 and 6. Note that their shape are very similar and increase uniformly over time although at different levels. The closer to the end of the working life, the larger the autocorrelation coefficients are.[3] This provides strong evidence that wages are becoming more stable as employees progress in their life-cycle.

3 Empirical strategy

We specify a linear factor model of wage dynamics. Our starting point is the model of human capital investments after leaving school that is developed in Magnac et al. (2018) and that provides exact theoretical foundations à la Ben Porath (1967), for a Mincer (1974) reduced form wage equation at the individual level. This specification is more tractable to estimate, than alternatives when heterogeneity is pervasive (Browning, Ejrnæs and Alvarez, 2012, Polachek et al., 2015). Magnac et al. (2018) spell out the conditions under which the (log) wage equation can be written as a linear factor model where the three observed factors are $f_t = \left(1, t, \beta^{-t}\right)$:

$$\ln w_{it} = \eta_{i1} + \eta_{i2} t + \eta_{i3} \beta^{-t} + v_{it}, \tag{1}$$

in which $\beta < 1$ is the assumed homogenous discount factor. This reduced form delivers the familiar increasing and concave shape when $\eta_{i2} > 0$ and $\eta_{i3} < 0$.

When wages are appropriately discounted, as in the flat spot approach that we describe below, the first three terms measure the logarithm of the current human capital stock (net of current investments) while v_{it} can be interpreted as the logarithm of the price of human capital, net of accumulated depreciations. Implicitly, the influence of job search, the job ladder, or dismissals, are hidden in the latter component.

Factor loadings or individual specific effects η_{i1}, η_{i2} and η_{i3} have a structural interpretation. The first two ones are related respectively to the initial level of human capital i.e. the ability to earn (see Browning et al., 1999) and to the rate of return to investments i.e. the ability to learn. Furthermore, the ratio between η_{i3} and η_{i2}, i.e. the ratio of the "curvature" relative to the growth of log wages, can be structurally interpreted as the value given to human capital at an arbitrary terminal period. The longer the horizon of investment, the smaller the curvature (Lillard and Reville, 1999). Note that this model nests the Heterogenous Income Profile (HIP)

[3] In the Supplementary Appendix, graph S.ii dispalys that this cohort has nothing specific when compared to younger cohorts entering later into the labor market

and the Restricted Income Profile (RIP) as in Guvenen (2007) since v_{it} can be any time-series process.

Estimating this very heterogenous reduced form leads to the decomposition of the wage profile heterogeneity in each of its structural components. Deriving the relative importance of each component is the object of Section 6 which extends to an heterogeneous setting, the calibration exercise of Huggett et al. (2011) using US data.

Before that, the first step of our empirical strategy decomposes the net log price of human capital, v_{it}, into aggregate and individual specific components. Aggregate components are constructed using the subsamples defined by the education groups that we constructed before, from skill and age at labor market entry. They can be interpreted as market (log) prices net of depreciation for education "types" when we adopt the framework of Heckman et al. (1998), in which human capital stocks of different education groups are imperfect substitutes in the aggregate production function. In contrast, perfect substitution holds within groups, and individual specific shocks are interpreted as frictions.

The mechanisms that underlie the specific dynamics of aggregate and individual specific components are allowed to differ, and are left unrelated. In consequence, we handle education group and individual specific effects separately and recompose them afterwards to recover the full effects.

3.1 Aggregate components

Equation (1) can be linearly aggregated, in each education group, as:

$$\overline{\ln y}_{gt} = \overline{\eta}_{g1} + \overline{\eta}_{g2} t + \overline{\eta}_{g3} \beta^{-t} + \bar{v}_{gt}, \tag{2}$$

in which g denotes a group defined by skill and age of entry and $\overline{\eta}_{gk} = E(\eta_{ik} \mid i \in g)$ for $k = 1, 2$ or 3, $\bar{v}_{gt} = E(v_{it} \mid i \in g)$. The term $\bar{v}_{gt}$ stands for the market log-prices of human capital of group g at time t.

Aggregate log prices are given by :

$$E(\bar{v}_{gt} \mid f_t = \left(1, t, \beta^{-t}\right)) = \varphi_{gt}, \tag{3}$$

in which φ_{gt} is an estimable series. Condition (3) requires that the net log price dynamics is driven by factors orthogonal to the ones which govern average human capital accumulation and is the key restriction that separates quantities from prices. To identify φ_{gt} (up to a constant term), Heckman et al. (1998) and Bowlus and Robinson (2012) use a "flat spot" condition by which (2) is satisfied with $\overline{\eta}_{g2}t + \overline{\eta}_{g3}\beta^{-t} \simeq 0$ for a restricted window of periods close to the end of the working life (around 50). This means that investments and depreciation shocks exactly balance each other at those ages. This is a structural way of solving the well known impossibility of separately identifying age, cohort and time effects in a linear setting. We apply this flat spot technique and assess robustness of our results by using various wage or productivity deflators as estimates of φ_{gt}. The stability of the between and within group wage inequality in France over this period supports the credibility of assumption (3).

3.2 Individual specific components

Turning to within group variation, define centered individual effects as their deviations with respect to their means, $\eta_{ik}^c = \eta_{ik} - \overline{\eta}_{gk}$, for $k = 1$, 2, or 3 and $v_{it}^c = v_{it} - \bar{v}_{g(i)t}$. Centering the log wage equation (1) yields log-wage residuals:

$$u_{it} = \ln y_{it} - \overline{\ln y}_{g(i)t} = \eta_{i1}^c + \eta_{i2}^c t + \eta_{i3}^c \beta^{-t} + v_{it}^c, \tag{4}$$

in which u_{it} is the deviation of individual log-wages with respect to their group averages. Individual specific deviations, v_{it}^c, stand for frictions in a model of search and mobility (see e.g. Postel-Vinay and Turon, 2010). The dynamics of the wage process is indeed partly bounded from below and from above by two processes which are individual productivity in the current match and outside offers that the agent receives while on the job. At times, bounds are binding and wages are: Either equal to the productivity process because adverse shocks on that process make employee and employer renegociate the wage contract. Or equal to the outside offer in the case the employee can either renegociate with his employer or take the outside offer if the productivity is lower that the outside option.

These frictions are here described by a stochastic process which is mean independent of factors and factor loadings:[4]

$$E(v_{it}^c \mid f_t = \left(1, t, \beta^{-t}\right), \eta_i^c) = 0. \tag{5}$$

Note that it lets other moments of v_{it}^c depend freely on factors and individual effects η_i^c. The mean independence of frictions, v_{it}, with respect to factors, f_t, and individual effects, η_i^c, is the main assumption underpinning the identification of individual-specific structural parameters.

[4] We slightly relax the assumption of mean independence of frictions with respect to individual effects by authorizing general initial conditions in the random effect model that we estimate, in the empirical application below.

The issue of missing data is potentially important in our empirical application since we observe wages if and only if individuals are employed by the private sector. Missingness could be due to periods spent in the public sector, as self-employed or as non employed. Absent credible instruments or structural assumptions, we assume in the following that missingness is at random. The existence of missing data becomes unconsequential for consistently estimating aggregate effects while we show below that the random effect likelihood approach we adopt, deals with missing data easily.

4 Econometric method

In this section we summarize our econometric estimation method, step by step. Most of the technical details are relegated to appendices. Our main objective is to recover estimates of individual effects $(\eta_{i1}, \eta_{i2}, \eta_{i3})$ in the linear factor structure (1) by using identifying restrictions presented in Section 3.

Stacking log wages $\ln y_{it}$ and the stochastic component v_{it} into $T \times 1$-vectors $\ln y_i$ and v_i as well as η_{ik} into a 3×1-vector η_i, equation (1) writes:

$$\ln y_i = M(\beta)\eta_i + v_i \tag{6}$$

in which $M(\beta)$ is a $T \times 3$ matrix in which a constant, a linear and a geometric term are stacked.

As explained above, we split the estimation in two stages. First, we estimate aggregate equation (2) group by group. At this aggregate level, we have 28 observations per group and those have their own aggregate dynamics. This is why we estimate parameters in each group by simple OLS as justified by condition (3). This provides consistent estimates of $\overline{\eta}_g$, say $\widehat{\overline{\eta}}_g$, and standard errors are computed using a Newey West procedure.

Second, within-group residuals, u_i, can be expressed as a function of centered individual specific parameters, η_i^c as:

$$u_i = M(\beta)\eta_i^c + v_i^c. \tag{7}$$

in which $v_i^c = v_i - \overline{v}_{g(i)}$. Because some data are missing, inference might be poor if we estimate this last equation individual by individual since there are at most 28 observations in our data. Let $T_i \leq T$ denote the number of observations for individual i, estimates of parameters η_i for every individual profile are consistent when $T_i \to \infty$ but, in small samples, the bias is of the order $1/T_i$ (e.g. Arellano and Bonhomme, 2012). To overcome this difficulty, η_i^c are estimated using a two-step strategy which consists first in estimating a flexible random effect model using the whole sample and second, in estimating equation (7) by FGLS, individual by individual. In the second step fixed effect procedure, the FGLS weight is the inverse of the population covariance matrix of v_i^c over time estimated in the first random effect step.

Arguments underlying such an estimation strategy are based on trading off the consistency properties of random effect methods when the time dimension is small and the flexibility of the fixed effect methods. On the one hand, estimating serial correlation by random effects in a first step, allows general serial correlation to be controlled for, in the fixed effect estimation, so that the latter is presumably more precise. Gaining precision is also likely even if the random effect specification is only an approximation of a more complicated data generating process. Furthermore, random effect estimation provides a benchmark against which we can assess the amount of bias in the fixed effect estimation due to the finite length of the observation period for each profile. On the other hand, random effects methods provides estimates of means and covariances only. Fixed effect estimates bring about richer information on the underlying distribution of individual specific parameters.

Combining aggregate and individual specific estimates yields fixed effect estimates of the original factor loadings:

$$\hat{\eta}_i = \widehat{\overline{\eta}}_g + \hat{\eta}_i^c.$$

We now present random and fixed effect methods in more detail.

4.1 Random Effects

Equation (7) and mean independence restrictions (5) lead to:[5]

$$\begin{aligned} E(u_i \mid \eta_i^c) &= M(\beta)\eta_i^c, \\ V(u_i \mid \eta_i^c) &= V(v_i^c \mid \eta_i^c) \equiv \Omega(\eta_i^c), \end{aligned}$$

and:

$$V(u_i) = V(E(u_i \mid \eta_i^c)) + E(V(u_i \mid \eta_i^c)) = M(\beta)V(\eta_i^c)M(\beta)' + E(\Omega(\eta_i^c)). \quad (8)$$

Our parameter of interest in this equation is the covariance matrix of the individual effects, $V(\eta_i^c)$ standing for the covariances between level, slope and curvature parameters. Identifying the covariance matrix requires restrictions on the average variance of the idiosyncratic errors,

[5] Appendix B presents the full specification of the process for v_{it}^c in which we also deal with initial conditions in the most general way used in the dynamic panel data literature. The covariance matrix of initial conditions is free as well as the covariance between those initial conditions and the individual specific parameters η_i^c. Further details are given in the supplementary Appendix S.II.

$E(\Omega(\eta_i^c))$. An ARMA specification is common in the earnings dynamic literature and generally low orders are used (see Guvenen, 2009, or Hryshko, 2012) whereas an alternative is a composition of permanent and transitory shocks with specific structures (Blundell, 2014, Bonhomme and Robin, 2009, Lochner and Shin, 2014). We refrain from decomposing v_i^c into its persistent and transitory components since those are not identified absent additional restrictions and consumption data (Ejrnæs and Browning, 2014).

Arellano and Bonhomme (2012) show that a finite lag ARMA specification is sufficient to identify $V(\eta_i^c)$. We use this result and proceed by specifying that the processes v_{it}^c belong in the family of time-heteroskedactic ARMA processes although we limit the orders of the AR and MA to vary between 1 and 3. This allows the robustness of the estimated covariance of individual effects, $V(\eta_i^c)$, to the orders of the ARMA process to be assessed. Moreover, we allow for time heteroskedasticity of the innovations whose importance is argued by Alvarez and Arellano (2004). What the decomposition (8) shows in addition, is that a restricted form of individual heterogeneity, possibly dependent on parameters, η_i, could be allowed in the ARMA process, provided that the expected value, $E(\Omega(\eta_i^c))$, remains in the ARMA family that we consider.

The most commonly used minimum distance method for estimating equation (8) is severely small-sample biased since the range of moments involved when the time dimension becomes large makes first order asymptotics a poor guide in empirical research. Okui (2009) derives the small sample biases not only in the mean but also in the variance of GMM estimates due to the presence of too many moments and he suggests some moment selection mechanism. This is why some researchers proposed to return to an OLS set up adding a bias correction step (Fernandez-Val and Weidner, 2018) or to maximum or pseudo-maximum likelihood methods that reduce the number of moments available (Alvarez and Arellano, 2004).

Specifically, the estimation method proposed by Alvarez and Arellano (2004) seems to dominate in Monte Carlo experiments other fixed T consistent estimators such as the maximum likelihood estimator using differenced data, and the corrected within group estimator. This method is particularly well adapted to the case in which there are missing data in wage dynamics. For any missing data configuration, it consists in deleting the rows and columns of the covariance matrix corresponding to missing data and write the likelihood function accordingly. Under a normality assumption, the implicit moment selection underlying this estimation method is optimal, and though the method loses optimality in the non-normal cases, it is still useful for moment selection and for small-sample bias reduction (Okui, 2009).

4.2 Fixed effects

Random effect estimates can now be exploited to construct individual specific estimates of parameters η_i^c. First, if the ARMA model that we retained above is the correct specification of every stochastic individual profile, fixed effect estimates are linear combinations of residuals u_{it} and those FGLS estimates are optimally weighted to account for serial dependence. Supplementary Appendix S.III indeed proves that the individual effect estimates are given by:

$$\hat{\eta}_i^c = \hat{B}u_i, \tag{9}$$

in which matrix B is a function of random effect parameters, such as those governing serial correlation of transient errors and the covariance between the individual parameters and the initial conditions. Its estimate, $\hat{B}$, is obtained by plugging-in the expression of B, their respective random effect estimates.

Second, even if the ARMA model is incorrect, those estimates are still consistent when $T_i \to \infty$ because what matters is the mean independence of individual effects with respect to factors stated in equation (5), and not the specific form of serial dependence. Their standard errors should, however, be corrected. We use Newey-West robust standard errors in the empirical section. Nonetheless, FGLS relying on serial dependence as estimated by random effects, exploits the information that we have about "aggregate" serial dependence, as opposed to a simple OLS or non-linear least square estimate (Polachek et al, 2015). It enhances the quality of the estimates if the term $\Omega\left(\eta_i^c\right)$ in equation (8), is not too heterogeneous and this will be checked after estimation.

Consistency properties could, however, be misleading since T_i varies in our sample between 4 and 28. To assess the magnitude of the bias, we shall compare the estimates of the covariance matrix of η_i^c that we obtained by random effect and by fixed effect methods by grouping individual profiles according to the length of the observation periods.

Abstracting first from estimating errors, an unfeasible estimate of individual effects is defined as:

$$\tilde{\eta}_i^c = Bu_i = \eta_i^c + B\xi_i,$$

in which random vector ξ_i has mean zero conditionally on η_i^c and covariance matrix, Ω_ξ. This new notation, ξ_i, is introduced since it differs from v_i^c in equation (7) because of the correlation of initial conditions with η_i^c. These objects and this expression are defined and derived in the supplementary appendix S.III. We have:

$$\begin{aligned} V(\tilde{\eta}_i^c) &= EV(\tilde{\eta}_i^c \mid \eta_i^c) + VE(\tilde{\eta}_i^c \mid \eta_i^c) \\ &\Longrightarrow V(\tilde{\eta}_i^c) = B\Omega_\xi B' + V(\eta_i^c), \end{aligned} \tag{10}$$

a biased estimate of $V(\eta_i^c)$. The bias term is $B\Omega_\xi B'$ and it is easy to show that the dominating term is of order $1/T_i$.

Our feasible estimate has an additional bias given by the measurement equation,

$$\hat{\eta}_i^c = \hat{B}u_i = \tilde{\eta}_i^c + (\hat{B} - B)\xi_i,$$

although this term is in $1/\sqrt{N}$ and thus dominated, in large N and moderate T_i samples, by the bias in $1/T_i$.

We estimate the bias in equation (10) by replacing, in the expression, $B\Omega_\xi B'$, the unknowns by their corresponding random effect estimates and derive a bias-corrected estimate of the true variance of fixed effects, $V(\eta_i^c)$ (e.g. Arellano and Bonhomme, 2012, and Jochmans and Weidner, 2018).

