Human capital investments and the life cycle variance of earnings

Thierry Magnac[*], Nicolas Pistolesi[†], Sébastien Roux[‡]

Preliminary, All comments welcome

29th July 2011

Abstract

We propose a model of on-the-job human capital investments in which individuals differ in their initial human capital, their rate of return, their costs of human capital investments and their terminal values of human capital at retirement. We derive a tractable reduced form Mincerian model of log wage profiles along the life cycle which is written as a function of three individual specific factors. The model is estimated by pseudo maximum likelihood using panel data for a single cohort of French wage earners observed over a long span of 30 years. This structure allows us to compute counterfactual profiles in which returns and terminal values are modified and we show how wage inequality is affected by these changes over the life-cycle.

JEL Codes: J22, J24, J31

Keywords: life cycle human capital investment, on-the-job training, eraning dynamics, dynamic panel data

[*]Toulouse School of Economics (Université Toulouse Capitole, GREMAQ, IDEI)

[†]Toulouse School of Economics (Université Toulouse Capitole & GREMAQ)

[‡]CREST INSEE, Paris

1 Introduction

Since the seminal work by Lillard and Willis (1978) on the estimation of reduced form earnings dynamics an extensive literature has emerged. While a very large set of empirical studies estimating ARMA models on earnings residuals have been conducted, the literature has not reached any consensus on a unique specification of the earnings process (see Meghir and Pistaferri, 2010 for a survey). Most authors admit that a mixed process with individual-specific effects, autoregressive and moving average components seems necessary to fit the longitudinal change in earnings dispersion that is commonly observed although they do not agree on the description of earnings growth. Several papers have considered a beauty contest between a specification in which earnings growth is random and a specification in which earnings growth is governed by a linear trend multiplied by a fixed individual effect (see Baker, 1997 and Guvenen, 2009 for instance). In most of these papers the theoretical background for such reduced form models are nevertheless unclear while additional structure might be useful so as to distinguish different reduced forms.

In this paper we develop a simple theoretical model of on-the-job human capital investments accomodating substantial unobserved heterogeneity and derive a tractable and convenient reduced form for earnings dynamics. Following Mincer (1974) *Accounting identity model* as presented by Heckman Lochner and Todd (2006), we explain differences in earnings trajectories by heterogenous choices derived from heterogeneous individual characteristics. What interests us is the second part only of the research by Mincer that is the post schooling wage growth as taken from the Ben Porath (1967) model used to explain the shape in the mean earnings profile: earnings increase at the beginning of the working career then decrease slightly before retirement. It is commonly interpreted as reflecting individuals economic decisions to acquire skills mostly at the beginning of their career whereas they stop investing during the final years because their horizon of investment is shortened.

There are two other interesting predictions of the human capital setting which are tested (Rubinstein and Weiss, 2006). First, the variance of earnings should have an inverted U-shape along the life-cycle. Comparing earnings trajectories between large-returns investors having a steep earnings profile and low-returns individuals experiencing a flatter profile provides indications on the way earnings dispersion increases over time. Second, the autocorrelation of earnings along the life cycle should be negative. Because investments in human capital are more intensive at the beginning of the life cycle for the high return investors, there tends to be a negative correlation between earning growth and level in cross section at the beginning of the life cycle and this correlation fades out with time to become positive. A simple endogeneous search model would predict the contrary. The better paid tend to search less because it is more costly for them and the level and earnings growth tend to be negatively associated all along the life-cycle.

We start from the main intuition of the post schooling wage growth model describing differences in trajectories by, on the one hand, heterogenous characteristics and on the other, heterogenous choices of investment. Instead of focusing on the mean we investigate the implications of the theory for the covariance of earnings along the life-cycle profile. We consider as given school investments and we treat them as an additional source of individual heterogeneity. We are allowing for a lot of heterogeneity as Alvarez, Browning and Erjnaes (2010) do not only because it has been recognized that unobserved heterogeneity would bias the rates of return but also because the amount of unobserved heterogeneity conditions the diagnostics about life-cycle inequality. We are building up as well on what has been developed times ago by Heckman (see Heckman, Lochner and Todd, 2006, for a survey) and Card (for instance in the Econometrica lecture in 2001) for schooling investments in human capital.

In this paper, we specify a model in which individuals differ in three main respects. Firstly, individuals have different initial human capital levels when they enter the labor market. Secondly, individuals differ in their returns to skill investments. It can be interpreted as individuals being more of less productive in transforming invested time in productive skills. As in Mincer's original model, heterogeneity in rates of return to investment play a crucial role explaining why individual earnings trajectories differ. Our model also assumes that the marginal cost of producing skills is heterogenous within the population. Finally, we allow the terminal value of human capital to vary across individuals and infer from these values the implicit horizon of investment that agents condier from the curvature of the earnings profile. This follows a suggestion by Lillard and Reville (1999) insisting on this crucial aspect of earnings growth. As a consequence, since most of these characteristics are not observable for the econometrician, this translates into an error component structure of the earnings equation, that is highly persistent and whose variance increases over time.

We treat search and job mobility as frictions under the form of exogenous shocks. Indeed what Postel-Vinay and Turon (2010) nicely explicits in their presentation is that the dynamics of the earnings process is partly controled by two other processes which are individual productivity in the current match and outside offers that the individual receives while on the job. Three things can happen: either the earnings remains within the two bounds defined by these processes;

or the earnings is equal to the productivity process because adverse shocks on that process made employee and employer renegociate the wage contract; or finally, the wage is 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. We do not impose these structural constraints in this paper and we treat them as an element of idiosyncratic shocks.

We estimate the model on a very long panel for a single cohort of male French wage earners observed from 1977 to 2007. DADS data is an administrative dataset collecting earnings in the private sector for social security records and that has many advantages for our purpose. First, it includes enough observations so that we can study a single cohort of individuals who enter the labor market simultaneously and face the same economic environment over their life-cycle, contrary to most studies of earnings dynamics that must pull different cohorts to collect samples large enough. Secondly, as the data come from social security records, we expect fewer measurement errors than in usual surveys or other administrative data. Finally, the DADS data are long and homogeneous enough to study the dynamics of earnings over a long period of time. It has also some shortcomings as well since first, few other individual characteristics than age and broad skill groupings. Second, the panel data is incomplete at the periods during which individuals leave the private sector because of unemployment, self-employment, non-participation or because they are working in the public sector. This explains why we choose to use male earning data only.

We first estimate the model by random effect maximum likelihood (Alvarez and Arellano, 2004) and derive the fixed effect estimates. Using the latter estimates, we evaluate structural restrictions and compute estimates of the structural unobserved factors. We can construct counterfactuals measuring the impact of changes in those structural estimates. We find that surprisingly an increase in post retirement returns to human capital decrease the variances of earnings in late years although increases in pre retirement returns unambiguosly increase this variance.

In the next section we describe the model of human capital accumulation. In section 3 we consider the econometric framework and offer a literature review of empirical earning equations and the way dynamic panel data methods are used to estimate them. Data are described in section 4. Section 5 presents the results. After a discussion of a possible extension, a final section concludes.