5 Results

We first describe the estimated parameters of the aggregate equation (2) in Section 5.1 and turn to the estimation results of the within group wage equation by random effect methods in Section 5.2. In Section 5.3 we comment on the estimates of fixed individual effects and present descriptive statistics of these estimates. We wind up the section with robustness checks and other diagnostics.

5.1 Average effects estimation

For estimating series φ_{gt} in equation (3), we use a simplication of the flat spot approach proposed by Heckman, et al. (1998) and developed by Bowlus and Robinson (2012). Details of the estimation of a single series of human capital prices are presented at the end of the Data Appendix A. In a nutshell, human capital prices are estimated using a population of older males whose potential experience ranges from 25 to 40 (e.g. whose age stands between 43 and 58, see Appendix A) as in Bowlus and Robinson (2012). Those prices are used to deflate real wages.

Table 2 presents for each of the 20 groups the estimated aggregate group parameters from equation (2). They exhibit expected patterns. The first factor loading average $\bar{\eta}_{g1}$ ranges from

2.4 for the lowest skill groups to 3.4 for the highest skill groups – a 100% difference. The estimated average slope , $\bar{\eta}_{g2}$ ranges from 0.017 to 0.07 in the range of Mincer's estimates (e.g. Lagakos et al., 2018). As the previous average, it is larger for the high-skilled groups than for the low-skilled ones although the evidence is weaker. The geometric factor loading average $\bar{\eta}_{g3}$ is negative as expected or non significantly different from zero.

Interestingly the pattern of correlations of the average effects across education groups (weighted by group size) is very close to the correlation pattern of centered factor loadings, estimated by random effects, presented next in Table 3.

5.2 Random effect estimates

Random effect estimates when disturbances, v_{it}^c, are ARMA(p, q) in which p and q vary between 1 and 3 are reported in the Supplementary Appendix in Tables S.iv and S.v. Those results and the Akaike criterion reported in Table S.iii made us choose an ARMA$(3, 1)$ as our preferred specification.[6] We comment below ARMA$(3, 1)$ results as well as other parameters that can be constructed from those.

The estimated covariance matrix of the centered individual effects, η_i^c is stable across the different specifications of ARMA processes as shown in Table 3. Their standard deviations are very precisely estimated at around .30 for the "level" factor loading, η_{i1}^c, and .25 for the curvature one, η_{i3}^c, and at around .04 for the slope, η_{i2}^c. The correlation between the slope and curvature factor loadings is very strongly negative and equal to $-.95$ consistently across ARMA specifications. The magnitude of this correlation and its sign are consistent with the structural model that ties in parameters η_{i2} and η_{i3}: the larger the slope, the more curved the wage profile. The retraction force due to the horizon is stronger for high wage growth workers.

The correlation coefficient between the curvature, η_{i3}^c, and the level, η_{i1}^c, factor loadings is also significantly negative – around -0.6 – and confirms that the retraction force at the end of the life cycle is also stronger for highly skilled workers. The correlation between the level and slope factor loadings, η_{i1}^c and η_{i2}^c, is positive and around .5. This is related to the result in the recent literature (e.g. Lagakos et al, 2018, Engbom, 2018) that wage growth is significantly higher for more educated individuals in a very heterogeneous setting (also Guvenen, 2007).

Returning to the correlation across aggregate effects, the coefficient of correlation between $\bar{\eta}_{g1}$ and $\bar{\eta}_{g2}$ (i.e. by varying g) is equal to 0.66 and close to the random effect estimate of the

[6] Increasing the order of the MA decreases the AIC criterion very moderately and some of the parameters are very imprecisely estimated (see Table S.iv).

correlation between η^c_{i1} and η^c_{i2}, which is equal to 0.5. The estimated correlation between $\bar{\eta}_{g1}$ and $\bar{\eta}_{g3}$ is negative, -0.64 and very close to the random effect estimate, -0.636. The estimated correlation between $\bar{\eta}_{g2}$ and $\bar{\eta}_{g3}$ of -0.96, is also very close to the random effect estimate, -0.95. This result evinces that human capital investment patterns between and within groups are similar in France in contrast to what was found in the US (Heckman et al., 1998). We attribute this difference to the stability of relative human capital prices across groups over this period in France. Our specification nonetheless slightly deviates from previous ones, by the introduction of a curvature factor and the control of initial conditions.

Irrespective of the order of the ARMA process, the initial conditions are negatively correlated with the level factor loadings, η^c_{i1}, positively with the slope effect, η^c_{i2}, and negatively with the curvature one, η^c_{i3}.[7] These initial conditions account for the strong transitory conditions that seem to affect the wage process at the beginning of the working life (as well as the impact of our data selection process).[8] Even if the (log) wage process is asymptotically stationary, initial conditions are definitely not set on the stationary path that corresponds to this process. Moreover, in all ARMA specifications, the standard deviation of individual and time specific transitory shocks is decreasing over time. Individual specific frictions decrease over time and this result is found across different countries (e.g. Bagger et al., 2014, Bowlus et Liu, 2012).

Goodness-of-fit is examined in different graphs. In Figure 2, we report how the estimated variances as well as the observed variances evolve over time. They fit very nicely in the first part of the sample (until 1994) but this breaks down after 1994 after which the shape of the evolution of variances is similar, albeit at a level which is higher than the observed level. It confirms that 1994 is an abnormal year even if the goodness-of-fit for autocorrelations is good as reproduced in Figures 3 and 4.

We tried different mechanisms in order to understand better the discrepancy between observed and predicted variance profiles. One possibility is to allow for an additional measurement error term in 1994 for instance, like in Guvenen (2009) or to drop this year altogether. These attempts did not affect goodness-of-fit. A more disturbing explanation for those discrepancies is that it reflects a failure in the missing at random hypothesis. When one represents the evolution of the variance of wages over the life-cycle using fixed effect estimates (see below), it clearly

[7] Estimates of the covariances between the factor loadings and the initial conditions are reported in Table S.iv in the Supplementary Appendix.

[8] The strong decrease of the variance observed during the first years might partly be due to the very stringent selection made in the 1977 entry cohort. The very flexible initial conditions, as they are accounted for in the random effect estimation, also control for this selection.

appears that the level of these profiles negatively depends on the number of periods in which we observe each individual. Variances are larger for individuals who have shorter spells in the panel. Nevertheless, correcting for non random attrition seems out of the scope of this paper and we leave it for further research.[9]

5.3 Fixed effect estimation

We now turn to the estimation results by FGLS of the three individual factor loadings from individual wage profiles. Technical details are given in Section 4.2 and completed in the Supplementary Appendix S.III. Estimated group averages are added to within group estimates to reconstruct the final estimates of individual effects, $\hat{\eta}_i$.

Table 4 reports estimates of the bias-corrected covariance matrix of centered individual effects, obtained by fixed effect methods, and compare them to the random effect benchmark. Standard errors for any function of fixed effects are computed using sampling variability to which is added the effects of parameter uncertainty due to random effect estimation. We use Monte Carlo simulations to compute the latter by sampling 1,000 times in the asymptotic distribution of random effects estimates.

Our working sample to be used from now on, gathers individuals observed over more than 21 periods only, because small-T_i bias issues seem to be lingering for observations $T_i \leq 20$. Tables S.vi and S.vii in the Supplementary Appendix report raw and bias-corrected estimates in the full sample and justify such a selection, to which we return in the robustness section below. As random and fixed effects do not strictly refer to the same population because of this selection, discrepancies seem very moderate between random and fixed effect estimates – as Figure 5 reporting variance estimates confirm.

Table 4 also reports these estimates by subsamples indexed by a varying number of periods of observation from 21 to 28. It clearly appears that the longer the observation period, the less variable individual effects. It might be due to subsisting small-T bias that we imperfectly control. It might also indicate that subpopulations, with exits or/and reentries, are more heterogenous that the subpopulation of those who remains in the private sector all along. Random effect estimates would reflect the mixture of these subpopulations.

[9] Furthermore, the conditions for consistency of fixed effect estimates described below are less stringent since the missing at random assumption can be weakened and taken as conditional on individual effects.

5.4 Predictions

We can assess whether predictions of human capital models conform with these estimates. We look at the Mincer dip, the prediction about the covariance between wage and wage growth and the correlation between wage growth and initial wage level.

Mincer dip First, all estimates generate a neat Mincer dip at the beginning of the life-cycle after 4 to 8 years – a little longer than expected (Mincer, 1974). This is shown by plotting, in Figure 5, the profile of predicted variances of wages along the life-cycle computed using random and, raw and bias-corrected, fixed effect estimates. We display, in this graph, the life-cycle profile of variances of the permanent effects, given by the combination of factors and factor loadings ($V(M(\beta)\eta_i^c)$) in which matrix $M(\beta)$ is composed of a constant, trend and geometric terms, see equation (6). Stochastic transitory earnings, v_i^c, obscure the comparison between estimates and are excluded from the graph. Wage profiles using raw and bias-corrected fixed effect estimates reproduce the random effect variance profile at a slightly higher level during the first years of working life in Figure 5. Discrepancies with random effect estimates seem however second order in particular at later periods in the life-cycle and this validates the use of this selected sample in the decomposition of inequalities and counterfactual variations in Section 6.

Correlation between wage and subsequent wage growth Second, as developed in Rubinstein and Weiss (2006), a human capital model predicts that the covariance between wage levels and subsequent wage growth should be negative when the person enters the labour market, and should turn positive after some time, in contrast to a job search model in which the correlation remains negative. This is what Table 5 confirms. In the working sample, covariances are negative in years 1977 and 1982 and turn positive and significant from 1987 onwards, although the increase over time is not monotonic. This is slightly more pronounced for the high-skilled group. While Rubinstein and Weiss (2006) use data on residual wages, smoothed over three years and show that this covariance increases over time, it remains negative (their Figures 8a-8e in Section 4.2). Our focus on the permanent effects that filter out transitory components allows to go a step further and produces a prediction in contradiction with a pure job-search rationale at least after ten years in the labour market.

Initial wage levels and returns to experience Initial skills and skills acquired during the education stage, are shown to be complementary (see Heckman, Humphries and Veramendi, 2017). It is thus interesting to measure the complementarity over the life cycle by computing the correlation between the initial log wage and the returns at each point of the life-cycle as in Gladden and Taber (2009). By filtering out the transitory components at the beginning of the life cycle, we are able to look more precisely at the complementarity between estimated initial wage levels – by equation (1), computed as the estimate of $\eta_{i1} + \eta_{i3}$ – and returns.

Returns to experience are computed as the marginal effects of potential experience, t, on the permanent component – i.e. $\eta_{i2} - \log(\beta)\beta^{-t}\eta_{i3}$. These returns are decreasing because $\log(\beta) < 0$ and estimated η_{i3} are mostly negative. As shown in Table 6, while the correlation is significant and negative at the beginning of the life-cycle (the first twenty years) as in Gladden and Taber (2009) and Sorensen and Vejlin (2014), it turns positive and significant afterwards. This result is true in the working sample and in various skill subsamples. Yet, the correlation between the long-run log wage and the corresponding returns remain negative and significant.

The negative correlation at the beginning of the life-cycle is due to the negative correlation between the initial log wage, $\eta_{i1}+\eta_{i3}$, and the growth effect, η_{i2}. The positive correlation between the second curvature term– $-\log(\beta)\beta^{-t}\eta_{i3}$ – and the initial log wage kicks in after 20 years to reverse this foregoing negative correlation. Individuals with high initial log wages tend to have higher returns later in the life-cycle. It emphasizes the importance of considering horizon effects (Lillard and Reville, 1998) in the analysis of wage profiles.

5.5 Robustness and other diagnostics

We tested various departures from our baseline estimates to check that our results are robust. We also comment on additional goodness-of-fit diagnostics.

The first issue is the estimation of the series by a flat spot approach that underpins the identifying restriction (3). The dynamics of human capital accumulation depends on whether the average wage or productivity profile is attributed to human capital only or to other factors (physical capital for instance). To control for this issue, we also repeated our procedures by deflating real wage by a series of average labor productivity. Results change only marginally with respect to the results that were presented above.

As serial correlation affects inference, we also vary the number of lags in the Newey West procedures without much impact overall.

Another issue is related to the quality of the correction of the bias in the fixed effect estimates that we reported. In Tables S.vi and S.vii in the Supplementary Appendix which report raw and bias-corrected estimates, the magnitude of all covariance estimates decreases with the number of period of observations, as expected by the computations of Section 4.2. Bias-correction flattens these estimates by factors of 2 to 3 when the number of periods is small but it decreases to 10-20% when the number of periods is 20. There is a clear break in these Tables between the estimates below and above 20 and this is why we chose to work with the 21+ sample. As expected the bias correction becomes negligible when $T_i = 28$. The counterfactual analyses that we present below are robust to a change in this threshold.

The comparison between random effect and fixed effect estimates implicitly relies on an homogeneity assumption of the residuals, $\hat{v}_i^c$ as a function of η_i. When plotting the variance profiles of these residuals in groups defined either by skills or by age of entry, we find very little differences between those groups (see Figure S.iii in the Supplementary appendix). The three-factor structure seems to be sufficient to describe the individual permanent heterogeneity in our data and this partly justifies ex-post the homogeneity assumption of the covariance matrix of transitory terms in the random effect specification as well as the homogeneity assumption of the discount factor which is used to measure the curvature or horizon effect.

6 Wage inequalities: decompositions and counterfactual impacts

We now exploit those estimates to study the impact of heterogeneity on life-cycle wages and decompose wage inequalities at a point in time and over the life-cycle. When we use fixed effect bias-corrected estimates, we only use information on the most stable trajectories ($T_i \geq 21$) for which we compute various counterfactuals of interest.

6.1 Counterfactual impacts on means

The log wage equation (6) writes:

$$\log y_i = M(\beta)(\bar{\eta}_{g(i)} + \eta_i^c) + v_i^c,$$

and we assess the effect on average log wages of increasing each component of $\eta_i = \bar{\eta}_{g(i)} + \eta_i^c$ by one tenth of its standard deviation. Those experiments can be easily expressed as a transformation of η_i^c into $\phi(\eta_i^c)$ and we have that:

$$\begin{aligned}\Delta \log y_i &= \log y_i(\phi(\eta_i^c)) - \log y_i, \\ &= M(\beta)(\phi(\eta_i^c) - \eta_i^c).\end{aligned}$$

We report mean impacts, every five years, starting in 1977 – their year of entry into the labor market – and finishing the last year of observation in 2007. We refrain from extrapolating to further periods in the future. We also consider a longer-run measure of wages over the observed life-cycle, an appropriately discounted aggregate of log wages – that happens to correspond to intertemporal utility over the observed profile of wages, normalized in such a way that it can be compared with the annual values (see Magnac et al., 2018):

$$\log y_i^{(LR)} = \frac{1-\beta}{\beta(1-\beta^T)} \sum_{t=1}^{T} \beta^t \log y_{it}. \tag{11}$$

Those impacts are reported in Table 7. Unsurprisingly because only levels of log wages are affected, the impact of increasing the level, η_{i1}, by one tenth of its standard deviation has a constant short-run and long-run impact around 0.03 – i.e. 3% on the wage level – except in the first year because initial conditions blur the impact. This is somewhat below the average slope of profiles (Table 2) and thus, somewhat below the effect of an additional year of experience at the beginning of the life cycle. In contrast, the impacts of increases of one tenth of a standard deviation in η_{i2} (returns) and η_{i3} (horizon) increase from 1977 to 2007 as they accumulate over the life-cycle. At the end of the period of observation, they have sizeable magnitudes slightly above 0.15 on the log wage in the terminal year 2017. Effects on the long-run value are less important and around .06 (returns) or .07 (horizon). Overall, these results mean that the heterogeneity in level, slope and curvature that we consider is economically significant.

We now address two questions related to the decomposition of the variance of log wages into its different components.

6.2 Short-run and long-run inequality decompositions

Table 8 provides decompositions of within-cohort inequalities and long-run inequality into observed heterogeneity, unobserved heterogeneity and transitory components. Using random effect estimates, decompositions are derived from writing the log wage equation as:

$$\log y_i = M(\beta)\bar{\eta}_{g(i)} + D\eta_i^c + \xi_i,$$

in which the three components $\eta_{g(i)}$, η_i^c and ξ_i are orthogonal between each other and are interpreted as observed heterogeneity, unobserved individual heterogeneity and individual-and-time specific frictions. This expression deals with the absence of orthogonality between permanent and transitory components because of initial conditions. Initial conditions are negatively correlated with η_{i1} and η_{i3}, and these effects play an important role during the first years of the working life (Table S.iv of the Supplementary appendix). To make them orthogonal, initial conditions are first projected onto individual effects η_i^c and this projection is aggregated with the impact of individual effects η_i^c in the term, $D\eta_i^c$, and the residuals with the transitory ones in the term ξ_i. Supplementary appendix S.III gives more details.

We report results obtained by random effects in the first panel of Table 8 and by fixed effects in the second panel. Results obtained using random or fixed effect estimates are very close to each other and we will comment the latter only. The last column of these Tables reports the predicted variance of log-wages $V(\log y_{it})$ every five years from 1977 to 2007 as well as the long-run value defined in equation (11). These results conform with the inverted U-shape of variances as in Figure 1. The variance of long-run log wages is lower since transitory components over the life cycle are partly averaged out. The ratio of short-run inequalities to the long run one is varying between 1.22 in 1982 and 1.82 in 2007, well in line with Bonke et al. (2014).

The first three columns respectively report the share of the variance due to observed heterogeneity, $V(M(\beta)\bar{\eta}_{g(i)})$, the share of the variance attributable to unobserved individual heterogeneity, $V(D\eta_i^c)$, and the share of the variance generated by transitory components, $V(\xi_i)$. Note that the decomposition in 1977 seems to be quite different from the one in other years because of the eventful process at the beginning of a working life. The weight in percentage terms of the transitory component is similar to other years but the variance of unobserved heterogeneity is almost absent. This is partly due to the orthogonalisation procedure that we have just discussed above, whereby negative correlations between initial conditions and fixed effects lower the contribution of permanent unobserved heterogenity to the variance of log wages.

Nonetheless, on average, 68% of the variance is due to the combination of the observed and unobserved heterogeneity factors in 1977 and 1982. This share displays an increase over the life cycle from 68% to 91% thirty years later. As a mirror image, the share of the variance explained by transitory components decreases sharply from 32% in 1982 to 8.5% in 2007 as well as the share due to the observed heterogeneity component, albeit at a lesser degree from 35% to 25%. This is the consequence of an increase in the importance of unobserved individual heterogeneity, which doubles its contribution to the variance of log wages from 33% to 66%. As expected, these effects are weighted differently when the long-run value of log wages (i.e. equation (11)) is computed. The transitory effect is lower (less than 6%) and observed and unobserved heterogeneity components have roughly equal contributions (45 and 49%).

The sum of the observed and unobserved heterogeneity contributions is larger than the ones found by Huggett, Ventura and Yaron (2011), who find that the initial endowments related to human capital (initial human capital and learning ability as well as initial wealth) account for 60% of the variance of lifetime wage. Their framework however allows for less pervasive unobserved heterogeneity, than we do here, so that we capture more unobserved heterogeneity than they do.

We now turn to decompositions involving each specific factor loading.

6.3 Heterogeneity components

We can further decompose observed and unobserved heterogeneity into its constituent parts: level, η_{i1}, slope, η_{i2}, and curvature, η_{i3}. Denote the permanent heterogeneity variable, as defined in equation (6):

$$p_i = M(\beta)^{[1-p,T]}\eta_i.$$

If observed heterogeneity only is present (i.e. $\eta_i = \bar{\eta}_{g(i)}$), the variance of p_i is the variance of an homogenous Mincer equation. If there is as much heterogeneity as is estimated, this delivers the estimated $V(p_i)$ as might be computed from previous Table 8. Between these two benchmarks, we can compute hypothetical variances of the permanent effect by shutting down some of its components. For instance, in experiment 2 below, we consider that η_{i1} is equal to the estimated value while η_{i2} and η_{i3} are set to the sum of the observable components, $\bar{\eta}_{g(i)2}$ and $\bar{\eta}_{g(i)3}$, and the predictable components of η_{i2} and η_{i3} given η_{i1}. In other words we restrict heterogeneity to a single component, the level in order to assess its impact. We do the same for other components, slope and curvature.

We report variances of log wages, every 5 years, as well as the long-run value of log-wages, in the following five experiments:

1. **Observable benchmark**: heterogeneity in level, slope and curvature restricted to observables – $\eta_{i1} = \bar{\eta}_{g(i)1}$, $\eta_{i2} = \bar{\eta}_{g(i)2}$ and $\eta_{i3} = \bar{\eta}_{g(i)3}$.

2. **Heterogeneity in level:** heterogeneity in slope and curvature restricted to observables. Analytically, $\eta_{i1} = \bar{\eta}_{g(i)1} + \hat{\eta}_{i1}$,the estimated value and $\eta_{i2} = \bar{\eta}_{g(i)2} + \omega_{21}\hat{\eta}^c_{i1}$ and $\eta_{i3} = \bar{\eta}_{g(i)3} + \omega_{31}\hat{\eta}^c_{i1}$ as predicted using the correlations between the components of η^c_i.

3. **Heterogeneity in slope:** heterogeneity in level and curvature restricted to observables – η_{i2} = estimated value, η_{i1} and η_{i3} equal to the sum of group values and predictions given $\hat{\eta}_{i2}$ using the correlations between the components of η^c_i.

4. **Heterogeneity in curvature**: heterogeneity in level and slope restricted to observables – η_{i3} = estimated value, η_{i1} and η_{i2} equal to the sum of group values and predictions given $\hat{\eta}_{i3}$ using the correlations between the components of η_i^c.

5. **Heterogeneity in slope and curvature**: heterogeneity in level restricted to observables – η_{i2} and η_{i3} = estimated values, η_{i1} equal to the sum of group values and predictions given $\hat{\eta}_{i2}$ and $\hat{\eta}_{i3}$ using the correlations between the components of η_i^c.

Table 9 reports variances of the counterfactual permanent variables, p_i, defined by such a sequence of experiments as they can be constructed from fixed effect estimates in the working sample. Rows report them every five years from 1977 to 2007 as well as the long-run variance. The last column reports the estimated variance of p_i. The first to the fifth column report variances for every experiment 1 to 5 described above.

First, as shown in the previous section, observables explain most of the permanent heterogeneity in 1977, which is hardly surprising, since observables are age at entry and observed skills in the first job. This explanatory power declines afterwards down to less that 30% in 2007 (column 1, Table 9). Second, the heterogeneity in level explains most of the remaining variance of the permanent component in 1982 but this contribution remains almost constant over the years while the variance of the permanent component increases (column 2). This shows the limit of panel data analyses in which a single unobserved heterogeneity component in levels is considered. Third, the heterogeneity in slope or in curvature contributes less to the variance of the permanent component at the beginning of the life cycle but their contributions increase until a single heterogeneity dimension in the growth or the curvature almost perfectly predict the variance of the permanent component in 2007.

Nonetheless, explaining heterogeneity in the long run variance requires the presence of the three factor loadings. None of them alone, or combined for η_{i2} and η_{i3}, succeeds in explaining fully the permanent wage component variance. This is because the long-run variance combines elements at the beginning of the life cycle which are well explained by the level while the second and third components are associated with what happens later in the life-cycle.

The same Table for the low and medium skill group is displayed in the Supplemental Appendix S.viii and they exhibit the same patterns. Table 10 reports the same exercise for the high skill group.[10] In this group, the variance of the permanent components seem less predictable by only one or two heterogeneous components and the three factor loadings seem necessary.

[10] For this group, the importance of observables is lower since the skill in the first job is less well described.

Overall, these results confirm the importance of taking into account an heterogeneous slope coefficient for experience as well as allowing for an horizon effect.

7 Conclusion

In this paper, we analyze wage profiles by using an observable microfounded factor model, based on human capital investments, and ARMA individual-and-time errors. Three factor loadings – level, slope and curvature – describe wage profiles. We propose an estimation method of the factor loadings based on a sequence of intertwined random and fixed effect methods. We show the importance of considering pervasive heterogeneity to model wage profiles and offer a set of decompositions of wage inequalities in terms of observed and unobserved heterogeneity as well as in terms of level, slope and curvature of the individual wage profiles. In particular, heterogeneous horizon effects, which are not considered in the literature, are shown to be quite important, at least after 20 years of experience.

Much remains of be done at the methodological level. The issue of missing data seems at the top of the agenda since the selection of balanced panels, or the missing at random assumption, might bias our view of inequalities since we select more stable subpopulations. Other ways of modelling transitory components might be in order and in particular, more attention might be paid to what can be anticipated by agents as in Cunha, Heckman and Navarro (2005). We leave all these issues for future research.

REFERENCES

Abowd, J. M., Kramarz, F., & Margolis, D. N., 1999, "High wage workers and high wage firms", *Econometrica*, 67(2), 251-333.

Alvarez, J. and Arellano, M., 2004, "Robust likelihood estimation of dynamic panel data models", Working Paper 0421.

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

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

Atkinson, A.B. and S. Morelli, 2014, "Chartbook of Economic Inequality: 25 Countries 1911-2010". *ECINEQ* WP 2014-324

Autor, D. H., Katz, L. F., & Kearney, M. S., 2008, "Trends in US wage inequality: Revising the revisionists." *The Review of economics and statistics,* 90(2), 300-323.

Bagger, J., Fontaine, F., Postel-Vinay, F., & Robin, J. M., 2014, "Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics", *American Economic Review*, 104(6), 1551-96.

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

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.

Belley, P., 2017, "Understanding Wage Growth: Estimating and Testing Learning-by-Doing", *Journal of Human Capital*, 11(2), 213-246.

Blandin, A., 2018, "Learning by doing and Ben-Porath: Life-cycle predictions and policy implications", *Journal of Economic Dynamics and Control*, 90, 220-235.

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

Blundell, R., & Etheridge, B., 2010, "Consumption, income and earnings inequality in Britain", *Review of Economic Dynamics*, 13(1):76-102.

Bonhomme, S., and J.M., Robin, 2009, "Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990-2000", *Review of Economic Studies,* 76(1), 63-92.

Bönke, T., Corneo, G., & Lüthen, H., 2014, "Lifetime earnings inequality in Germany", *Journal of Labor Economics*, 33(1), 171-208.

Bowlus, A. J., & Liu, H., 2013, "The contributions of search and human capital to earnings growth over the life cycle", *European Economic Review*, 64, 305-331.

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

Browning, M., Ejrnæs, M. and Alvarez, J., 2012, "Modelling income processes with lots of heterogeneity", *Review of Economic Studies,* 77(4):1353-1381

Burdett, K., Carrillo-Tudela, C., & Coles, M., 2016, "Wage inequality: A structural decomposition", *Review of Economic Dynamics*, 19, 20-37.

Charnoz, P. Coudin, E. and Gaini, M., 2011, "Wage inequalities in France 1976-2004: a quantile regression analysis", WP DESE n°G2011/06.

Charnoz, P. Coudin, E. and Gaini, M., 2014, "Forty Years of Decreasing Wage Inequality in France: The Roles of Supply and of Hidden Skill-Biased Technical Change", WP CREST n° 2014-20.

Cunha, F., J.J., Heckman and S. Navarro, 2005, "Separating uncertainty from heterogeneity in life cycle earnings", *Oxford Economic Papers*, 57(2):191-261.

Daly, M., Hryshko, D., & Manovskii, I., 2016, "Reconciling estimates of income processes in growth rates and levels". Working paper, University of Pennsylvania.

Ejrnæs, M., and M., Browning, 2014, "The persistent–transitory representation for earnings processes", *Quantitative Economics*, 5(3), 555-581.

Engbom, N., 2017, "Worker Flows and Wage Growth over the Life-Cycle: A Cross-Country Analysis", Available at SSRN 2976359.

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

Gladden, T., & Taber, C., 2009, "The relationship between wage growth and wage levels", *Journal of Applied Econometrics*, 24(6), 914-932.

Gottschalk, P., & Moffitt, R., 2009, "The rising instability of US earnings", *Journal of Economic Perspectives*, 23(4), 3-24."

Guvenen, F., 2007, "Learning your earnings: Are labor income shocks really very persistent?", *American Economic Review* 97(3): 687-712.

Guvenen, F., 2009, "An empirical investigation of labor income processes", *Review of Economic Dynamics,* 12:58-79.

Heckman, J. J., Humphries, J. E., & Veramendi, G., 2018, "Returns to education: The causal effects of education on earnings, health, and smoking", *Journal of Political Economy*, 126(S1), S197-S246.

Heckman, J., Lochner, L., & Cossa, R., 2002, "Learning-by-doing vs. on-the-job training: Using variation induced by the EITC to distinguish between models of skill formation", NBER 9083, National Bureau of Economic Research.

Heckman, J.J., L. Lochner and 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.

Hoffmann, F., 2019, "HIP, RIP and the robustness of empirical earnings processes", *Quantitative Economics*, 10, 1279–1315.

Hospido, L., 2012, "Modelling heterogeneity and dynamics in the volatility of individual wages", *Journal of Applied Econometrics*, 27(3), 386-414

Hryshko, D., 2012, "Labor income profiles are not heterogeneous: Evidence from income growth rates", *Quantitative Economics*, 3(2)177-210.

Huggett, M., G., Ventura and A., Yaron, 2011, "Sources of Lifetime Inequality", *American Economic Review*, 101:2923-2954.

Jochmans, K., & Weidner, M., 2018, "Inference on a distribution from noisy draws", arXiv preprint arXiv:1803.04991.

Keane, M. P., & Wolpin, K. I., 1997, "The career decisions of young men", **Journal of Political Economy**, 105(3), 473-522.

Lagakos, D., Moll, B., Porzio, T., Qian, N., & Schoellman, T., 2018, "Life cycle wage growth across countries", *Journal of Political Economy*, 126(2), 797-849.

Le Minez, S. and Roux, S., 2002, "Les différences de carrières salariales à partir du premier emploi", *Economie et Statistique* 351:31-63.

Lillard, L. A. and Reville, R. T., 1999, "Life cycle human capital investment: New evidence on and old issue", unpublished manuscript.

Lochner, L., and Y. Shin, 2014, "Understanding earnings dynamics: Identifying and estimating the changing roles of unobserved ability, permanent and transitory shocks", National Bureau of Economic Research, WP-w20068.

Magnac, T., N., Pistolesi and S., Roux, 2018, "Human Capital Investments and the Life Cycle of Earnings", *Journal of Political Economy,* 126(3), 1219-1249.

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

Milanovic, B., 2016, *Global Inequality,* The Belknap Press of Harvard University Press, Cambridge.

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

Moffitt, R. A., & Gottschalk, P., 2011, "Trends in the covariance structure of earnings in the US: 1969–1987", *The Journal of Economic Inequality*, 9(3), 439-459.

Okui, R., 2009, “The Optimal Choice of Moments in Dynamic Panel Data Models”, *Journal of Econometrics*, 151:1-16.

Piketty, T., 2013, *Le capital au XXIe siècle,* Le Seuil: Paris.

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

Pora, P., & Wilner, L. (2017). Heterogeneous Exposure to Labor Earnings Risk.

Postel-Vinay, F. and Turon, H., 2010, “On the job search, productivity shocks and the individual earnings process”, *International Economic Review* 51(3):599-629.

Rubinstein, Y. and Weiss, Y., 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.

Sanders, C., & Taber, C., 2012, "Life-cycle wage growth and heterogeneous human capital", *Annual Review of Economics*, 4(1), 399-425.

Song, J., Price, D. J., Guvenen, F., Bloom, N., & Von Wachter, T., 2018, "Firming up inequality", *The Quarterly Journal of Economics*, 134(1), 1-50.

Sørensen, K. L., & Vejlin, R., 2014, "Return to experience and initial wage level: Do low wage workers catch up?", *Journal of Applied Econometrics*, 29(6), 984-1006.

Verdugo, G., 2014, “The great compression of the wage structure in France 1969-2008”, *Labour Economics* 28:131–144.

APPENDICES

A Data Appendix

As in Le Minez and Roux (2002), we consider individuals right from their entry into the labor market and onwards. Labor market entry is defined as being employed for more than 6 months and being still employed the following year, possibly in different firms. For the entry cohort of interest which starts in 1977, this leads us to select from the administrative data $36,883$ individuals who were employed more than 6 months in 1977 and at least one day in 1978. Among them, 53% have worked but not permanently before. Conversely, individuals who have worked in 1977 are not considered as entrants if their jobs are not permanent enough. They may however enter with a subsequent cohort.

In addition, we aim at keeping employees with a permanent full-time attachment to the private sector only. Firstly, we consider workers employed full time only and we censor information about part-time jobs. In addition to the condition which requires workers to work in the private sector during the year of entry and the following one, we further restrict the sample to men also working in 1982 and 1984. This is because we want to avoid dealing with non participation issues for females and with too many exits from the sample since the bulk of entries into public service occurs at the beginning of the working life. These restrictions lead us to retain in the 1977 entry cohort $16,091$ men who entered the labor market in 1977 in a full-time position for more than 6 months and who were also full-time employed in 1978. Adding the condition on the presence in a full-time position in 1982 and 1984 further restricts the sample to $8,288$ individuals. Finally, we keep only workers who were aged between 16 and 30 at their entry in the labor market and this restricts the sample to $7,446$ workers.