2 The Model

We present a model of human capital investment in which agents face individual specific costs, individual specific rates of return and individual specific terminal values. We characterize the optimal sequence of human capital investments over the life cycle and we derive the reduced form of life cycle earnings equation.We then analyze the transformation between parameters of the reduced and structural forms.

2.1 The set up

As in Ben Porath (1967) and Mincer (1974) we suppose that the retirement date is fixed at $t = R$. The model starts when individuals enter the labor market normalized at time $t = 0$. The entry decision in the labour market is endogenous and depends on previous human capital accumulation. We take these initial conditions as given and depending on a unobserved variable, the human capital stock at entry, which is potentially correlated with all shocks affecting the life-cycle dynamics of earnings.

From period 0 to R agents can acquire human capital through part-time on-the-job training. Human capital is supposed to be single-dimensional and potential individuals earnings, $y_i^P(t)$ are given by individual human capital times an individual specific rental rate that is $y_i^P(t) = \exp(\delta_i(t))H_i(t)$. Individuals face uncertainty through the variability of the rental rate of human capital $\delta_i(t)$ which is mainly affected by aggregate shocks but also by individual ones if there are some frictions in the labor market. Firms might temporarily value human capital differently than the market in order to attract, retain or discourage specific individuals. The rental rate is supposed to follow a stochastic process and $\delta_i(t)$ is fully revealed at period t to the agent. We do not provide a market analysis of the wage equilibrium process and take it as a given (in terms of its distribution).

By deducting chosen human capital investments, actual individual earnings are assumed to be given by:

$$y_i(t) = \exp(\delta_i(t))H_i(t)\exp(-\tau_i(t))$$

where $1 - \exp(-\tau_i(t))$ can be interpreted as the fraction of working time devoted to investing in human capital as in the original Ben Porath formulation. It might also be interpreted as the level of effort put in the acquisition of human capital at the cost of losing some potential earnings. With no loss of generality, we denote $\tau_i(t)$ at time t the level of investment in human capital instead of working with investment time. Note in particular that if $\tau_i(t) = 0$, actual earnings are equal to potential earnings.

Because of these investments, individuals accumulate human capital in a way that is described by the following equation

$$H_i(t+1) = H_i(t)\exp[\rho_i \tau_i(t) - \lambda_i(t)] \tag{1}$$

where $H_i(t)$ is the stock of human capital, ρ_i an individual specific rate of return of human capital investments and $\lambda_i(t)$ is the depreciation of human capital in period t. This latter component embeds innovations at the economy level as these innovations depreciate previous vintages of human capital or embeds individual-specific shocks. The latter can be negative because of unemployment periods or of layoffs followed by mobility across sectors. These shocks can also be positive when certain components of human capital acquire more value or because of voluntary moves across firms or sectors. As $\delta_i(t)$, the variable $\lambda_i(t)$ is supposed to be revealed at period t to the agent and is uncertain before. We also take the stochastic process $\lambda_i(t)$ as a given.

Current-period utility is assumed to be equal to actual log earnings net of investment costs,

$$u_i(t) = \delta_i(t) + log\ H_i(t) - \gamma_i\left(\tau_i(t) + c_i\frac{\tau_i(t)^2}{2}\right)$$

where γ_i and c_i represent between-individual differences in the cost of human capital accumulation in utility terms and the cost is quadratic. There are two aspects that depart from a standard formulation. The simplest objective function would be a function of actual earnings or their logarithm only:

$$\delta_i(t) + log\ H_i(t) - \tau_i(t). \tag{2}$$

We neither assume at this stage that $\gamma_i = 1$ nor that the objective function is linear ($c_i = 0$). It adds richness to the setting and it fits well with the interpretation of $\tau_i(t)$ in terms of effort exerted for human capital investments and not only time as in the simple specification. Nonetheless, the costs of investments do not depend on the level of human capital $H_i(t)$. Section 6 proposes a convenient generalization of our setting to the case of increasing costs of investment with the level of human capital. It comes at the price of having additional factors in the econometric model.

As individuals maximize their future discounted utility stream, their decision program is given by the following Bellman equation:

$$V_t(H_i(t), \tau_i(t)) = \delta_i(t) + \log H_i(t) - \gamma_i\left(\tau_i(t) + c_i\frac{\tau_i(t)^2}{2}\right) + \beta E_t\left[W_{t+1}(H_i(t+1))\right] \tag{3}$$

where:

$$W_{t+1}(H_i(t+1)) = V_{t+1}(H_i(t+1), \tau_i^*(t+1)) = \max_{\tau_i(t+1)} V_{t+1}(H_i(t+1), \tau_i(t+1)),$$

and where β is the discount factor. We do not tackle the case where the discount factor is heterogenous between agents.

The dynamic program is completed by the returns to human capital after working life. We assume directly that the value function or the discounted value of utility stream from date R onwards is given by:

$$W_R(H_i(R)) = \delta^* + \kappa_i \log H_i(R), \tag{4}$$

where κ is the capitalized value of one euro over remaining life. It includes the heterogenous survival probabilities from R onwards that agents anticipate and we restrict our setting so that:

$$\kappa_i < \frac{1}{1-\beta}.$$

This condition is justified by the fact that the discount rate after retirement is smaller than β because the survival probability is smaller after retirement.

2.2 The life-cycle profile of investments

We first consider the case where human capital investments are always positive over the life-cycle and the profile of investments is summarized in:

Proposition 1 *Suppose that :*

$$\beta\rho_i\kappa_i > \gamma_i, \tag{5}$$

then:

$$\tau_i(t) = \frac{1}{c_i}\left\{\frac{\rho_i}{\gamma_i}\left[\frac{\beta}{1-\beta} + \beta^{R-t}(\kappa_i - \frac{1}{1-\beta})\right] - 1\right\} > 0, \quad \forall t < R \tag{6}$$

Proof. The first order condition of the maximization problem for $t < R$ is

$$-\gamma_i\left[1 + c_i\tau_i(t)\right] + \beta\rho_i H_i(t+1) E_t\left[\frac{\partial W_{t+1}}{\partial H_i(t+1)}\right] = 0. \tag{7}$$

The marginal value of human capital is the derivative of the Bellman equation so that by the envelope theorem:

$$\frac{\partial W_t}{\partial H_i(t)} = \frac{1}{H_i(t)} + \beta E_t\left[\frac{\partial W_{t+1}}{\partial H_i(t+1)}\right]\frac{H_i(t+1)}{H_i(t)} \tag{8}$$

For $t = R$, condition (8) writes more simply as:

$$\frac{\partial W_R}{\partial H_i(R)} = \frac{\kappa_i}{H_i(R)} \Longrightarrow H_i(R)\frac{\partial W_R}{\partial H_i(R)} = \kappa_i,$$

so that, by backward induction, we obtain:

$$H_i(R-1)\frac{\partial W_{R-1}}{\partial H_i(R-1)} = 1 + \beta\kappa_i, H_i(R-2)\frac{\partial W_{R-2}}{\partial H_i(R-2)} = 1 + \beta(1 + \beta\kappa_i)$$

and so on. This yields:

$$H_i(t+1)\frac{\partial W_{t+1}}{\partial H_i(t+1)} = \frac{1 - \beta^{R-(t+1)}}{1-\beta} + \beta^{R-(t+1)}\kappa_i.$$

Replacing in equation (7) yields:

$$\gamma_i\left(1 + c_i\tau_i(t)\right) = \beta\rho_i\left[\frac{1}{1-\beta} + \beta^{R-(t+1)}(\kappa_i - \frac{1}{1-\beta})\right] = \rho_i\left[\frac{\beta}{1-\beta} + \beta^{R-t}(\kappa_i - \frac{1}{1-\beta})\right],$$

and equation (6) follows. Furthermore, as the second term in (7) is constant, the second order condition is satisfied if and only if $\gamma_i c_i > 0$.