We impose these restrictions in order to concentrate on a relatively homogeneous sample of workers with a long term attachment to the private firms' labor market. Admittedly, it does not represent the full working population. Because of the lack of a credible identification strategy to correct for selection, we shall assume that selection is at random or can be conditioned on individual-specific effects only. The distribution functions of unobserved factor loadings or idiosyncratic components that we estimate in the following refer to this subpopulation.

The empirical analysis uses "annualized" earnings which are thus better called wages. It is defined as the sum of all earnings during the year divided by the number of days worked and remultiplied by 360 (total number of days during the year in the administrative data). Accounting only for total yearly earnings would miss other earnings from employment in the public sector, self-employment income or unemployment benefits that are not observed in the data. Considering annualized earnings instead limits this problem, although it may lead to overestimating yearly income. In order to weaken the possible impact of measurement error, we coded as missing the first and last percentiles of the earnings distribution in every period.

Inflation, as measured by consumer prices, leads to subtracting a factor equal to 1.17 to current log-wages over the whole period. This can be roughly subdivided into two sub-periods between 1977 and 1986 in which this factor is equal to .77 and between 1986 and 2007 during which inflation levelled off and this factor is equal to .40.

Age at labor market entry (in 1977) can only take odd integer values from 17 to 29, i.e seven different values because of the specific sampling of the dataset. As groups formed by age at entry and skills are defined according to characteristics recorded at the entry on the labor market, individuals are attached to the same group during their whole working life.

Estimation of human capital prices by a flat spot condition and robustness checks We follow Bowlus and Robinson (2012). From the DADS, average log daily real wages by age and year can be computed on full-time males employees in the Private Sector from 1976 to 2010. To identify the "flat spot" region where human capital remains stable, we run regressions of the average log daily real wage on potential experience (difference between current age and 16), an exponential term reflecting the curvature of the wage profile with respect to potential experience, and year dummies. We have run different regression changing the contributing individuals with respect to their potential experience and selected the population with the broader range of potential experience for which the coefficients an potential experience and the exponential term were statistically non significant. This leads us to select individuals who are aged between 43 and 58 whose average log-wage profile did not exhibit any slope or curvature. The results of the regressions and the human capital prices values are available upon request.

We then repeat the procedures that lead to the estimates of Table 2.

B Random effect specification

Redefining the time index accordingly, we shall assume that initial conditions of the process $(u_{i(1-p)}, ., u_{i0})$ are observed. The dynamic process is thus a function of the random variables $z_i = (v_{i(1-p)}, ., v_{i0}, \zeta_{i(1-q)}, ., \zeta_{iT})$ which collect initial conditions of the autoregressive process $(v_{i(1-p)}, ., v_{i0})$, initial conditions of the moving average process $(\zeta_{i(1-q)}, ., \zeta_{i0})$ and the idiosyncratic shocks affecting random shocks between 1 and T. We write the quasi-likelihood of the sample using a multivariate normal distribution

$$z_i \rightsquigarrow N(0, \Omega_z)$$

We define v_{it} as

$$v_{it} = \alpha_1 v_{i(t-1)} + ... + \alpha_p v_{i(t-p)} + \sigma_t w_{it},$$

where w_{it} is $MA(q)$:

$$w_{it} = \zeta_{it} - \psi_1 \zeta_{it-1} - ... - \psi_q \zeta_{it-q}.$$

The construction of the structure of Ω_z is detailed in the Supplementary Appendix S.II (Magnac et al, 2014) although it can be summarized easily. The correlations between initial conditions and individual effects are not constrained, while innovations ζ_{it} are assumed orthogonal to any previous terms including initial conditions. However, the initial conditions $(v_{i(1-p)}, ., v_{i0})$ can be correlated with previous shocks as $\zeta_{i0}, ., \zeta_{i(1-q)}$.

As for the individual effects $(\eta^c_{i1}, \eta^c_{i2}, \eta^c_{i3})$ we assume that they are independent of the idiosyncratic shocks $\zeta_{i(1-q)}, ., \zeta_{iT}$ while they can be correlated with the initial conditions of the autoregressive process $(v_{i(1-p)}, ., v_{i0})$ in an unrestricted way. From these restrictions it is possible to build the covariance matrix of the observed variables

$$Vu_i = (u_{i(1-p)}, ., u_{i0}, u_{i1}, ., u_{iT}) \equiv \Omega_u.$$

This covariance matrix, Ω_u, is a function of the parameters of the model that are the autoregressive parameters $\{\alpha_k\}_{k=1,\ldots,p}$, the moving average parameters $\{\psi_k\}_{k=1,\ldots,q}$, the covariance matrix (conditional on groups) of η^c, Σ_η, the heteroskedastic components $\{\sigma_t\}_{t=1,\ldots,T}$ and the covariance of fixed effects and initial conditions, $\Gamma_{0\eta}$ (see Supplementary Appendix S.II).

A pseudo likelihood interpretation can always be given to this specification. As in Alvarez and Arellano (2004), the estimates remain consistent under the much weaker assumption that:

$$E(\zeta_{it} \mid \eta_i, u_i^{t-1}) = 0,$$

although optimality properties of such an estimation method are derived under the normality assumptions only.

Tables and Figures

Table 1: Sample size

	Age of Entry in 1977			
	Below 20	Between 20 and 23	Above 23	All
1977	4460	2112	874	7446
1978	4460	2112	874	7446
1979	3855	1923	787	6565
1980	3748	1930	785	6463
1982	4460	2112	874	7446
1984	4460	2112	874	7446
1985	3792	1808	724	6324
1986	3683	1800	726	6209
1987	3569	1741	678	5988
1988	3402	1654	637	5693
1989	3486	1657	644	5787
1991	3319	1598	613	5530
1992	3299	1581	603	5483
1993	3330	1620	627	5577
1994	2508	1316	503	4327
1995	3256	1566	578	5400
1996	3236	1557	579	5372
1997	3202	1529	556	5287
1998	3208	1521	543	5272
1999	3218	1503	547	5268
2000	3180	1506	536	5222
2001	3117	1480	517	5114
2002	3018	1463	511	4992
2003	2800	1323	467	4590
2004	2844	1387	463	4694
2005	2851	1399	467	4717
2006	2896	1382	442	4720
2007	2864	1377	429	4670

Table 2: Group averages of individual factor loadings η_g

Skill group	Age group	Nb Obs	$\overline{\eta}_{g1}$	$\overline{\eta}_{g2}$	$\overline{\eta}_{g3}$
2	17	1268	2.4	0.04	-0.15
			(0.032)	(0.0067)	(0.051)
3	17	1224	2.4	0.039	-0.15
			(0.039)	(0.0056)	(0.04)
1	19	41	2.7	0.07	-0.33
			(0.038)	(0.0057)	(0.042)
2	19	934	2.6	0.044	-0.17
			(0.035)	(0.0046)	(0.034)
3	19	994	2.5	0.042	-0.17
			(0.04)	(0.007)	(0.051)
1	21	117	2.9	0.052	-0.22
			(0.086)	(0.0085)	(0.068)
2	21	710	2.7	0.047	-0.2
			(0.014)	(0.0024)	(0.018)
3	21	512	2.6	0.041	-0.19
			(0.015)	(0.0026)	(0.019)
1	23	171	3.1	0.055	-0.24
			(0.018)	(0.0036)	(0.027)
2	23	348	2.7	0.05	-0.21
			(0.026)	(0.0037)	(0.028)
3	23	254	2.7	0.051	-0.25
			(0.046)	(0.0053)	(0.04)
1	25	191	3.3	0.061	-0.29
			(0.056)	(0.0066)	(0.05)
2	25	146	2.8	0.038	-0.14
			(0.059)	(0.0065)	(0.046)
3	25	93	2.6	0.033	-0.09
			(0.018)	(0.0031)	(0.024)
1	27	114	3.4	0.047	-0.21
			(0.019)	(0.0045)	(0.034)
2	27	87	3	0.061	-0.32
			(0.02)	(0.0034)	(0.025)
3	27	63	2.7	0.03	-0.079
			(0.036)	(0.005)	(0.039)
1	29	58	3.2	0.041	-0.14
			(0.052)	(0.0058)	(0.041)
2	29	67	2.8	0.038	-0.2
			(0.084)	(0.013)	(0.11)
3	29	55	2.6	0.017	0.0061
			(0.048)	(0.0074)	(0.059)

Note: Estimation of equation (2). A flat spot deflator is used. Newey West standard errors in parentheses (5 lags).

Table 3: Estimated standard errors and correlations of individual effects η_i^c: Random effect estimation

	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3
σ_{η_1}	.302	.302	.301	.310	.306	.304	.306	.300	.298
	(.001)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.004)
σ_{η_2}	.038	.039	.039	.038	.039	.036	.038	.037	.037
	(.005)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
σ_{η_3}	.255	.259	.256	.263	.260	.248	.258	.247	.242
	(.005)	(.006)	(.006)	(.004)	(.005)	(.005)	(.005)	(.006)	(.007)
ρ_{η_1,η_2}	.473	.413	.454	.571	.486	.610	.505	.485	.365
	(.016)	(.021)	.021	(.013)	(.017)	(.013)	(.017)	(.020)	(.030)
ρ_{η_1,η_3}	- .604	- .548	- .586	- .694	- .618	- .729	- .636	- .620	- .509
	(.003)	(.020)	.019	(.011)	(.015)	(.012)	(.016)	(.019)	(.029)
ρ_{η_2,η_3}	- .946	- .948	- .947	- .945	- .946	- .941	- .946	- .943	- .944
	(.023)	(.003)	.003	(.002)	(.002)	(.003)	(.002)	(.003)	(.004)

Note: The first line corresponds to the ARMA specification (AR-MA) used for the random effect estimation. Standard errors in parentheses.

Table 4: Bias corrected covariance matrix of individual effects: fixed and random effect estimation

Sample periods	$Var(\eta_1)$	$Cov(\eta_1, \eta_2)$	$Cov(\eta_1, \eta_3)$	$Var(\eta_2)$	$Cov(\eta_2, \eta_3)$	$Var(\eta_3)$
21	0.18	0.012	-0.13	0.0034	-0.026	0.22
	(0.029)	(0.0032)	(0.03)	(0.00061)	(0.0049)	(0.043)
22	0.15	0.015	-0.13	0.0035	-0.027	0.22
	(0.019)	(0.0031)	(0.026)	(0.00067)	(0.0053)	(0.042)
23	0.16	0.012	-0.11	0.0035	-0.024	0.19
	(0.017)	(0.0024)	(0.02)	(5e-04)	(0.0038)	(0.03)
24	0.14	0.014	-0.12	0.0041	-0.03	0.23
	(0.017)	(0.0027)	(0.022)	(0.00061)	(0.0047)	(0.037)
25	0.13	0.01	-0.089	0.0028	-0.02	0.16
	(0.014)	(0.0023)	(0.019)	(0.00041)	(0.0033)	(0.027)
26	0.097	0.0066	-0.059	0.0025	-0.017	0.12
	(0.0072)	(0.00093)	(0.0072)	(0.00023)	(0.0016)	(0.012)
27	0.077	0.0046	-0.04	0.0017	-0.011	0.079
	(0.0046)	(0.00064)	(0.0047)	(0.00015)	(0.001)	(0.0074)
28	0.067	0.0036	-0.031	0.0015	-0.0097	0.067
	(0.0049)	(0.00067)	(0.0047)	(0.00016)	(0.0011)	(0.0074)
21+ sample	0.11	0.0078	-0.07	0.0025	-0.017	0.13
	(0.0073)	(0.00095)	(0.0092)	(0.00018)	(0.0015)	(0.013)
Random effects	0.093	0.0059	-0.05	0.0015	-0.0093	0.066
	(0.0036)	(0.00051)	(0.004)	(0.00011)	(0.00079)	(0.0059)

Notes: The first lines are obtained using fixed effect estimates. Sample periods = number of observed periods. Standard errors (heteroskedastic-consistent sampling and parameter uncertainty, 1000 MC simulations) between brackets. The working sample (21+) has 4873 observations

Table 5: Covariances of wage level and subsequent growth

Years	All	Low skilled	Medium skilled	High skilled
1977	-0.00594	-0.00573	-0.00665	-0.00688
	(0.000639)	(0.000564)	(0.000885)	(0.00214)
1982	-0.000972	-0.00137	-0.00102	1.04e-05
	(0.000388)	(0.000303)	(0.00052)	(0.00149)
1987	0.00178	0.00113	0.00219	0.00336
	(0.000271)	(0.000198)	(0.000402)	(0.00119)
1992	0.00233	0.0018	0.00299	0.00334
	(0.000207)	(0.000162)	(0.000357)	(0.00089)
1997	0.00135	0.00115	0.00204	0.00133
	(0.000174)	(0.000139)	(0.000324)	(0.000637)
2002	0.000812	0.000838	0.00141	0.00108
	(0.000269)	(0.000229)	(0.00042)	(0.00108)
2007	0.00524	0.00459	0.00583	0.0109
	(0.000666)	(0.000593)	(0.00114)	(0.00304)
Observations	4873	2942	1433	498

Notes: The coefficient is the covariance between $\eta_{i1}+\eta_{i2}t+\eta_{i3}\beta^{-t}$ and $\eta_{i2}+\beta^{-t}(1/\beta-1)\eta_{i3}$. Standard errors (heteroskedastic-consistent sampling and parameter uncertainty, 1000 MC simulations) between brackets. The working sample (21+) has 4873 observations

Table 6: Time varying correlation of initial levels and returns

Year	All	Low	Med	High
1977	-0.498	-0.674	-0.561	-0.379
	(0.0521)	(0.0425)	(0.0682)	(0.13)
1982	-0.505	-0.683	-0.564	-0.38
	(0.0529)	(0.043)	(0.0705)	(0.134)
1987	-0.509	-0.687	-0.561	-0.375
	(0.0539)	(0.0436)	(0.0742)	(0.139)
1992	-0.47	-0.635	-0.507	-0.31
	(0.0527)	(0.0436)	(0.0776)	(0.136)
1997	-0.186	-0.265	-0.201	0.000268
	(0.0385)	(0.042)	(0.0664)	(0.0796)
2002	0.156	0.207	0.186	0.212
	(0.0346)	(0.0392)	(0.0458)	(0.0765)
2007	0.286	0.389	0.338	0.276
	(0.0387)	(0.0387)	(0.0444)	(0.0888)
Long-run value	-0.373	-0.507	-0.411	-0.247
	(0.0407)	(0.0335)	(0.0596)	(0.0997)

Note: The correlation is $\rho = Corr(\eta_{i1} + \eta_{i3}, \eta_{i2} - \log(\beta)\beta^{-t}\eta_{i3})$ Only observations with more than 21 periods. 4873 observations. First column reports results for centered individual effects while the other columns include aggergate effects. The last three columns for low, medium and high skills.

Table 7: Impacts of unobserved heterogeneity on mean log wages

Impact of $\{\sigma_j\}_{j=1,.,3}$ on:	Level $\eta_1 \to \eta_1 + \sigma_1$	Slope $\eta_2 \to \eta_2 + \sigma_2$	Curvature $\eta_3 \to \eta_3 + \sigma_3$
Log-wage 1977	0.0112	0.0116	0.0199
	(0.000203)	(0.000336)	(0.00062)
Log-wage 1982	0.0303	0.0263	0.0448
	(0.000549)	(0.000765)	(0.0014)
Log-wage 1987	0.0325	0.0499	0.0601
	(0.000587)	(0.00145)	(0.00187)
Log-wage 1992	0.0326	0.0747	0.0778
	(0.00059)	(0.00217)	(0.00242)
Log-wage 1997	0.0326	0.0996	0.101
	(0.00059)	(0.0029)	(0.00313)
Log-wage 2002	0.0326	0.124	0.13
	(0.00059)	(0.00362)	(0.00405)
Log-wage 2007	0.0326	0.149	0.168
	(0.00059)	(0.00435)	(0.00523)
Long-run value	0.0287	0.0573	0.0672
	(0.00052)	(0.00167)	(0.00209)

Note: Average impact on log wages of an increase of a tenth of the standard deviation of: 2nd column, unobserved heterogeneity in the initial human capital; third column, wage growth or returns; fourth column, curvature or horizon.