Furthermore and given that $c_i > 0$, the condition that investments are always positive yields:

$$\frac{\rho_i}{\gamma_i}\left[\frac{\beta}{1-\beta} + \beta^{R-t}(\kappa_i - \frac{1}{1-\beta})\right] - 1 \geq 0. \quad \forall t < R$$

As $\kappa_i - \frac{1}{1-\beta} < 0$ and $\beta < 1$, $\tau_i(t)$ is decreasing in t because of the term β^{-t} and the RHS attains its minimum at $t = R-1$. This yields condition (5) since:

$$\frac{\rho_i}{\gamma_i}\left[\frac{\beta}{1-\beta} + \beta(\kappa_i - \frac{1}{1-\beta})\right] - 1 \geq 0 \Longleftrightarrow \frac{\rho_i}{\gamma_i} \geq \frac{1}{\beta\kappa_i}.$$

■

It is now easy to analyze cases in which investments in human capital stop before the penultimate period. Indeed, the level of investment $\tau_i(t)$ is deterministic and decreasing in t because of the term in β^{-t}. As expected, it is always better to invest earlier than later given that the horizon over which investments are valuable is becoming smaller and smaller. No investments in period t, $\tau_i(t) = 0$, imply that no investments would take place later on, $\tau_i(t') = 0$, $\forall t' \geq t$. In consequence, we can proceed backwards and analyze the conditions under which human capital investments stop at period t.

Proposition 2 *The investment sequence is such that for any* $t \in [1, R-1]$

$$\tau_i(t') = 0, \forall t' \geq t+1, \tau_i(t) > 0$$

if and only if:

$$\frac{1}{\kappa_{it}} < \beta\frac{\rho_i}{\gamma_i} \leq \frac{1}{\kappa_{i,t+1}}, \tag{9}$$

where $\frac{1}{\kappa_R} = +\infty, \kappa_{i,R-1} = \kappa_i$ *and* $\kappa_{it} = 1 + \beta\kappa_{i,t+1}$. *Additionally, when condition (9) is satisfied we can replace period* R *by period* $S_i = t+1$ *in equation (6) to derive human capital investments before and including period* $S_i - 1$. *Period* S_i *is the optimal stopping period for human capital investments.*

Proof. Condition (9) is consistent since $\kappa_{it} = 1+\beta\kappa_{i,t+1} > \kappa_{i,t+1} \Leftrightarrow \kappa_{i,t+1} < \frac{1}{1-\beta} \Leftrightarrow \kappa_{i,t+2} < \frac{1}{1-\beta}$ and by repetition $\kappa_{i,R-1} = \kappa_i < \frac{1}{1-\beta}$. We now proceed by backward induction. The content of Proposition 2 is true at period $t = R-1$ because of Proposition 1 and condition (5). Suppose that this condition is true at periods $t' \geq t+1$.

First, assume that $\tau_i(t') = 0, \forall t' \geq t+1$ so that the condition $\tau_i(t') > 0$ is violated for any $t' \geq t+1$ and therefore $\beta\frac{\rho_i}{\gamma_i} \leq 1/\kappa_{i,t+1}$. Conversely, if $\beta\frac{\rho_i}{\gamma_i} \leq 1/\kappa_{i,t+1}$ then $\tau_i(t') = 0, \forall t' \geq t+1$ because Proposition 2 is true for $t' \geq t+1$. Furthermore, the condition $\tau_i(t') = 0$ implies simple forms for the Bellman equation (3):

$$W_t(H_i(t')) = \delta_i(t') + \log H_i(t') + \beta E_{t'} W_{t'+1}(H_i(t'+1)),$$

and the accumulation equation (1):

$$\log H_i(t'+1) = \log H_i(t') - \lambda_i(t').$$

Using equation (4) where we set $\kappa_{iR-1} = \kappa_i$ and the linearity of the previous two equations lead to the condition derived by induction again:

$$W_{t'}(H_i(t')) = \delta^*(t') + \kappa_{i,t'-1} \log H_i(t'). \tag{10}$$

for any $t' \geq t+1$ and where $\kappa_{it} = 1 + \beta\kappa_{i,t+1}$.

Second in addition to the previous statements, assume that $\tau_i(t) > 0$. Proposition 1 can be recast in a set-up where the last period becomes $S_i = t+1$ instead of R since there are no further human capital investments after this date and since the value function can be written as in equation (10) evaluated at $t' = t+1$. We can adapt equation (6) and obtain:

$$\tau_i(t) = \frac{1}{c_i}\left\{\frac{\rho_i}{\gamma_i}\left[\frac{\beta}{1-\beta} + \beta(\kappa_{it} - \frac{1}{1-\beta})\right] - 1\right\} > 0,$$

which is equivalent to $\beta\frac{\rho_i}{\gamma_i} > \frac{1}{\kappa_{it}}$.

Therefore the equivalence stated in the Proposition is true at period t. It is therefore true at any date until $t = s+1$. ∎

It is easy to prove that the sequence κ_{it} of the previous proposition is:

Lemma 3

$$\kappa_{it} = \frac{1}{1-\beta} + \beta^{R-t-1}(\kappa_i - \frac{1}{1-\beta}) \tag{11}$$

Proof. By induction. It is true when $t = R-1$ since $\kappa_{i,R-1} = \kappa_i$. Assume that it is true at $t+1$ and prove it at t. ■

We can then summarize the two propositions into the following:

Corollary 4 *Suppose that there exists $S = 1, ., R$ such that:*

$$\frac{1}{\kappa_{i,S-1}} < \beta \frac{\rho_i}{\gamma_i} \leq \frac{1}{\kappa_{iS}},$$

then:

$$\tau_i(t) = \frac{1}{c_i}\left\{\frac{\rho_i}{\gamma_i}\left[\frac{\beta}{1-\beta} + \beta^{R-t}(\kappa_i - \frac{1}{1-\beta})\right] - 1\right\} > 0, \quad \forall t < S$$

Proof. Form Proposition 2, we know that human capital investments stop the period before S. We can then use Proposition 1 to derive from equation (6) adapted to period S that:

$$\tau_i(t) = \frac{1}{c_i}\left\{\frac{\rho_i}{\gamma_i}\left[\frac{\beta}{1-\beta} + \beta^{S-t}(\kappa_{i,S-1} - \frac{1}{1-\beta})\right] - 1\right\} > 0, \quad \forall t < S.$$

Replace $\kappa_{i,S-1}$ by its expression (11) to prove the Proposition. ■

This corollary proves that the profile of life-cycle investments is truncated at zero but there are no dynamic effects of the truncation. The profile remains similar even if investments stop. We can also use this corollary by default. In order to have some investments in human capital along the life cycle we shall have that $\beta\frac{\rho_i}{\gamma_i} > \frac{1}{\kappa_{iS}}$ and therefore $\beta\frac{\rho_i}{\gamma_i} > 1-\beta$ since $\kappa_{iS} < \frac{1}{1-\beta}$. We will assume that:

$$\frac{\rho_i}{\gamma_i} > \frac{1-\beta}{\beta}. \tag{12}$$

2.3 The Lifecycle Profile of Earnings

We start by deriving earnings equations when human capital investments remain positive over the life-cycle. First, the stock of human capital in period t depends on previous investment choices and past depreciation that is

$$H_i(t) = H_i(0)\exp\left[\sum_{l=0}^{t-1}\rho_i\tau_i(l) - \sum_{l=0}^{t-1}\lambda_i(l)\right] \text{ for } t > 0.$$

We can write the logarithm of actual earnings in period t as

$$log\ y_i(t) = \delta_i(t) + log\ H_i(0) + \sum_{l=0}^{t-1}\rho_i\tau_i(l) - \sum_{l=0}^{t-1}\lambda_i(l) - \tau_i(t). \tag{13}$$

First, it shows that returns to human capital $\delta_i(t)$ cannot be distinguished from depreciation effects $\sum_{l=0}^{t-1}\lambda_i(l)$ and we will therefore write that transitory earnings are equal to:

$$\delta_i^y(t) = \delta_i(t) - \sum_{l=0}^{t-1}\lambda_i(l).$$

Furthermore, inserting the reduced form for $\tau_i(\cdot)$ from equation (6) into the first sum we get :

$$\begin{aligned}
E\sum_{l=0}^{t-1}\rho_i\tau_i(l) &= \frac{\rho_i^2}{c_i\gamma_i}\sum_{l=0}^{t-1}\left[\frac{\beta}{1-\beta} + \beta^{R-l}(\kappa_i - \frac{1}{1-\beta})\right] - \frac{\rho_i}{c_i}(t), \\
&= \frac{\rho_i^2}{c_i\gamma_i}\frac{\beta}{1-\beta}(t) + \frac{\rho_i^2}{c_i\gamma_i}(\kappa_i - \frac{1}{1-\beta})\beta^R\sum_{l=0}^{t-1}\beta^{-l} - \frac{\rho_i}{c_i}(t) \\
&= \left(\frac{\rho_i^2}{c_i\gamma_i}\frac{\beta}{1-\beta} - \frac{\rho_i}{c_i}\right)(t) + \frac{\rho_i^2}{c_i\gamma_i}(\kappa_i - \frac{1}{1-\beta})\beta^R\frac{1-(1/\beta)^t}{1-1/\beta} \\
&= -\frac{\rho_i^2}{c_i\gamma_i}(\kappa_i - \frac{1}{1-\beta})\frac{\beta^{R+1}}{1-\beta} + \left(\frac{\rho_i^2}{c_i\gamma_i}\frac{\beta}{1-\beta} - \frac{\rho_i}{c_i}\right)(t) \\
&\quad + \frac{\rho_i^2}{c_i\gamma_i}(\kappa_i - \frac{1}{1-\beta})\frac{\beta^{R+1}}{1-\beta}\beta^{-t},
\end{aligned}$$

which writes as the sum of three factors whereas one factor is in levels, the second one is linear and the last one is geometric.

Finally, using equation (6):

$$\tau_i(t) = \frac{1}{c_i}\left(\frac{\rho_i}{\gamma_i}\frac{\beta}{1-\beta} - 1\right) + \frac{\rho_i}{c_i\gamma_i}\beta^R(\kappa_i - \frac{1}{1-\beta})\beta^{-t}$$

and rearranging expression (13) we have the following reduced form expression for log earnings

$$\log y_i(t) = \eta_{i1} + \eta_{i2}t + \eta_{i3}\beta^{-t} + \delta_i^y(t), \tag{14}$$

where:

$$\eta_{i1} = \log H_i(0) - \frac{\rho_i^2}{c_i\gamma_i}\left(\kappa_i - \frac{1}{1-\beta}\right)\frac{\beta^{R+1}}{1-\beta} - \frac{1}{c_i}\left(\frac{\rho_i}{\gamma_i}\frac{\beta}{1-\beta} - 1\right), \tag{15}$$

$$\eta_{i2} = \frac{\rho_i^2}{c_i\gamma_i}\frac{\beta}{1-\beta} - \frac{\rho_i}{c_i}, \tag{16}$$

$$\eta_{i3} = \frac{\rho_i^2}{c_i\gamma_i}\left(\kappa_i - \frac{1}{1-\beta}\right)\frac{\beta^{R+1}}{1-\beta} - \frac{\rho_i}{c_i\gamma_i}\beta^R(\kappa_i - \frac{1}{1-\beta}). \tag{17}$$

From that reduced form equation, it is clear that different permanent and transitory factors contribute to individual earnings trajectories. On the one hand, three types of permanent heterogeneities drive earnings dynamics. Firstly, differences in initial capital investment at school, $H_i(0)$, lead to permanent differences in log earnings. Secondly, between-individual differences in marginal return to investment, ρ_i, and marginal costs γ_i and c_i make earnings growth rate to be individual specific. Thirdly, the interaction between marginal return and between-individual differences in the cost of accumulation, $\rho_i/(c_i\gamma_i)$, leads earnings trajectories to differ in amplitude. We shall look below at the form of transitory earnings.

In the case in which human capital investments stop before the penultimate period, the previous results can be adapted by replacing period R by period S_i as developed in Proposition 2. This affects the definitions of the factors $(\eta_{i1}, \eta_{i2}, \eta_{i3})$ as derived in equations (15) to (17) although it does not affect the form of the earnings equation (14) before and including period $S_i - 1$. Nonetheless after period S_i, human capital investments are equal to zero and the earnings equation (14) is derived by using potential earnings and the accumulation equation:

$$log\ y_i(t) = \delta_i(t) + log\ H_i(t), log\ H_i(t+1) = log\ H_i(t) - \lambda_i(t), \forall t \geq S_i$$

so that we have:

$$log\ y_i(t+1) = log\ y_i(t) + \delta_i(t+1) - \delta_i(t) - \lambda_i(t).$$

Earnings growth becomes stochastic and is no longer determined by the terms η_{i2} and η_{i3}.