Table 8: Variance decomposition: random and fixed effects

Random Effects	Obs. het. %	Unobs. het. %	Transitory %	Total var.
Log-wage 1977	65.2	1.18	33.6	0.481
Log-wage 1982	34.8	33.1	32.1	0.133
Log-wage 1987	32.8	42	25.3	0.153
Log-wage 1992	29.7	49.9	20.4	0.179
Log-wage 1997	27.2	56	16.8	0.201
Log-wage 2002	26.2	60.8	13	0.206
Log-wage 2007	25.1	66.4	8.47	0.202
Long-run value (1977-2007)	50.2	44.5	5.31	0.118

Note: The variance (5th column) of each variable (1st column) is decomposed into its component shares which are reported in percentages in column 2 (observed heterogeneity), column 3 (unobserved heterogeneity) and column 4 (transitory component). The share of variance of log 1982 wage (0.133) explained by observed heterogeneity is 34.8

Fixed Effects	Obs. het. %	Unobs. het. %	Transitory %	Total var.
Log-wage 1977	65	1.45	33.5	0.483
Log-wage 1982	33.5	35.7	30.9	0.138
Log-wage 1987	33.7	40.4	25.9	0.149
Log-wage 1992	31.2	47.3	21.5	0.17
Log-wage 1997	29	53	17.9	0.189
Log-wage 2002	28	58	14	0.193
Log-wage 2007	24.5	67.2	8.25	0.207
Long-run value	52.2	42.3	5.51	0.113

Note: See Table above. Bias corrected statistics. Only observations with more than 21 periods. 4873 observations.

Table 9: Counterfactual variances of log wages

Years	$\eta_1 = \bar{\eta}_{1g}$ $\eta_2 = \bar{\eta}_{2g}$ $\eta_3 = \bar{\eta}_{3g}$	Heterogeneity in: Levels	Growth	Curvature	Growth and Curvature	All
1977	0.306	0.312	0.307	0.308	0.311	0.313
	(0.0169)	(0.0169)	(0.0169)	(0.0169)	(0.017)	(0.017)
1982	0.0449	0.0924	0.0489	0.0547	0.0679	0.094
	(0.0181)	(0.0181)	(0.0181)	(0.0181)	(0.0181)	(0.0182)
1987	0.0491	0.103	0.0636	0.0667	0.0673	0.109
	(0.0197)	(0.0197)	(0.0197)	(0.0197)	(0.0198)	(0.0198)
1992	0.0523	0.107	0.0849	0.0818	0.085	0.133
	(0.0221)	(0.0221)	(0.0221)	(0.0221)	(0.0221)	(0.0222)
1997	0.0539	0.109	0.112	0.103	0.113	0.154
	(0.0257)	(0.0257)	(0.0257)	(0.0257)	(0.0257)	(0.0258)
2002	0.0533	0.108	0.144	0.136	0.144	0.165
	(0.0313)	(0.0313)	(0.0313)	(0.0313)	(0.0313)	(0.0313)
2007	0.05	0.105	0.18	0.188	0.188	0.189
	(0.0402)	(0.0402)	(0.0402)	(0.0402)	(0.0402)	(0.0402)
Long-run	0.0579	0.1	0.0771	0.0799	0.0801	0.106
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.0201)

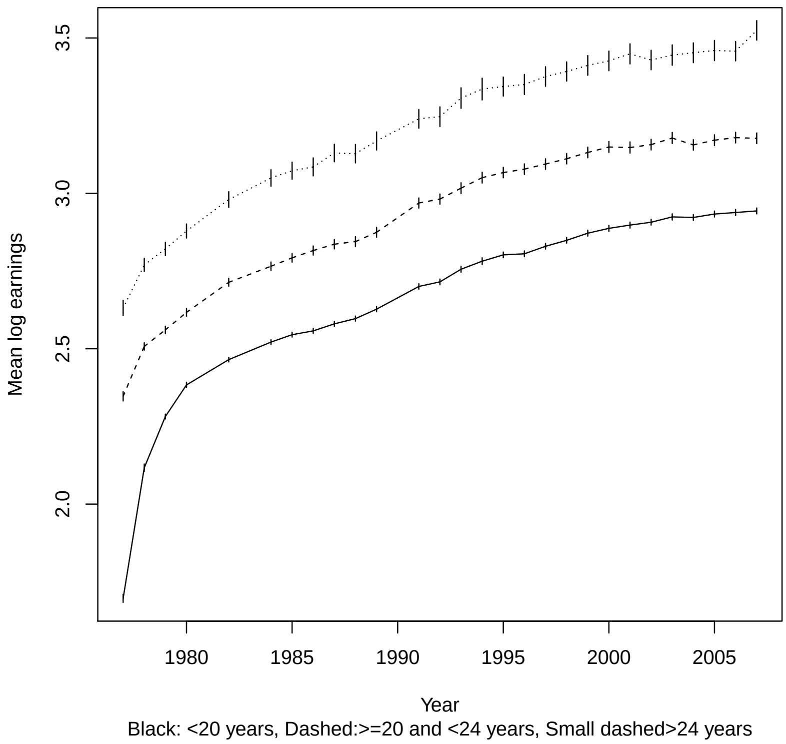
Note: Only observations with more than 21 periods. 4873 observations. The counterfactuals are described in the text and measure the influence of each component of heterogeneity, in levels, growth and curvature.

Table 10: Counterfactual variances of log wages: High skills

Years	$\eta_1 = \bar{\eta}_{1g}$ $\eta_2 = \bar{\eta}_{2g}$ $\eta_3 = \bar{\eta}_{3g}$	Heterogeneity in: Levels	Growth	Curvature	Growth and Curvature	All
1977	0.049	0.0636	0.0501	0.0522	0.0582	0.0642
	(0.0677)	(0.0677)	(0.0677)	(0.0677)	(0.0678)	(0.0681)
1982	0.0643	0.171	0.0698	0.0802	0.11	0.172
	(0.074)	(0.074)	(0.074)	(0.074)	(0.0741)	(0.0743)
1987	0.0767	0.199	0.0966	0.105	0.112	0.209
	(0.0814)	(0.0814)	(0.0814)	(0.0814)	(0.0815)	(0.0818)
1992	0.0871	0.211	0.132	0.135	0.135	0.248
	(0.0901)	(0.0901)	(0.0901)	(0.0901)	(0.0902)	(0.0904)
1997	0.0929	0.216	0.172	0.173	0.174	0.276
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
2002	0.092	0.216	0.216	0.226	0.226	0.287
	(0.112)	(0.112)	(0.112)	(0.112)	(0.112)	(0.112)
2007	0.0837	0.207	0.262	0.307	0.322	0.333
	(0.126)	(0.126)	(0.126)	(0.126)	(0.126)	(0.127)
Long-run	0.0725	0.168	0.0988	0.108	0.114	0.177
	(0.0814)	(0.0815)	(0.0814)	(0.0814)	(0.0816)	(0.0818)

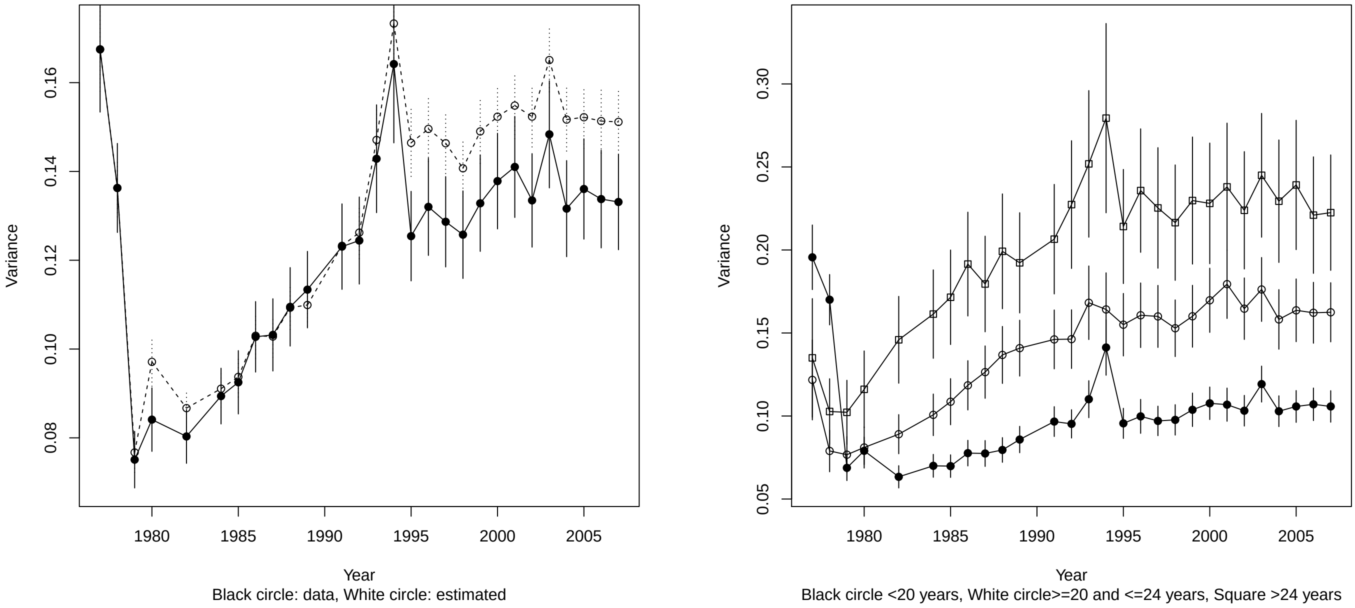
Note: High skills. Only observations with more than 21 periods. 498 observations. The counterfactuals are described in the text and measure the influence of each component of heterogeneity, in levels, growth and curvature.

Figure 1: Mean log earnings by age at entry: 1977-2007



Note: The small vertical lines represent the 95% confidence intervals.

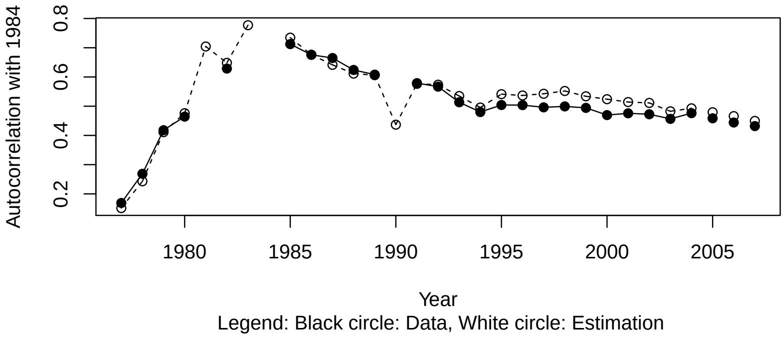
Figure 2: Cross-sectional variance of log wage residuals: 1977-2007



(A) full sample (B) by age of entry

Note: The small vertical lines represent the 95% confidence intervals. Log wage residuals are obtained by regressing log wages on a saturated set of dummies for skill groups and years.

Figure 3: Autocorrelations with 1986 and 2007



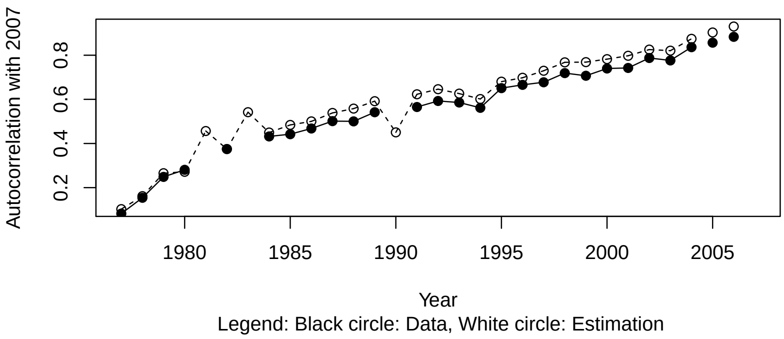
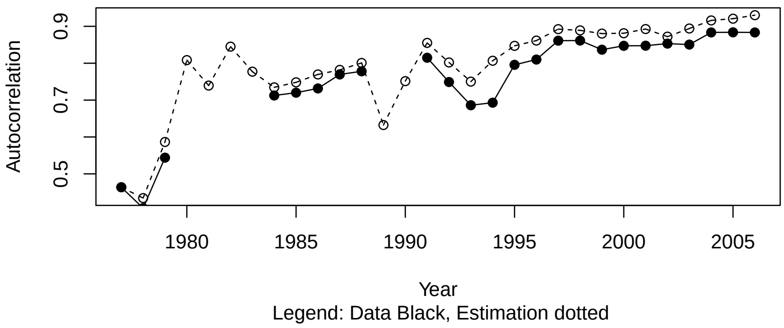


Figure 4: Forward autocorrelations of order 1 and of order 6



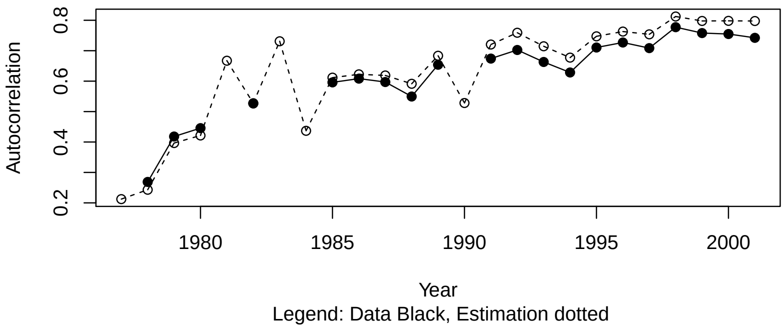
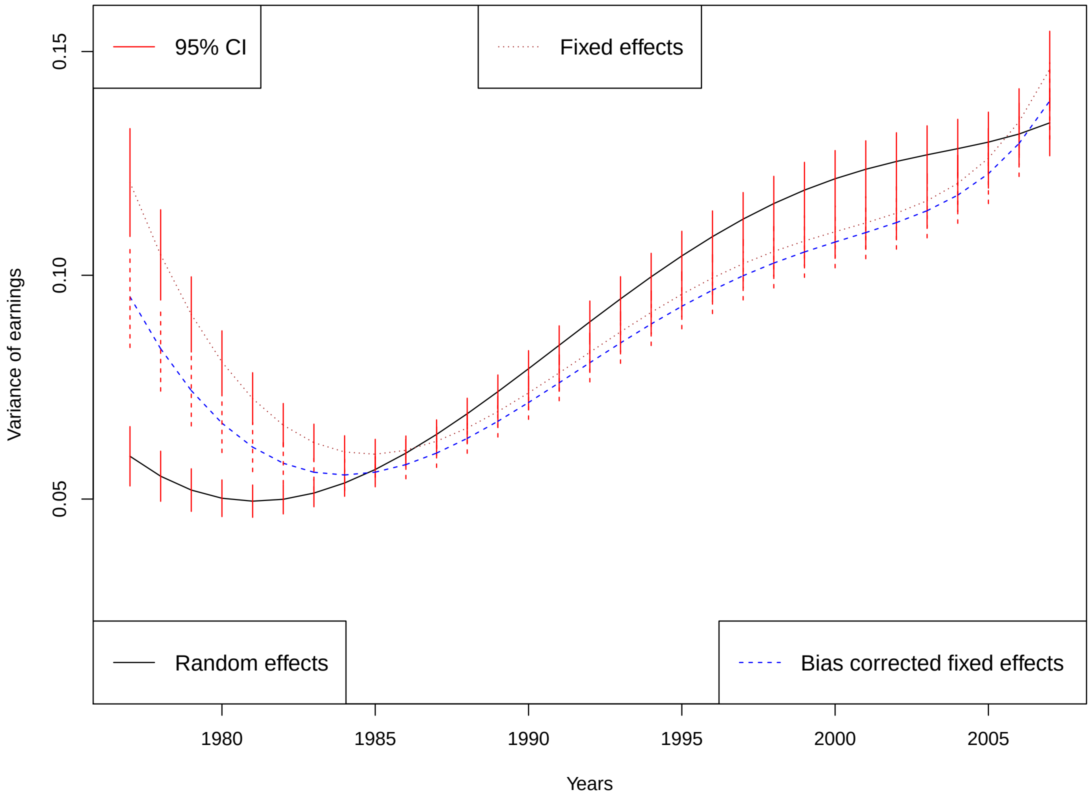


Figure 5: Variance of the permanent components



Note: The permanent component is $M(\beta)\eta$ defined in equation (6). The sample is restricted to long history profiles (more than 21 periods). "Random effects" are using estimates derived from random effect estimation. "Fixed effects" are using estimates derived from raw fixed effect estimation and "Bias corrected f.e." are the bias corrected version of them.

Supplementary Appendices
Only for Web publication

S.I Notation

S.I.1 The model

- t: time elapsed since the entry in the labor market.
- i: index for individuals.
- β: homogenous discount rate
- T: Arbitrary date at which we examine whether individuals goes on investing in human capital, last date of observation in the empirical application.
- η_{i1}: individual-specific fixed level of log-wages.
- η_{i2}: individual-specific growth rate of log-wages.
- η_{i3}: individual-specific degree of curvature of log-wages.
- v_{it}: (log) price of human capital net of cumulative depreciation.
- g: group of workers, defined by their age at entry and their skills
- $\overline{\ln y}_{gt}$: average of $\ln y_{it}$ over the group g
- η_{gk}: average of η_{ik} over g, for $k = 1, 2, 3$
- v_{gt}: average of v_{it} over g
- η_{ik}^c: centered individual effect of η_{ik}, for $k = 1, 2, 3$
- u_{it}: centered wages, with respect to group g
- v_{it}^c: individual-specific variations of human capital prices

S.I.2 Econometric Modeling

- $M(\beta)$: $T, 3$ matrix of factors.
- $\Omega(\eta_i^c)$: covariance matrix of centered individual fixed effects.
- $\hat{\eta}_i^c$: estimate of the centered individual fixed effect.
- B: matrix $3, T$ establishing the relationship between the centered individual fixed effects and the wages residuals.