In the empirical section, we will assume that we never observe the second regime since at the end of the period of observation we are still far from any retirement period.

2.4 Reduced and Structural Forms

One of the purpose of the paper is not only to impose the three-factor structure on the reduced form but also to recover the distribution of unobserved heterogeneity components. Equation (15) which refers to unobserved hetogeneity in levels allows us to identify the level of initial human capital if the other individual specific terms are fixed. It thus imposes no constraint in the data. Equations (16) and (17) are more interesting and can be specifically rewritten as:

$$\eta_{i2} = \frac{\rho_i}{c_i}\left(\frac{\rho_i}{\gamma_i}\frac{\beta}{1-\beta} - 1\right), \tag{18}$$

$$\eta_{i3} = \frac{\rho_i}{c_i\gamma_i}\beta^R(\kappa_i - \frac{1}{1-\beta})\left(\rho_i\frac{\beta}{1-\beta} - 1\right). \tag{19}$$

This is a non linear system of two equations with four unknowns: ρ_i, γ_i, c_i and κ_i. The structural form cannot be identified as such and some restriction must be adopted. Because it has a natural interpretation in terms of costs, we are going to assume that $\gamma_i = 1$. The two equations above as well as the investment equation simplify to:

$$\begin{cases} \eta_{i2} = \frac{\rho_i}{c_i}\left(\rho_i\frac{\beta}{1-\beta} - 1\right), \\ \eta_{i3} = \frac{\rho_i}{c_i}\beta^R(\kappa_i - \frac{1}{1-\beta})\left(\rho_i\frac{\beta}{1-\beta} - 1\right), \\ \tau_i(t) = \frac{1}{c_i}\left\{\rho_i\left[\frac{\beta}{1-\beta} + \beta^{R-t}(\kappa_i - \frac{1}{1-\beta})\right] - 1\right\} \\ \quad = \frac{1}{c_i}\left(\rho_i\frac{\beta}{1-\beta} - 1\right) + \frac{\rho_i}{c_i}\beta^R(\kappa_i - \frac{1}{1-\beta})\beta^{-t}. \end{cases}$$

Structural restrictions are:

$$\kappa_i \in [0, \frac{1}{1-\beta}], c_i > 0, \rho_i > 0, \tau_i(t) \geq 0 \text{ for any } t \leq R - 10.$$

As taking the ratio of the second and the first equation yields:

$$\frac{\eta_{i3}}{\eta_{i2}} = \beta^R(\kappa_i - \frac{1}{1-\beta})$$

we derive the restriction that:

$$\frac{\eta_{i3}}{\eta_{i2}} \in [-\frac{\beta^R}{1-\beta}, 0]. \tag{20}$$

Conversely, if this restriction is valid then $\kappa_i \in [0, \frac{1}{1-\beta}]$.

The remaining equations are:

$$\begin{cases} \eta_{i2} = \frac{\rho_i}{c_i}\left(\rho_i\frac{\beta}{1-\beta} - 1\right), \\ \tau_i(t) = \frac{1}{c_i}\left(\rho_i\frac{\beta}{1-\beta} - 1\right) + \frac{\rho_i}{c_i}\beta^{-t}\frac{\eta_{i3}}{\eta_{i2}}. \end{cases}$$

Solving each equation in ρ_i yields:

$$\begin{cases} \rho_i = \frac{1-\beta}{2\beta}(1 \pm \sqrt{1 + 4\frac{\beta}{1-\beta}\eta_{i2}c_i}), \\ \rho_i = \frac{1-\beta}{\beta}\mu_i(t)(c_i\tau_i(t) + 1), \end{cases}$$

where we defined:

$$\mu_i(t) = \left(1 + \frac{1-\beta}{\beta}\beta^{-t}\frac{\eta_{i3}}{\eta_{i2}}\right)^{-1} \in \left[1, \left(1 - \frac{1-\beta}{\beta}\beta^{R-t}\right)^{-1}\right] \subset \left[1, \left(1 - \frac{1-\beta}{\beta}\beta^{10}\right)^{-1}\right],$$

if we use bounds (20). Note that the argument of the inverse function is decreasing with t because β^{-t} is increasing and $\frac{\eta_{i3}}{\eta_{i2}}$ is negative so that $\mu_i(t)$ is maximum at the end of the panel $t = R - 10$. It corresponds to the fact that $\tau_i(t)$ is decreasing. Let the upper bound be:

$$\bar{\mu}_i \equiv \mu_i(R - 10) = \left(1 + \frac{1-\beta}{\beta}\beta^{-(R-10)}\frac{\eta_{i3}}{\eta_{i2}}\right)^{-1}. \tag{21}$$

Furthermore, as $c_i > 0$ and $\tau_i(t) \geq 0$, then $\rho_i \geq \frac{1-\beta}{2\beta}$ and therefore $\eta_{i2} \geq 0$ and the only solution is the positive root. Thus:

$$\begin{cases} \frac{\beta}{1-\beta}\rho_i = \frac{1}{2}(1 + \sqrt{1 + 4\frac{\beta}{1-\beta}\eta_{i2}c_i}), \\ \frac{\beta}{1-\beta}\rho_i = \mu_i(t)(c_i\tau_i(t) + 1). \end{cases}$$

Parameter ρ_i is given by the intersection of these two curves in the positive orthant for (ρ_i, c_i). Let the equilibrium condition for ρ_i be the zeroes of this function:

$$\phi_i(c_i, \tau_i(t)) \equiv \mu_i(t)(c_i\tau_i(t) + 1) - \frac{1}{2}(1 + \sqrt{1 + 4\frac{\beta}{1-\beta}\eta_{i2}c_i}).$$

This gives easily the relationship between c_i and $\tau_i(t)$:

$$\tau_i(t) = \frac{1}{c_i}\left[\frac{1}{2\mu_i(t)}(1 + \sqrt{1 + 4\frac{\beta}{1-\beta}\eta_{i2}c_i}) - 1\right],$$

although the inversion of this function (c_i as a function of $\tau_i(t)$) is less easy to obtain.

To invert this function, note first that $\lim_{c_i \to 0} \phi_i(c_i, \tau_i(t)) = \mu_i - 1 \geq 0$. Furthermore, for $\tau_i(t) > 0$, we have:

$$\lim_{c_i \to \infty} \phi_i(c_i, \tau_i(t)) > 0.$$

Finally, note that

$$\phi_i(c_i, 0) = \mu_i(t) - \frac{1}{2}(1 + \sqrt{1 + 4\frac{\beta}{1-\beta}\eta_{i2}c_i})$$

is equal to zero if and only if:

$$c_i = c_i^L(t) \equiv \frac{\mu_i(t)(\mu_i(t) - 1)}{\nu_i}, \nu_i \equiv \frac{\beta}{1-\beta}\eta_{i2}.$$

Since $\mu_i(t) \geq 1$, a solution is acceptable if $\eta_{i2} > 0$.