- $\hat{B}$: estimate of B
- $\tilde{\eta}_i^c$: unfeasible estimator of η_i^c using B
- ξ_i: T-vector of residuals, orthogonal to η_i^c
- Ω_ξ: covariance matrix of ξ_i
- T_i: number of actual observations for the individual i.

S.II The random effect model : Model Specification and Likelihood function

The main difference with standard specifications lies in the introduction of three individual heterogeneity factors that interact in a specific way with factors depending on time. Equation (7) writes

$$u_i^{[1,T]} = M(\beta)^{[1,T]} \eta_i^c + v_i^{c[1,T]}$$

where $u_i^{[1,T]} = (u_{i1}, ..., u_{iT})'$, $v_i^{c[1,T]} = (v_{i1}^c, ..., v_{iT}^c)'$, $\eta_i^c = (\eta_{i1}^c, \eta_{i2}^c, \eta_{i3}^c)$ are the centered versions of the ηs and:

$$M(\beta)^{[1,T]} = \begin{bmatrix} 1 & 1 & 1/\beta \\ \vdots & \vdots & \vdots \\ 1 & T & 1/\beta^T \end{bmatrix},$$

is a $[T, 3]$ matrix. The system is further completed by initial conditions, the number of which depends on the order of the autoregressive process. Denote p this order and write the initial conditions as:

$$u_i^{[1-p,0]} = v_i^{c[1-p,0]}$$

since unrestricted dependence between $v_i^{[1,T]}$, η_i^c and those initial conditions will be allowed for. We can rewrite the whole system as:

$$u_i^{[1-p,T]} = M(\beta)^{[1-p,T]} \eta_i^c + v_i^{c[1-p,T]}$$

in which the matrix $M(\beta)^{[1-p,T]}$ is completed by p rows equal to zero, $M(\beta)^{[1-p,0]} = 0$.

We now go further and specify the correlation structure. A comment is in order. Usually, the autoregressive structure directly applies to wage shocks u_{it} and in the absence of covariates, this is equivalent to specifying it through the residual part v_{it}^c because there is a single individual effect. This equivalence still holds when another heterogeneity factor interacted with a linear trend is present. Nevertheless, our specification includes a third factor interacted with a geometric term and this breaks the equivalence. To circumvent this problem, we posit that v_{it}^c is a (time heteroskedastic) ARMA process whose innovations are independent of the individual heterogeneity terms, η_i^c. As a consequence, our variable of interest, u_{it}, is the sum of two processes, the first one being related to fixed individual heterogeneity and the second one to the pure dynamic process. These processes are assumed to be independent between them after period 1 although they are both correlated with initial conditions, $u_i^{[1-p,0]}$.

We now derive the covariance matrix of $u_i^{[1-p,T]}$ as a function of the parameters of these processes in two steps . We first describe the ARMA process and then include the individual heterogeneity factors.

S.II.1 Time heteroskedastic ARMA specification

Following Alvarez and Arellano (2004) or Guvenen (2009), we specify

$$v_{it}^c = \alpha_1 v_{it-1}^c + ... + \alpha_p v_{it-p}^c + \sigma_t w_{it}$$

where w_{it} is $MA(q)$:

$$w_{it} = \zeta_{it} - \psi_1 \zeta_{it-1} - ... - \psi_q \zeta_{it-q}.$$

Let $\alpha = (\alpha_1, ., \alpha_p)$ and $M_T(\alpha)$ a matrix of size $[T, T+p]$ where $p = \dim(\alpha)$:

$$M_T(\alpha) = \begin{pmatrix} -\alpha_p & ... & -\alpha_1 & 1 & 0 & ... & 0 \\ 0 & -\alpha_p & ... & -\alpha_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & 0 \\ 0 & ... & 0 & -\alpha_p & ... & -\alpha_1 & 1 \end{pmatrix}.$$

If $v_i^{c[1-p,T]} = \left(v_{i1-p}^c, ..., v_{iT}^c\right)$, we have:

$$\begin{pmatrix} \begin{pmatrix} I_p & 0 \end{pmatrix} \\ M_T(\alpha) \end{pmatrix} v_i^{[1-p,T]} = \begin{pmatrix} v_i^{[1-p,0]} \\ \sigma_t w_i^{[1,T]} \end{pmatrix}$$

Since w_{it} is $MA\,(q)$, we have

$$w_i^{[1,T]} = M_T(\psi).\zeta_i^{[1-q,T]}$$

where $\zeta_i^{[1-q,T]} = (\zeta_{i1-q}, ..., \zeta_{iT})$.

Denote Λ a diagonal matrix whose diagonal is $(\sigma_1, ., \sigma_T)$ to get the following description of the stochastic process as a function of initial conditions and idiosyncratic errors:

$$\begin{pmatrix} I_p \quad 0 \\ M_T(\alpha) \end{pmatrix} .v_i^{c[1-p,T]} = \begin{pmatrix} I_p & 0_{p,T+q} \\ 0_{T,p} & \Lambda.M_T(\psi) \end{pmatrix} \begin{pmatrix} v_i^{c[1-p,0]} \\ \zeta_i^{[1-q,T]} \end{pmatrix}. \tag{S.II.1}$$

To compute the covariance of $v_i^{c[1-p,T]}$, we derive the covariance matrix of $\begin{pmatrix} v_i^{c[1-p,0]} & \zeta_i^{[1-q,T]} \end{pmatrix}$. Since $\zeta_i^{[1-q,T]}$ are i.i.d and are of variance 1, the South-East corner of the matrix is the identity matrix of size $(1+q+T)$. The North West corner is assumed to be an unrestricted covariance matrix $Vu_i^{[1-p,0]} = \Gamma_{00}$. Assuming as usual that $E(u_{i\tau}\zeta_{it}) = 0$ for any $\tau < t$, we have that $E(v_i^{c[1-p,0]}.(\zeta_i^{[1,T]})') = 0$. Only $E(u_i^{[1-p,0]}.(\zeta_i^{[1-q,0]})')$ remains to be defined:

$$E(v_i^{c[1-p,0]}.(\zeta_i^{[1-q,0]})') = \Omega = [\omega_{rs}]$$

where $r \in [1-p,0]$ and $s \in [1-q,0]$ and where:

$$\begin{array}{rc} r < s: & \omega_{rs} = 0 \\ r \geq s: & \omega_{rs} \text{ is not constrained} \end{array}$$

because the innovation ζ_{is} is drawn after r and is assumed to be not correlated with y_i^r.

Hence the covariance matrix of $z_i = \begin{pmatrix} v_i^{c[1-p,0]} \\ \zeta_i^{[1-q,T]} \end{pmatrix}$ writes :

$$\Omega_z = V \begin{pmatrix} v_i^{c[1-p,0]} \\ \zeta_i^{[1-q,T]} \end{pmatrix} = V \begin{pmatrix} v_i^{c[1-p,0]} \\ \zeta_i^{[1-q,0]} \\ \zeta_i^{[1,T]} \end{pmatrix} = \begin{pmatrix} \Gamma_{00} & \Omega & 0 \\ \Omega' & I_q & 0 \\ 0 & 0 & I_T \end{pmatrix}.$$

S.II.2 Individual heterogeneity

The covariance matrix of the individual heterogeneity factors is denoted Σ_η. as said above, we assume that the fixed heterogeneity terms are independent of the whole innovation process $\zeta_i^{[1-q,T]}$. As for the covariance structure between initial conditions and those factors, we assume that:

$$E\left(v_i^{c[1-p,0]} \eta_i^{c\prime}\right) = \Gamma_{0\eta}$$

Consider the covariance matrix of initial conditions Σ :

$$\Sigma = V \begin{pmatrix} v_i^{c[1-p,0]} \\ \eta_i^c \\ \zeta_i^{[1-q,0]} \end{pmatrix} = \begin{pmatrix} \Gamma_{00} & \Gamma_{0\eta} & \Omega \\ \Gamma'_{0\eta} & \Sigma_\eta & 0 \\ \Omega & 0 & I_q \end{pmatrix}.$$

and define,

$$\begin{aligned} R_T(\alpha) &= \begin{pmatrix} \begin{pmatrix} I_p & 0 \end{pmatrix} \\ M_T(\alpha) \end{pmatrix}^{-1} \\ S_{T,p}(\psi, \Lambda) &= \begin{pmatrix} I_p & 0_{p,T+q} \\ 0_{T,p} & \Lambda.M_T(\psi) \end{pmatrix} \end{aligned}$$

Write the covariance matrix of vector $u_i^{[1-p,T]}$:

$$\begin{aligned} \Omega_u &= V\left(u_i^{[1-p,T]}\right) = V\left(v_i^{c[1-p,T]} + M(\beta)^{[1-p,T]} \eta_i^c\right) \\ &= V\left(\left[M(\beta)^{[1-p,T]}, R_T(\alpha).S_{T,p}(\psi,\Lambda)\right] \begin{pmatrix} \eta_i^c \\ v_i^{c[1-p,0]} \\ \zeta_i^{[1-q,T]} \end{pmatrix}\right) \end{aligned}$$

Since $v_i^{c[1-p,T]} = R_T(\alpha).S_{T,p}(\psi,\Lambda) \begin{pmatrix} v_i^{c[1-p,0]} \\ \zeta_i^{[1-q,T]} \end{pmatrix}$, the matrix

$$V\left(v_i^{c[1-p,T]}\right) = R_T(\alpha).S_{T,p}(\psi,\Lambda).\Omega_z.S_{T,p}(\psi,\Lambda)' R_T(\alpha)'$$

and

$$
\begin{aligned}
E\left(v_i^{c[1-p,T]}\eta_i'\right) M\left(\beta\right)^{[1-p,T]\prime} &= R_T(\alpha).S_{T,p}(\psi,\Lambda)E\begin{pmatrix} v_i^{c[1-p,0]}(\eta_i^c)' \\ \zeta_i^{[1-q,T]}(\eta_i^c)' \end{pmatrix} M\left(\beta\right)^{[1-p,T]\prime} \\
&= R_T(\alpha).S_{T,p}(\psi,\Lambda)\begin{pmatrix} \Gamma_{0\eta} \\ 0_{T+q,3} \end{pmatrix} M\left(\beta\right)^{[1-p,T]\prime} \\
&= R_T(\alpha).\begin{pmatrix} I_p & 0_{p,T+q} \\ 0_{T,p} & \Lambda.M_T(\psi) \end{pmatrix}\begin{pmatrix} \Gamma_{0\eta} \\ 0_{T+q,3} \end{pmatrix}\left(0_{3,p}, M\left(\beta\right)^{[1,T]\prime}\right) \\
&= R_T(\alpha).\begin{pmatrix} I_p & 0_{p,T+q} \\ 0_{T,p} & \Lambda.M_T(\psi) \end{pmatrix}\begin{pmatrix} 0_{p,p} & \Gamma_{0\eta}M\left(\beta\right)^{[1,T]\prime} \\ 0_{T+q,p} & 0_{T+q,T} \end{pmatrix} \\
&= R_T(\alpha).\begin{pmatrix} 0_{p,p} & \Gamma_{0\eta}M\left(\beta\right)^{[1,T]\prime} \\ 0_{T,p} & 0_{T,T} \end{pmatrix}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\Omega_u &= R_T(\alpha).S_{T,p}(\psi,\Lambda).\Omega_z.S_{T,p}(\psi,\Lambda)'R_T(\alpha)' + M\left(\beta\right)^{[1-p,T]}\Sigma_\eta M\left(\beta\right)^{[1-p,T]\prime} \\
&\quad + R_T(\alpha).\begin{pmatrix} 0_{p,p} & \Gamma_{0\eta}M\left(\beta\right)^{[1,T]\prime} \\ 0_{T,p} & 0_{T,T} \end{pmatrix} + \begin{pmatrix} 0_{p,p} & 0_{p,T} \\ M\left(\beta\right)^{[1,T]}\Gamma_{0\eta}' & 0_{T,T} \end{pmatrix} R_T(\alpha)'
\end{aligned}
$$

The two first terms correspond to variances of the dynamic process and the individual heterogeneity factors, the other terms correspond to the correlation between the two processes induced by initial conditions. Note that the parameters of the MA process don't appear in the correlation between the two processes since innovations are assumed to be independent with individual heterogeneity factors. Initial conditions are given by $\zeta_i^{[1-q,0]}$, η^c and $v_i^{c[1-p,0]}$.

The Choleski decomposition of matrix Σ can be parametrized expressing the following matrix into a polar coordinate basis.

$$
\begin{pmatrix}
1 & 0 & \cdots & & & & & \cdots & 0 & \\
0 & \ddots & \ddots & \ddots & & & & & \vdots & \\
\vdots & 0 & 1 & 0 & & & \ddots & & & \\
 & \cdots & 0 & 1 & 0 & & 0 & \ddots & & \\
 & & & \omega_\eta^{12} & 1 & 0 & & & & \\
 & & 0 & \omega_\eta^{13} & \omega_\eta^{23} & 1 & 0 & & & \\
\vdots & & \theta^{(1)}_{1-q,1-p} & \theta_{\eta 1,1-p} & \theta_{\eta 2,1-p} & \theta_{\eta 3,1-p} & 1 & & & \\
0 & & & \vdots & \vdots & \vdots & \theta_{2-p,2-p} & \ddots & 1 & \\
\theta^{(1)}_{0,0} & & & \theta_{\eta 1,0} & \theta_{\eta 21,0} & \theta_{\eta 3,0} & \cdots & \ddots & \theta_{0,0} & 1
\end{pmatrix}
$$

where $\theta^{(1)}_{1-q,1-p} = 0$ if $p > q$ and, more generally, $\theta^{(1)}_{l,m} = 0$ if $l > m$.

S.III Fixed Effect Estimation

The main equation is:

$$u_i^{[1-p,T]} = M(\beta)^{[1-p,T]}\eta_i^c + v_i^{c[1-p,T]},$$

where η_i^c and $v_i^{c[1-p,T]}$ are centered by construction and where a row of $M(\beta)$ is defined as $M(\beta)^{[t]} = (1, t, 1/\beta^t)$ as in Appendix S.II (with some 0s between $1-p$ and 0).

Later on, we shall reintroduce the estimated averages, $\overline{\eta}_g$, of the individual effects that we estimate by OLS using the sub-groups defined by age of entry and skill level (21 groups). Define the average in each group as $\bar{y}_g^{[1-p,T]}$ and define:

$$\widehat{\overline{\eta}}_g = (M(\beta)^{[1-p,T]\prime} M(\beta)^{[1-p,T]})^{-1} M(\beta)^{[1-p,T]\prime} \bar{y}_g^{[1-p,T]}.$$

We now present the fixed effect estimation of η_i^c. We consider first the case with no missing values and extend it to the case with missing values.

Assume first that there are no missing values. To deal with the correlation between η_i^c and v_i, we can always write:

$$v_i^{c[1-p,T]} = C\eta_i^c + \xi_i^{[1-p,T]},$$

where $E((\eta_i^c)'\xi_i^{[1-p,T]}) = 0$ so that we get:

$$C = E(v_i^{c[1-p,T]}(\eta_i^c)')(E(\eta_i^c(\eta_i^c)'))^{-1},$$

and:

$$\Omega_\xi = E(v_i^{c[1-p,T]} v_i^{c[1-p,T]\prime}) - E(v_i^{c[1-p,T]}(\eta_i^c)')(E(\eta_i^c(\eta_i^c)'))^{-1} E(\eta_i^c v_i^{c[1-p,T]\prime}).$$

This yields the estimating equation for η_i^c :

$$u_i^{[1-p,T]} = D\eta_i^c + \xi_i^{[1-p,T]} \text{ where } D = M(\beta)^{[1-p,T]} + C,$$

that we can estimate by GLS methods since D can be estimated using random effect methods. This yields:

$$\tilde{\eta}_i^c = Bu_i^{[1-p,T]},$$

in which:

$$B = (D'\Omega_\xi^{-1}D)^{-1}D'\Omega_\xi^{-1}.$$

Furthermore:

$$\tilde{\eta}_i^c = B(D\eta_i^c + \xi_i^{[1-p,T]}) = \eta_i^c + B\xi_i^{[1-p,T]},$$

is such that:

$$\begin{aligned} V(\tilde{\eta}_i^c) &= EV(\tilde{\eta}_i^c \mid \eta_i^c) + VE(\tilde{\eta}_i^c \mid \eta_i^c) \\ &\Longrightarrow V(\tilde{\eta}_i^c) = B\Omega_\xi B' + V\eta_i^c. \end{aligned}$$

The term $B\Omega_\xi B'$ goes to zero at least at the rate $1/T$ since matrix D is determined by different factors which are going to zero at least as fast as T^{-1}.

The feasible estimator is now given by:

$$\hat{\eta}_i^c = \hat{B}u_i^{[1-p,T]},$$

and by reinclusion of the estimated averages for each group, $\bar{\eta}_{g \ni i} = \bar{\eta}_g$, we have:

$$\hat{\eta}_i = \bar{\eta}_g + \hat{\eta}_i^c = \bar{\eta}_g + \hat{B} u_i^{[1-p,T]},$$

Finally, the case with missing values is as follows. Suppose that $u_i^{[1-p,T]}$ is not observable, only $S_i u_i^{[1-p,T]}$ is where S_i is the matrix of dimension $(T_i, T+p+1)$ selecting non missing values and where T_i is the number of such non missing values. Consequently,

$$\tilde{\eta}_i^c = B S_i u_i^{[1-p,T]},$$

and by analogy to results above, we have

$$\hat{\eta}_i^c = \hat{B}_i u_i^{[1-p,T]},$$

in which $\hat{B}_i$ is a plug-in estimate of:

$$B_i \quad = (D' S_i' (S_i \Omega_\xi S_i')^{-1} S_i D)^{-1} D' S_i' (S_i \Omega_\xi S_i')^{-1}.$$

Table S.i: Non Missing Values

	1977	1979	1980	1985	1986	1987	1988	1989	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
1977	1																							
1978	1																							
1979	.882	.882																						
1980	.868	.786	.868																					
1982	1	.882	.868																					
1984	1	.882	.868																					
1985	.849	.751	.743	.849																				
1986	.834	.739	.731	.75	.834																			
1987	.804	.714	.704	.718	.737	.804																		
1988	.765	.675	.668	.694	.690	.691	.765																	
1989	.777	.689	.677	.701	.694	.691	.689	.777																
1991	.743	.658	.65	.67	.663	.655	.649	.678	.743															
1992	.736	.653	.647	.663	.655	.649	.642	.662	.679	.736														
1993	.749	.665	.653	.657	.666	.654	.631	.652	.659	.673	.749													
1994	.581	.515	.506	.508	.518	.511	.492	.506	.513	.517	.544	.581												
1995	.725	.643	.634	.636	.644	.632	.609	.628	.63	.635	.661	.535	.725											
1996	.721	.641	.631	.631	.638	.627	.603	.622	.622	.627	.652	.521	.671	.721										
1997	.71	.629	.621	.622	.63	.619	.596	.613	.612	.618	.642	.511	.649	.661	.71									
1998	.708	.628	.619	.618	.625	.615	.591	.61	.609	.614	.636	.506	.642	.649	.667	.708								
1999	.708	.628	.617	.617	.623	.614	.59	.61	.605	.609	.63	.502	.635	.639	.652	.665	.708							
2000	.701	.622	.611	.612	.62	.61	.583	.6	.595	.601	.623	.497	.625	.629	.637	.649	.662	.701						
2001	.687	.61	.598	.599	.605	.595	.573	.589	.584	.587	.605	.479	.608	.612	.62	.629	.639	.65	.687					
2002	.67	.595	.586	.588	.591	.581	.559	.575	.568	.573	.592	.471	.59	.594	.597	.606	.613	.617	.621	.67				
2003	.616	.547	.539	.544	.542	.532	.516	.533	.526	.53	.539	.425	.538	.541	.546	.553	.561	.564	.563	.577	.616			
2004	.63	.559	.551	.552	.556	.545	.523	.541	.534	.539	.555	.441	.555	.557	.559	.567	.573	.574	.574	.584	.565	.63		
2005	.634	.560	.552	.554	.558	.548	.526	.544	.536	.541	.558	.446	.557	.558	.559	.566	.570	.574	.571	.574	.543	.574	.634	
2006	.634	.561	.553	.556	.557	.549	.525	.544	.535	.541	.556	.444	.553	.556	.557	.563	.568	.570	.567	.574	.538	.566	.586	.634
2007	.627	.557	.547	.55	.552	.542	.521	.538	.531	.535	.548	.436	.547	.549	.551	.556	.560	.562	.557	.561	.525	.552	.570	.591

Notes: Frequencies of observations present in the sample at years described by row and column, relative to the full sample

Table S.ii: Autocorrelation matrix of earnings residuals

	1978	1979	1980	1982	1984	1985	1986	1987	1988	1989	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
1977	.438																										
1978	.280	.424																									
1979	.241	.367	.563																								
1980	.211	.343	.478	.539																							
1982	.223	.326	.439	.499	.733																						
1984	.221	.306	.401	.411	.665	.814																					
1985	.216	.301	.368	.430	.643	.785	.807																				
1986	.161	.266	.386	.441	.634	.767	.772	.853																			
1987	.156	.260	.401	.459	.634	.756	.744	.809	.871																		
1988	.134	.254	.368	.421	.617	.733	.730	.776	.830	.874																	
1989	.135	.239	.321	.383	.557	.682	.681	.726	.790	.824	.857																
1991	.145	.221	.334	.370	.577	.685	.679	.721	.765	.798	.821	.887															
1992	.134	.193	.306	.333	.515	.619	.619	.667	.724	.738	.762	.831	.854														
1993	.111	.179	.274	.314	.482	.607	.606	.644	.695	.709	.723	.810	.803	.823													
1994	.102	.183	.280	.330	.480	.590	.580	.632	.696	.711	.735	.809	.815	.810	.792												
1995	.109	.197	.289	.319	.491	.589	.582	.624	.686	.711	.746	.802	.815	.804	.795	.836											
1996	.128	.192	.305	.315	.497	.623	.623	.653	.720	.741	.764	.826	.839	.827	.816	.854	.878										
1997	.129	.198	.308	.336	.507	.625	.614	.656	.716	.737	.761	.828	.842	.833	.816	.862	.883	.932									
1998	.108	.194	.294	.316	.496	.618	.610	.651	.707	.735	.756	.819	.835	.813	.797	.838	.859	.904	.939								
1999	.117	.160	.294	.291	.478	.600	.594	.638	.689	.714	.730	.791	.815	.799	.784	.812	.837	.881	.908	.904							
2000	.124	.179	.293	.310	.501	.619	.613	.635	.696	.715	.741	.808	.822	.802	.795	.820	.830	.885	.919	.913	.908						
2001	.122	.180	.294	.296	.463	.588	.591	.616	.656	.685	.707	.776	.787	.767	.751	.779	.798	.855	.884	.880	.874	.912					
2002	.122	.179	.257	.261	.415	.543	.558	.568	.577	.605	.622	.695	.720	.694	.697	.716	.720	.785	.810	.811	.811	.844	.875				
2003	.128	.168	.291	.299	.469	.589	.585	.616	.669	.697	.715	.780	.794	.770	.763	.787	.799	.858	.887	.883	.877	.916	.914	.862			
2004	.108	.170	.289	.296	.462	.593	.584	.610	.666	.691	.707	.773	.784	.763	.757	.781	.792	.849	.876	.877	.873	.905	.903	.854	.950		
2005	.103	.155	.291	.287	.470	.595	.587	.619	.671	.698	.709	.776	.794	.771	.770	.790	.800	.853	.878	.878	.875	.903	.901	.857	.942	.957	
2006	.106	.157	.286	.279	.449	.572	.558	.591	.638	.670	.677	.738	.754	.745	.732	.757	.770	.819	.840	.845	.841	.872	.874	.828	.909	.931	.952

Note: In each cell, the correlation is computed using the individuals who are in the data both relevant years. Table S.i of the supplementary appendix presents the number of contributing individuals in each cell.

Table S.iii: AIC criterion

ARMA(p,q)	q=1	q=2	q=3
p=1	-344885	-344899	-344906
	(43)	(45)	(47)
p=2	-345301	-345447	-345733
	(47)	(50)	(53)
p=3	-345839	-346133	-346293
	(51)	(54)	(58)

AIC criterion computed as -2log$(L)+2K$, with L the likelihood and K the number of parameters. Number of parameters in brackets.

Table S.iv: Estimated parameters of the Random Effects Model

	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3
α_1	.702	.729	.711	.263	.186	.220	.200	.203	.194
	(.005)	(.006)	(.007)	(.011)	(.011)	(.011)	(.012)	(.011)	(.011)
α_2				.145	.324	.143	.191	.143	.161
				(.004)	(.008)	(.009)	(.005)	(.009)	(.009)
α_3							.022	.087	.187
							(.003)	(.004)	(.008)
ψ_1	.369	.391	.373	- .091	- .172	- .135	- .164	- .166	- .189
	(.005)	(.005)	(.007)	(.011)	(.011)	(.012)	(.012)	(.011)	(.011)
ψ_2		.020	.017		.170	- .028		- .046	- .046
		(.003)	(.003)		(.006)	(.008)		(.008)	(.008)
ψ_3			- .012			- .080			.114
			(.004)			(.004)			(.007)
σ_{η_1}	.302	.302	.301	.310	.306	.304	.306	.300	.298
	(.001)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.004)
σ_{η_2}	.038	.039	.039	.038	.039	.036	.038	.037	.037
	(.005)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
σ_{η_3}	.255	.259	.256	.263	.260	.248	.258	.247	.242
	(.005)	(.006)	(.006)	(.004)	(.005)	(.005)	(.005)	(.006)	(.007)
ρ_{η_1,η_2}	.473	.413	.454	.571	.486	.610	.505	.485	.365
	(.016)	(.021)	.021	(.013)	(.017)	(.013)	(.017)	(.020)	(.030)
ρ_{η_1,η_3}	- .604	- .548	- .586	- .694	- .618	- .729	- .636	- .620	- .509
	(.003)	(.020)	.019	(.011)	(.015)	(.012)	(.016)	(.019)	(.029)
ρ_{η_2,η_3}	- .946	- .948	- .947	- .945	- .946	- .941	- .946	- .943	- .944
	(.023)	(.003)	.003	(.002)	(.002)	(.003)	(.002)	(.003)	(.004)
σ_{y_0}	.491	.506	.496	.448	.479	.429	.442	.455	.494
	(.000)	(.007)	(.007)	(.004)	(.005)	(.004)	(.004)	(.005)	(.008)
$\sigma_{y_{-1}}$				.381	.424	.359	.387	.386	.428
				(.004)	(.005)	(.004)	(.004)	(.005)	(.008)
$\sigma_{y_{-2}}$							.264	.270	.299
							(.004)	(.006)	(.008)
$cov(\eta_1, y_0)$	- .227	- .257	- .237	- .156	- .214	- .149	-.186	- .201	- .282
	(.019)	(.017)	.017	(.015)	(.016)	(.016)	(.016)	(.017)	(.019)
$cov(\eta_1, y_{-1})$				- .127	- .183	- .113	- .153	- .168	- .253
				(.016)	(.017)	(.017)	(.017)	(.018)	(.020)
$cov(\eta_1, y_{-2})$							- .169	- .185	- .267
							(.018)	(.019)	(.022)
$cov(\eta_2, y_0)$	.358	.402	.374	.232	.335	.155	.219	.253	.361
	(.022)	(.020)	.021	(.017)	(.019)	(.021)	(.020)	(.022)	(.026)
$cov(\eta_2, y_{-1})$				.218	.331	.119	.242	.235	.352
				(.019)	(.021)	(.024)	(.022)	(.025)	(.029)
$cov(\eta_2, y_{-2})$							.239	.253	.351
							(.024)	(.027)	(.032)
$cov(\eta_3, y_0)$	- .290	- .333	- .305	- .179	- .270	- .107	- .163	- .195	- .291
	(.018)	(.023)	.023	(.020)	(.022)	(.023)	(.023)	(.024)	(.029)
$cov(\eta_3, y_{-1})$				- .169	- .272	- .077	- .190	- .181	- .287
				(.021)	(.023)	(.025)	(.023)	(.027)	(.032)
$cov(\eta_3, y_{-2})$							- .181	- .194	- .282
							(.026)	(.029)	(.035)
$cov(y_0, \zeta_0)$	.809	.036	- .024	- .823	.826	- .931	.841	- .795	.812
	(.023)	(8.525)	26.529	(.269)	(.059)	(.207)	(.061)	(.416)	(.096)
$cov(y_0, \zeta_{-1})$		.779	- .012		.408	- .352		- .208	.361
		(.438)	1.245		(.102)	(17.542)		(152.666)	(31.114)
$cov(y_{-1}, \zeta_{-1})$			.798		.722	- .066		.830	.234
			(.813)		(.062)	(.148)		(41.955)	(17.858)
$cov(y_0, \zeta_{-2})$						- .805			- .719
						(3.931)			(76.705)
$cov(y_{-1}, \zeta_{-2})$						- .382			- .202
						(11.249)			(44.061)
$cov(y_{-2}, \zeta_{-2})$									.752
									(.094)

Table S.v: Yearly standard deviation of earnings

	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3
1978	.311	.312	.312						
	(.001)	(.002)	(.002)						
1979	.254	.257	.255	.222	.232	.219			
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)			
1980	.223	.223	.223	.222	.227	.221	.224	.224	.230
	(.005)	(.001)	(.001)	(.001)	(.001)	(.001)	(.002)	(.002)	(.002)
1981	.264	.260	.263	.000	.103	.002	.004	.006	.001
	(.005)	(.005)	(.005)	(.096)	(.040)	(.066)	(.082)	(.076)	(.060)
1982	.152	.150	.150	.194	.193	.197	.193	.195	.198
	(.005)	(.005)	(.005)	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
1983	.244	.243	.247	.040	.175	.096	.023	.039	.193
	(.004)	(.005)	(.005)	(.063)	(.017)	(.037)	(.048)	(.049)	(.021)
1984	.154	.149	.149	.189	.184	.187	.188	.188	.182
	(.001)	(.004)	(.004)	(.002)	(.001)	(.002)	(.001)	(.001)	(.002)
1985	.182	.182	.182	.181	.183	.183	.181	.183	.183
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1986	.187	.187	.187	.189	.189	.190	.190	.190	.192
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1987	.181	.182	.181	.176	.176	.177	.176	.177	.177
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1988	.180	.180	.181	.181	.181	.181	.181	.182	.183
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1989	.171	.172	.172	.168	.170	.169	.169	.170	.171
	(.008)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1990	.012	.021	.005	.358	.303	.375	.349	.395	.363
	(.002)	(.007)	(.008)	(.012)	(.008)	(.015)	(.012)	(.016)	(.013)
1991	.182	.184	.180	.153	.167	.156	.161	.157	.163
	(.001)	(.002)	(.002)	(.002)	(.001)	(.002)	(.001)	(.002)	(.001)
1992	.162	.162	.162	.159	.155	.159	.157	.160	.161
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1993	.207	.207	.207	.209	.209	.209	.210	.209	.211
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1994	.237	.236	.237	.250	.250	.251	.252	.253	.254
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1995	.193	.195	.194	.177	.179	.177	.177	.178	.180
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1996	.177	.177	.177	.176	.178	.177	.177	.177	.178
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1997	.167	.167	.167	.162	.162	.162	.162	.162	.164
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1998	.137	.138	.138	.134	.137	.135	.135	.136	.138
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
1999	.152	.152	.152	.155	.157	.157	.156	.157	.158
	(.001)	(.001)	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.001)
2000	.159	.159	.159	.159	.159	.159	.159	.159	.160
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
2001	.158	.158	.158	.159	.159	.160	.159	.160	.161
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
2002	.153	.153	.153	.146	.146	.146	.146	.147	.149
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
2003	.168	.167	.168	.178	.178	.179	.179	.180	.181
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
2004	.147	.148	.148	.133	.133	.134	.133	.134	.135
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
2005	.128	.128	.128	.130	.132	.130	.131	.131	.133
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
2006	.123	.124	.123	.124	.124	.124	.125	.125	.127
	(.001)	(.001)	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
2007	.117	.117	.117	.115	.116	.116	.115	.117	.118
	(.003)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)