Given these elements, the condition $\phi_i(c_i, \tau_i(t)) = 0$ for fixed $\tau_i(t)$ can either have no solutions, 2 solutions or 1 solution. This result is the simple consequence that $\phi_i(c_i, \tau_i(t))$ is the difference between a linear function and a concave function where the intercept with the $c_i = 0$ axis of the former is above the intercept of the latter. Those are described by the following where $\tau_0 > 0$:

$$\begin{gathered} If\ \tau_i(t) > \tau_0,\ \forall c_i > 0, \phi_i(c_i, \tau_i(t)) > 0 \text{ (no solutions)} \\ If\ \tau_i(t) \in (0, \tau_0),\ \exists!(c_i^1, c_i^2) > 0, \phi_i(c_i^j, \tau_i(t)) = 0,\ \text{(2 solutions)} \\ If\ \tau_i(t) = \tau_0, \exists! c_i^U > 0, \phi_i(c_i^U, \tau_0) = 0 \text{ (1 solution)} \end{gathered}$$

and using the additional case described above when $\tau_i(t) = 0$. (Implicitly, we do make explicit the dependence of τ_0 and c_i^j on t).

Let us compute τ_0 which by definition satisfies two properties translating that both the function and its derivative are equal to zero at the point of tangency of the linear function and the concave function:

$$\begin{aligned} \phi_i(c_i^U, \tau_0) &= \mu_i(t)(c_i^U\tau_0 + 1) - \frac{1}{2}(1 + \sqrt{1 + 4\nu_i c_i^U}) = 0, \\ \frac{\partial \phi_i}{\partial c_i}(c_i^U, \tau_0) &= \mu_i(t)\tau_0 - \frac{\nu_i}{\sqrt{1 + 4\nu_i c_i^U}} = 0. \end{aligned}$$

Using the second equation yields:

$$c_i^U = \frac{1}{4\nu_i}\left[\left(\frac{\nu_i}{\mu_i(t)\tau_0}\right)^2 - 1\right].$$

Plugging into the first equation yields:

$$\mu_i(t)(\left[\left(\frac{\nu_i}{\mu_i(t)\tau_0}\right)^2 - 1\right]\tau_0 + 4\nu_i) = 2\nu_i(1 + \frac{\nu_i}{\mu_i(t)\tau_0})$$

and therefore:

$$\mu_i(t)(\left[\left(\frac{\nu_i}{\mu_i(t)}\right)^2 - \tau_0^2\right] + 4\nu_i\tau_0) = 2\nu_i(\tau_0 + \frac{\nu_i}{\mu_i(t)}).$$

The equation to solve is:

$$\mu_i(t)\tau_0^2 - 2(2\mu_i(t) - 1)\nu_i\tau_0 + \frac{\nu_i^2}{\mu_i(t)} = 0.$$

The (simplified) discriminant function is:

$$\Delta' = \nu_i^2(2\mu_i(t) - 1)^2 - \nu_i^2 = 4\nu_i^2(\mu_i(t) - 1)\mu_i(t) \geq 0,$$

and the solutions are:

$$\tau_0 = \nu_i\frac{2\mu_i(t) - 1 \pm 2\sqrt{(\mu_i(t) - 1)\mu_i(t)}}{\mu_i}.$$

There are two solutions although the largest one corresponds to a negative c_i^U (always because of the geometry of the problem). We thus have:

$$\tau_0 = \nu_i\frac{2\mu_i(t) - 1 - 2\sqrt{(\mu_i(t) - 1)\mu_i(t)}}{\mu_i}.$$

Replacing yields:

$$c_i^U = \frac{1}{4\nu_i}\left[\left(\frac{1}{2\mu_i(t) - 1 - 2\sqrt{(\mu_i(t) - 1)\mu_i(t)}}\right)^2 - 1\right].$$

As a summary, we can define a continuous one to one mapping, $c_i(\tau_i(t))$ between $\tau_i(t) \in [0, \tau_0]$ and $c_i \in [c_i^L, c_i^U]$ so that $\phi_i(c_i, \tau_i(t)) = 0$. It is simple to prove that:

$$\frac{\partial\tau_0}{\partial\mu_i(t)} < 0, \frac{\partial c_i^L}{\partial\mu_i(t)} > 0, \frac{\partial c_i^U}{\partial\mu_i(t)} < 0$$

so that as $\mu_i(t)$ is increasing with t, the smallest interval is obtained when $t = R - 10$, so that:

$$c_i \in [c_i^L, c_i^U] \Longrightarrow \rho_i \in \left[\rho_i^L, \rho_i^U\right],$$

in which:

$$\rho_i^j = \frac{1-\beta}{2\beta}(1 + \sqrt{1 + 4\nu_i c_i^j}).$$

This allows to show the converse result.

Let $\eta_{i2} > 0$ and $\frac{\eta_{i3}}{\eta_{i2}} \in [-\frac{\beta R}{1-\beta}, 0]$ so that $\mu_i(t) \geq 1$. Then $\nu_i > 0$ and c_i^L, c_i^U are positive. There thus exists a pair $(c_i, \tau_i(t))$ belonging to a non empty set in the positive orthant. ρ_i is well defined as well.

Note: One special condition is when $c_i = c_i^L$ so that $\tau_0 = 0$ exactly at $t = R - 10$.

2.5 Transitory earnings

In equation (14), transitory earnings $\delta_i^y(t)$ are due to individual specific and aggregate shocks, $\delta_i(t)$ net of human capital depreciations, $\lambda_i(t)$. To this we might add measurement errors $\zeta_i(t)$ to obtain that random shocks are described by:

$$\delta_i(t) - \sum_{l=s}^{t-1} \lambda_i(l) + \zeta_i(t).$$

Even if measurement errors are independent over time, the effects of the first two transitory components may persist across periods and generates autocorrelation in the earnings residuals. Indeed, the deviation of the rate of return $\delta_i(t)$ from the market rental rate is due to individual specific factors and the match the individual is with a specific firm and this is likely to persist over time. Depreciation factors included in $\sum_{l=0}^{t-1} \lambda_i(l)$ are highly persistent if $\lambda_i(t)$ is independent over time. It indeed generates a random walk if $\lambda_i(t)$ is iid over time. Nevertheless it needs not be so if $\sum_{l=0}^{t-1} \lambda_i(l)$ is stationary, that is that depreciation shocks are partly compensated in the future. Layoff shocks that force agents to change sectors might be an example of a long persistence in these factors. For the sake of generality we will not impose any structure on these shocks in the econometric model.

Next section describes how we deal empirically with this model of the earnings formation process.

3 Econometric Modelling of Earning Dynamics

In section 3.1 we review the previous literature on earnings dynamics and the estimation of it, then in section 3.3 we detail our estimation strategy. We first concentrate on the covariance structure implied by the reduced form earnings equation (14).

3.1 Literature on Earning Dynamics

The literature on earnings dynamics studies the covariance structure of earnings residuals. Modeling the increasing variance over the life cycle and fitting the residuals autocorrelation are the principal goals of this literature. Writing log earnings for individual i at time t Y_{it} as a function of observed individual characteristics X_{it} a vector of individual characteristics β and a residual component u_{it} orthogonal to X_{it}, we have:

$$Y_{it} = X_{it}\beta + u_{it}.$$

Economists have proposed several decompositions of the variances of u_{it} into permanent and transitory factors. The different empirical specifications differ in their degree of generality and in their implicit assumptions. The canonical model presented in Lillard and Willis (1978) is based on the following decomposition of log earnings residuals

$$u_{it} = \eta_i + v_{it}, \quad \eta_i \sim iid(0, \sigma_\eta^2),$$

with η_i an individual effect generating inter-individual differences in earnings levels. It can be interpreted as initial human capital. Considering v_{it} as *iid* means that the autocorrelation of an individual error term over time is only due to the presence of the permanent components, which is a strong assumption. Therefore, v_{it} capturing transitory differences in earnings residuals whose effect decreases over time as been modeled by Lillard and Wilis as an AR(1) process. Using PSID data they assume that the two components are orthogonal. Then, they compute the fitted probabilities of the model to the transition into and out of poverty in the U.S. This landmark paper has been the starting point of the literature.

Lillard and Weiss (1979) have extended the model allowing the permanent component to increase the variance with age. The *random growth* model introducing individual growth paths writes

$$u_{it} = \eta_{i1} + \eta_{i2}t + v_{it}.$$

In that framework η_{i2} is a mean zero random individual effect in experience. A similar approach has been followed by Hause (1980), Baker (1997), and Cappellari (2004). Importantly, this specification implies that it is possible to test Mincer's (1974) theoretical prediction that differences in earnings should increase at the beginning of the life cycle until high investors in human capital catch up low investors. Empirically this would translate into a negative covariance $cov(\eta_{i1}, \eta_{i2})$ between the two individual effects and this result has been confirmed by the previous studies.

Additionally, the transitory part of the model v_{it} has remained a low order moving average or an AR(1) process.

In an influencial paper MaCurdy (1982) has estimated a different specification allowing for time varying coefficients of the transitory process and heteroskedasticity of the white noise term. He proposed an ARMA structure without individual fixed effects. Hence, for example

$$\begin{aligned} \eta_{it} &= \eta_{it-1} + \epsilon_{it} \\ v_{it} &= \zeta_{it} + \psi_1 \zeta_{it-1} + \psi_2 \zeta_{it-2} \end{aligned}$$

where ϵ_{it} and ζ_{it} are independently and identically distributed. MaCurdy claims that the two specifications cannot really be distinguished in levels, but the ARMA structure represents a better fit for earnings in differences. In his application on the PSID data MaCurdy (1982) concludes to an ARMA(1,2) and cannot reject the hypothesis of a unit root for the permanent component. Therefore in the literature this model has been called the *random walk model* of earnings dynamics. The same specification has been estimated by Abowd and Card (1989), Moffit and Gottschalk (1995) and Lillard and Reville (1999) on US data, Dickens (2000) on U.K. data, Cappellari (2004) on Italian data, and Baker (1997) on Canadian data. Many other studies use the same framework such as Moffitt and Gottschalk (2002) and (2008), Kalwji and Alessie (2007) and Sologon and O'Donoghue (2009) but they favor a shorter dynamics with an ARMA(1,1) process.

Methodological contributions have generalized the model in another direction. Geweke and Keane (2000) investigate the normality assumption relative to the white noise in the ARMA structure. Implementing bayesian inference using the Gibbs sampler, they show that the share of the variance coming from permanent individual heterogeneity terms is larger than under a Gaussian model and that in the cross-section covariate effects are reduced. Hirano (2002) uses a Bayesian framework to propose a semi-parametric estimator for autoregressive panel data models. In his application, the normality assumption proves to be restrictive. Attention to this issue should be addressed in our empirical application. In a different framework Bonhomme and Robin (2009) focus on the same issue and model the change over time in earnings using copula. It make it possible to represent partially non parametrically the marginal distributions of earnings and it provides a flexible modelling of the joint distribution between different time periods. However, as they have only three-years panel data the dynamic is restricted to be short. Moreover, in this framework it is generally difficult to include unobserved heterogeneity except in the approximation of its distribution by a finite number of point of support.

Other recent contributions to the literature have relaxed the assumption that innovations of permanent and transitory processes are *iid*. These approaches model explicitely a dynamic form of heteroskedasticity. Hence, Meghir and Pistaferri (2004) postulate an ARCH(1) data generating process for the permanent and for the transitory shocks. That is

$$\begin{aligned} E(\epsilon_{it}^2) &= \kappa_t + \gamma\epsilon_{it-1}^2 + \lambda_i \quad \text{Permanent component} \\ E(\zeta_{it}^2) &= \phi_t + \phi\zeta_{it-1}^2 + \xi_i \quad \text{Transitory component} \end{aligned}$$

with κ_t and ζ_t representing year effects capturing the way the variance of transitory and permanent shocks vary over time and λ_i and ξ_i are individual fixed effects representing occupational choices for example. Estimating the model by educational group, Meghir and Pistaferri (2004) conclude that the variance of shocks persists in some education groups and that the ARCH effects are present in the data. In a similar framework Hospido (2010) models the variance of earnings but instead of implementing a GMM approach, she uses a likelihood estimator, however she does not model separately permanent versus transitory factors.

Guvenen (2007) has studied the implications of the form of the income process on consumption inequality. He compares the predictions of the *random walk* model with those of the *random growth* using a model of life cycle consumption on simulated data. Guvenen concludes that a model with heterogenous earnings growth paths is better able to replicate the observed change in consumption inequality than a model with a unit root. Therefore, he advocates for using the former. In Guvenen (2009) the sources of identification between the two income processes is more deeply investigated. A major difference between the model in which individuals have heterogenous earnings profile and the model in which they are subject to persistent shocks is that with the former, the autocorrelation in the earnings residuals in differences will persist because of the term $\eta_{i2}t$. While with the latter the autocorrelations will become insignificantly different from zero after some time (see Meghir and Pistaferri, 2004, p798 for a graphical example.) Empirical evidence in MaCurdy (1982), Abowd and Card (1989), Moffit and Gottschalk (1995) and Meghir and Pistaferri (2004) favor the hypothesis that autocovariance decline in absolute value and is after some time not statistically different from 0 contradicts the *random growth* model.

Finally most of these studies have used reduced form models. Very few studies have been able to present structural models reflecting the dynamics of earnings. Among the first analysis Farber and Gibbons (1996) consider a model in which firms discover over time the skills of their employees. They demonstrate that earnings will follow a stochastic process that can be represented by shocks compatible with a random walk model. More recently, Postel-Vinay and

Turon (2010) present a job-search model in which earnings determined by mutual consent are affected by productivity shocks. These shocks lead earnings to the usual earnings dynamics model. Lastly, two recent contributions model simultaneously the earnings dynamics and the job mobility decisions of workers. For instance, Altonji, Smith and Vidangos (2009) add to the linear earnings equation different selection equations to control for selection on the transitions into and out of the labor market. Hoffman (2010) proposes to the same exercise but using a dynamic discrete choice model of career progression. These studies underlines the potential selection effects of not considering the endogenous decision to participate to the labor market.