Table S.vi: Raw covariance matrix by number of non-missing periods

Sample periods	$Var(\eta_1)$	$Cov(\eta_1,\eta_2)$	$Cov(\eta_1,\eta_3)$	$Var(\eta_2)$	$Cov(\eta_2,\eta_3)$	$Var(\eta_3)$
4	4.7	0.41	-5.5	0.16	-1.1	9.4
	(889)	(68)	(956)	(5.3)	(74)	(1029)
5	23	0.8	-20	0.1	-1.1	19
	(101)	(7.6)	(108)	(0.59)	(8.2)	(115)
6	21	1.5	-22	0.15	-1.8	24
	(45)	(3.3)	(48)	(0.25)	(3.6)	(51)
7	18	1.3	-18	0.13	-1.5	20
	(36)	(2.8)	(39)	(0.21)	(3)	(42)
8	20	1.6	-21	0.15	-1.8	24
	(27)	(2.3)	(30)	(0.19)	(2.5)	(33)
9	16	1.3	-17	0.12	-1.5	19
	(14)	(1.2)	(16)	(0.11)	(1.4)	(18)
10	15	1.4	-17	0.14	-1.6	20
	(9.8)	(0.88)	(11)	(0.081)	(1)	(13)
11	9.4	0.86	-11	0.087	-1	12
	(5.1)	(0.47)	(5.9)	(0.044)	(0.54)	(6.8)
12	5.6	0.52	-6.3	0.055	-0.63	7.4
	(2.6)	(0.25)	(3)	(0.025)	(0.29)	(3.5)
13	3.9	0.37	-4.4	0.039	-0.44	5.1
	(1.3)	(0.13)	(1.5)	(0.013)	(0.15)	(1.8)
14	2.9	0.33	-3.6	0.041	-0.43	4.5
	(0.86)	(0.1)	(1.1)	(0.012)	(0.13)	(1.4)
15	2.4	0.25	-2.8	0.03	-0.32	3.4
	(0.72)	(0.075)	(0.86)	(0.0081)	(0.091)	(1)
16	0.93	0.1	-1.1	0.016	-0.15	1.5
	(0.24)	(0.026)	(0.29)	(0.0032)	(0.033)	(0.36)
17	1.1	0.12	-1.3	0.019	-0.18	1.7
	(0.26)	(0.032)	(0.33)	(0.0043)	(0.043)	(0.44)
18	0.75	0.089	-0.9	0.014	-0.13	1.2
	(0.17)	(0.021)	(0.22)	(0.0029)	(0.029)	(0.3)
19	0.53	0.064	-0.61	0.012	-0.1	0.91
	(0.085)	(0.012)	(0.11)	(0.0018)	(0.017)	(0.16)
20	0.33	0.04	-0.37	0.0084	-0.069	0.59
	(0.04)	(0.0056)	(0.052)	(0.001)	(0.0088)	(0.077)
21	0.22	0.017	-0.17	0.0051	-0.039	0.32
	(0.029)	(0.0032)	(0.03)	(0.00062)	(0.005)	(0.044)
22	0.17	0.018	-0.16	0.0048	-0.037	0.3
	(0.019)	(0.0031)	(0.026)	(0.00068)	(0.0053)	(0.042)
23	0.18	0.014	-0.13	0.0047	-0.033	0.25
	(0.017)	(0.0024)	(0.019)	(0.00051)	(0.0038)	(0.03)
24	0.16	0.015	-0.13	0.0049	-0.036	0.28
	(0.017)	(0.0027)	(0.022)	(0.00061)	(0.0047)	(0.037)
25	0.14	0.011	-0.098	0.0035	-0.025	0.19
	(0.014)	(0.0023)	(0.019)	(0.00042)	(0.0033)	(0.027)
26	0.1	0.0071	-0.064	0.003	-0.02	0.15
	(0.0072)	(0.00093)	(0.0071)	(0.00023)	(0.0016)	(0.012)
27	0.082	0.0048	-0.043	0.0021	-0.014	0.099
	(0.0046)	(0.00063)	(0.0047)	(0.00015)	(0.001)	(0.0072)
28	0.071	0.0037	-0.033	0.0018	-0.012	0.081
	(0.0049)	(0.00067)	(0.0047)	(0.00016)	(0.0011)	(0.0073)
Complete sample	2.6	0.22	-2.8	0.024	-0.27	3.3
	(3.4)	(0.27)	(3.7)	(0.022)	(0.29)	(4)

Note: Heteroskedastic consistent standard errors in parentheses.

Table S.vii: Bias corrected covariance matrix by number of sampling periods

Sample periods	$Var(\eta_1)$	$Cov(\eta_1, \eta_2)$	$Cov(\eta_1, \eta_3)$	$Var(\eta_2)$	$Cov(\eta_2, \eta_3)$	$Var(\eta_3)$
4	-22	-1.7	23	-0.0065	1.2	-22
	(744)	(56)	(796)	(4.3)	(60)	(852)
5	5.9	-0.54	-1.5	-0.0091	0.42	-0.98
	(69)	(5.2)	(73)	(0.4)	(5.6)	(78)
6	6.6	0.48	-6.9	0.064	-0.62	7.7
	(23)	(1.7)	(24)	(0.13)	(1.8)	(26)
7	3.9	0.14	-3.3	0.029	-0.23	3.3
	(23)	(1.7)	(24)	(0.13)	(1.8)	(26)
8	4.4	0.29	-4.4	0.034	-0.36	4.8
	(16)	(1.3)	(18)	(0.11)	(1.5)	(19)
9	4.1	0.29	-4.3	0.027	-0.33	4.6
	(8)	(0.69)	(8.9)	(0.061)	(0.77)	(9.9)
10	6.2	0.54	-6.9	0.054	-0.64	8
	(6.4)	(0.57)	(7.3)	(0.052)	(0.65)	(8.2)
11	4.2	0.35	-4.5	0.035	-0.41	5.2
	(3.7)	(0.33)	(4.2)	(0.03)	(0.38)	(4.7)
12	3.3	0.28	-3.6	0.027	-0.32	4
	(2.3)	(0.21)	(2.6)	(0.02)	(0.24)	(3)
13	1.6	0.11	-1.5	0.0089	-0.11	1.5
	(1.5)	(0.15)	(1.8)	(0.016)	(0.18)	(2.1)
14	1.7	0.18	-2	0.023	-0.24	2.5
	(0.76)	(0.088)	(0.95)	(0.011)	(0.11)	(1.2)
15	1.7	0.16	-1.9	0.019	-0.2	2.3
	(0.65)	(0.065)	(0.76)	(0.0068)	(0.077)	(0.89)
16	0.73	0.076	-0.81	0.011	-0.11	1.1
	(0.25)	(0.026)	(0.29)	(0.0031)	(0.033)	(0.35)
17	0.89	0.1	-1	0.015	-0.14	1.4
	(0.26)	(0.032)	(0.33)	(0.0042)	(0.042)	(0.43)
18	0.62	0.071	-0.72	0.01	-0.097	0.96
	(0.17)	(0.021)	(0.22)	(0.0028)	(0.028)	(0.29)
19	0.45	0.052	-0.49	0.0089	-0.077	0.7
	(0.085)	(0.012)	(0.11)	(0.0018)	(0.017)	(0.16)
20	0.27	0.032	-0.29	0.0062	-0.051	0.43
	(0.04)	(0.0057)	(0.052)	(0.001)	(0.0087)	(0.077)
21	0.18	0.012	-0.13	0.0034	-0.026	0.22
	(0.029)	(0.0032)	(0.03)	(0.00061)	(0.0049)	(0.043)
22	0.15	0.015	-0.13	0.0035	-0.027	0.22
	(0.019)	(0.0031)	(0.026)	(0.00067)	(0.0053)	(0.042)
23	0.16	0.012	-0.11	0.0035	-0.024	0.19
	(0.017)	(0.0024)	(0.02)	(5e-04)	(0.0038)	(0.03)
24	0.14	0.014	-0.12	0.0041	-0.03	0.23
	(0.017)	(0.0027)	(0.022)	(0.00061)	(0.0047)	(0.037)
25	0.13	0.01	-0.089	0.0028	-0.02	0.16
	(0.014)	(0.0023)	(0.019)	(0.00041)	(0.0033)	(0.027)
26	0.097	0.0066	-0.059	0.0025	-0.017	0.12
	(0.0072)	(0.00093)	(0.0072)	(0.00023)	(0.0016)	(0.012)
27	0.077	0.0046	-0.04	0.0017	-0.011	0.079
	(0.0046)	(0.00064)	(0.0047)	(0.00015)	(0.001)	(0.0074)
28	0.067	0.0036	-0.031	0.0015	-0.0097	0.067
	(0.0049)	(0.00067)	(0.0047)	(0.00016)	(0.0011)	(0.0074)
Complete sample	0.96	0.07	-0.95	0.0096	-0.094	1.1
	(2)	(0.16)	(2.1)	(0.013)	(0.17)	(2.3)
Random effects	0.093	0.0059	-0.05	0.0015	-0.0093	0.066
	(0.0036)	(0.00051)	(0.004)	(0.00011)	(0.00079)	(0.0059)

Note: Heteroskedastic consistent standard errors in parentheses.

Table S.viii: Counterfactual variances by skills

Years	$\eta_1 = \bar{\eta}_{1g}$ $\eta_2 = \bar{\eta}_{2g}$ $\eta_3 = \bar{\eta}_{3g}$	Heterogeneity in: Levels	Growth	Curvature	Growth and Curvature	All
Log-wage 1977	0.279	0.284	0.28	0.281	0.283	0.284
	(0.00691)	(0.00691)	(0.00691)	(0.00691)	(0.00703)	(0.00714)
Log-wage 1982	0.00486	0.0373	0.00809	0.0124	0.0227	0.0394
	(0.00771)	(0.00771)	(0.00771)	(0.00771)	(0.00781)	(0.00791)
Log-wage 1987	0.00569	0.0428	0.0173	0.0192	0.0195	0.0468
	(0.00945)	(0.00945)	(0.00945)	(0.00945)	(0.00953)	(0.00961)
Log-wage 1992	0.00641	0.0439	0.0324	0.029	0.0328	0.0639
	(0.0127)	(0.0127)	(0.0127)	(0.0127)	(0.0128)	(0.0128)
Log-wage 1997	0.00685	0.0444	0.0531	0.0446	0.0555	0.0808
	(0.0181)	(0.0181)	(0.0181)	(0.0181)	(0.0181)	(0.0182)
Log-wage 2002	0.00684	0.0444	0.0791	0.07	0.08	0.091
	(0.0263)	(0.0263)	(0.0263)	(0.0263)	(0.0264)	(0.0264)
Log-wage 2007	0.00627	0.0438	0.11	0.112	0.113	0.113
	(0.0389)	(0.0389)	(0.0389)	(0.0389)	(0.0389)	(0.039)
Long-run	0.013	0.0421	0.0283	0.0299	0.0299	0.0458
	(0.0102)	(0.0102)	(0.0102)	(0.0102)	(0.0103)	(0.0104)

Note: Low skills. Only observations with more than 21 periods. 2942 observations.

Years	$\eta_1 = \bar{\eta}_{1g}$ $\eta_2 = \bar{\eta}_{2g}$ $\eta_3 = \bar{\eta}_{3g}$	Heterogeneity in: Levels	Growth	Curvature	Growth and Curvature	All
Log-wage 1977	0.00935	0.0172	0.0103	0.0116	0.0144	0.0176
	(0.0161)	(0.0161)	(0.0161)	(0.0161)	(0.0162)	(0.0164)
Log-wage 1982	0.014	0.0716	0.0189	0.0256	0.0391	0.0728
	(0.0178)	(0.0178)	(0.0178)	(0.0178)	(0.0179)	(0.0181)
Log-wage 1987	0.0173	0.0832	0.0349	0.0381	0.0384	0.0913
	(0.0198)	(0.0198)	(0.0198)	(0.0198)	(0.0199)	(0.0201)
Log-wage 1992	0.0207	0.0872	0.0601	0.0555	0.0602	0.121
	(0.0222)	(0.0222)	(0.0222)	(0.0222)	(0.0223)	(0.0224)
Log-wage 1997	0.0239	0.0905	0.0939	0.0821	0.0959	0.149
	(0.0249)	(0.0249)	(0.0249)	(0.0249)	(0.025)	(0.0251)
Log-wage 2002	0.0268	0.0933	0.136	0.124	0.137	0.167
	(0.0282)	(0.0282)	(0.0282)	(0.0282)	(0.0282)	(0.0284)
Log-wage 2007	0.029	0.0956	0.187	0.191	0.193	0.197
	(0.0322)	(0.0323)	(0.0322)	(0.0322)	(0.0323)	(0.0324)
Long-run value	0.017	0.0686	0.0403	0.043	0.043	0.0766
	(0.02)	(0.02)	(0.02)	(0.02)	(0.0201)	(0.0203)

Note: Medium skills. Only observations with more than 21 periods. 1433 observations.

Figure S.i: Change over time in mean and variance of log earnings for cohorts 1977-2000

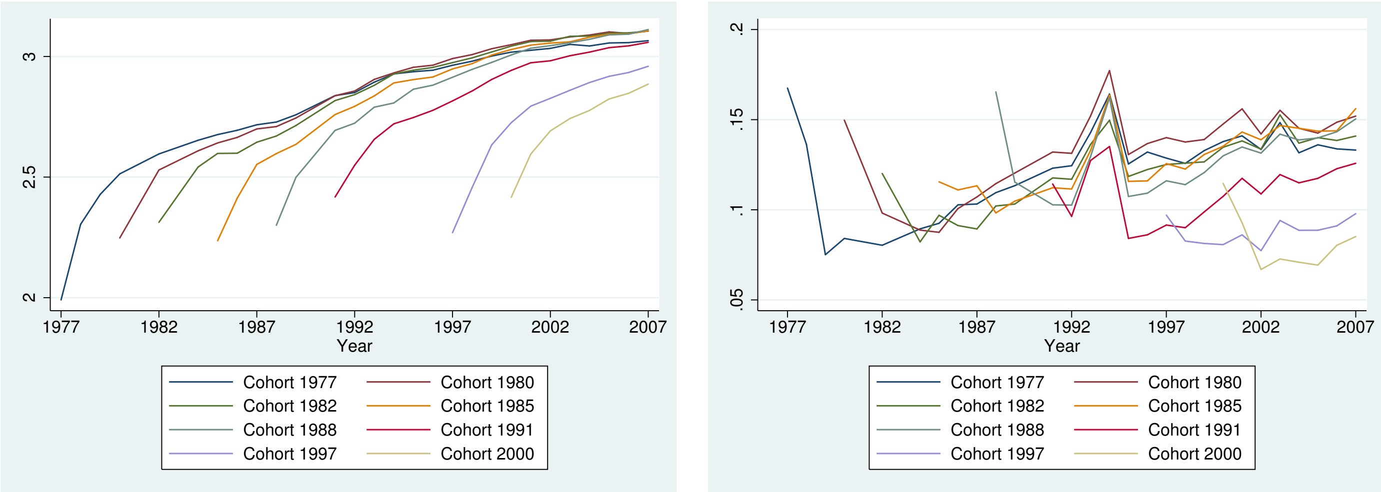


Figure S.ii: First order autocorrelation relative to potential experience for 1977, 1987 and 1997 entry cohorts

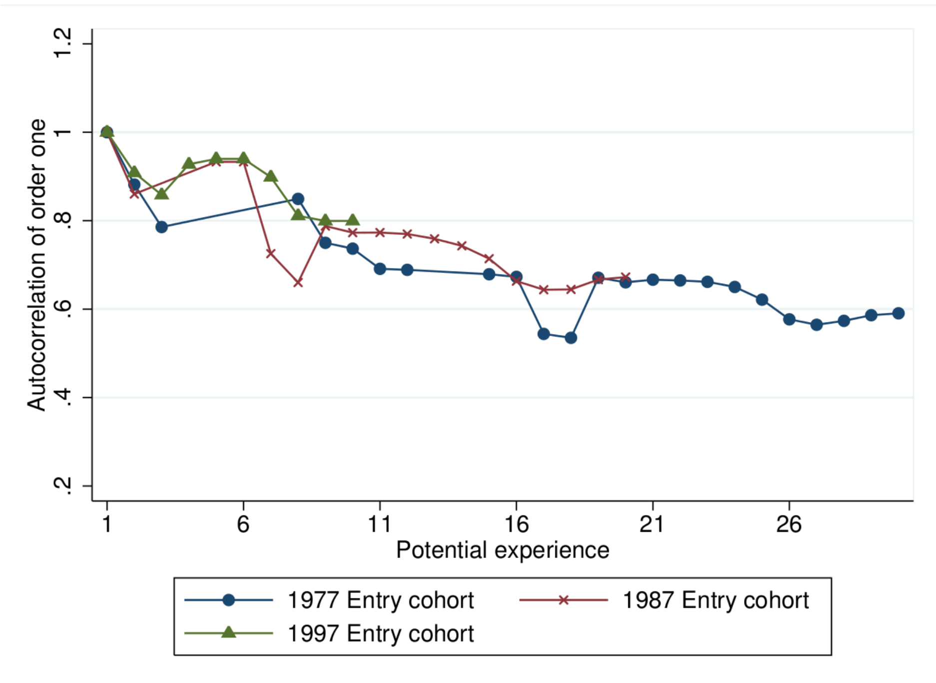


Figure S.iii: Unconstrained estimates: variance of residuals v_{it} by age and skill group