3.2 Covariance Structures and the Distributions of Individual Effects

The model of earnings that we specify in the subsection below, belong to the literature on covariance structures in the dynamic panel data literature. The difference between those set-ups is that in the dynamic panel data literature, papers emphasize the estimation of coefficients of exogenous or predetermined variables which are not present in our case. Nonetheless, the lessons from this literature are useful to remember here. As soon as GMM estimation was used to estimate dynamic models, it became clear that the range of moments involved was larger than the usual GMM case and that as a consequence, first order asymptotics were a poor guide in empirical research. Furthermore, the issue of weak instruments becomes more important under strong persistence or near unit root dynamics (Arellano and Bover, 1995) and this suggested to consider the reinforcement of identification assumptions (Blundell and Bond, 1998). This is why some researchers propsoed to return to an OLS set up adding a bias correction step (Hahn and Kuersteiner, 2004) or to maximum or quasi-maximum likelihood estimators (Hsiao, Pesaran and Tahmiscioglu, 2002, Dhaene and Jochmans, 2009). Another direction was recently proposed by Han, Philips and Sul (2010) in the case of AR(p) models under mean stationarity whose properties are robust and simple to derive under both stationary and non stationary cases.

As we stick to a framework in which the initial condition is supposed to have been generated by another stochastic process and that T is sufficiently small so that asymptotic stationarity properties are not satisfied, the GMM framework remains our reference. Alvarez and Arellano (2003) analyses the asymptotic properties of GMM estimators using double asymptotics in N and T. Okui (2009) derives the small sample biases not only in the mean but also in the variance of GMM estimates because of the presence of too many moments even in the case in which T is small. He suggests some moment selection mechanism in order to limit the importance of these biases by, to put it briefly, selecting out moments between variables which are too far apart in time. Those moments are far more likely to contribute to the bias and not the variance.

We use another route of maximum or quasi-maximum likelihood methods that reduces the number of moments available for estimation. We use normality assumptions although our estimates remain consistent in a non normal framework under the weak conditions of quasi-maximum likelihood (Gouriéroux, Monfort and Trognon, 1984). The selection of moments keeps being the optimal way, basically only under a normality assumption. We use an otherwise very flexible framework in what refers to initial conditions and the assumptions about the evolution of variances and autocovariances over time. Specifically, we use a random effect estimator as suggested by Alvarez and Arellano (2004) in a comparison with other fixed T consistent estimators. Specifically, this estimator seems to dominate in most Monte Carlo exercises the maximum likelihood estimator using differenced data (Hsiao et al., 2002) and the corrected within group estimator. Bai (2009) also derives MLE estimates in factor models in which the time factors are unknown and in the presence of covariates.

Horowitz and Markatou (1996) estimate semi-parametrically the distributions of the white noises and the individual effects. However, in their approach the dynamic dimension has to be restricted to be AR(1). Geweke and Keane (1998) and Hirano (2002) have generalized the model in the same direction by implementing a Bayesian approach to estimate posterior distributions of the parameters. Bonhomme and Robin (2010) construct an estimator of the distribution of factors using empirical characteristic functions and apply this estimator to analyze the distributions of permanent and transitory components of earnings using the PSID. Arellano and Bonhomme (2010) look in detail to the identification of individual effects when the time dimension is fixed and show that its variance is identified under restrictions of the dynamics as we do. They also propose the construction of non parametric estimates for the distribution of the individual factors. Finally, Cunha, Heckman and Schennach (2010) uses results from Schennach to show how non parametric estimates of moments of latent variables can be constructed from various measurements of these variables using empirical characteristic functions and inverse Fourier transforms.

3.3 Model Specification

Equation (14) can be written with respect to three individual factors $\eta_i = (\eta_{i1}, \eta_{i2}, \eta_{i3})$ such that with a slight abuse of notation

$$y_{it} \equiv \log y_{it} = \eta_{i1} + \eta_{i2}t + \eta_{i3}\frac{1}{\beta^t} + v_{it} \text{ for any } t = 1, ., T. \tag{22}$$

We now specify the stochastic process followed by the shock v_{it} whose variances and autocovariances are partly time heteroskedastic while they have a limited ARMA structure so as to allow the identification of the distribution of the main variables of interest which are the individual effects. Similar specifications of the dependence structure are developed in Alvarez and Arellano (2004), Guvenen (2009) and Arellano and Bonhomme (2010). 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 $MA(q)$:

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

Defining the time index accordingly, we shall assume that initial conditions of the process $(y_{i(1-p)}, ., y_{i0})$ 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

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

The structure of Ω_z structure is detailed in Appendix A although it can be summarized easily. The correlations between initial conditions and individual effects are not constrained, while innovations ζ_{it} are supposed 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 $(\eta_{i1}, \eta_{i2}, \eta_{i3})$ 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

$$Vy_i = (y_{i(1-p)}, ., y_{i0}, y_{i1}, ., y_{iT}) \equiv \Omega_y,$$

where Ω_y a function of the parameters of the model that are the autoregressive parameters $\{\alpha_k\}_{k=1,...,p}$, the moving average parameters $\{\psi_k\}_{k=1,...,q}$, the covariance matrix of η, Σ_η, the heteroskedastic components $\{\sigma_t\}_{t=1,\ldots,T}$) and finally, the covariance of fixed effects and initial conditions.

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:

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

although the optimality properties are derived from the normality assumptions.

The pseudo lilkelihood setting is particularly well adapted to the case in which there are mssing data in the earnings dynamics. Using GMM estimation procedures, we would have to rewrite each moment condition in which there are missing data by replacing the missing variables by their expressions as a function of observed variables. This is untractable in such a dataset in which the number of different missing structures is very large while this is handled with parsimony in a pseudo likelihood setting. 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.

3.4 Constraints and Structural Parameters

It is not possible to impose the constraints on the parameters at the estimation stage in the random effect model. It is nonetheless possible to use random effect estimators in order to construct estimates of individual effects after the estimation. In a log likelihood framework, we obtain estimates as linear combinations of residuals, the linear combinations being given by the covariance matrix estimated in the random effects model. Appendix B.1 develops the corresponding analytic computations that lead to define the individual effects estimates as:

$$\hat{\eta}_i = \bar{\eta}_{g \ni i} + \hat{B} u_i^{[1-p,T]}.$$

Replacing $u_i^{[1-p,T]}$ by its expression as a function of η_i, it is easy to show that these estimates are measured with errors and we have:

$$\hat{\eta}_i = \eta_i + \hat{B} w_i^{[1-p,T]}.$$

in which $w_i^{[1-p,T]}$ is supposed to be distributed as a multivariate normal distribution with mean zero and covariance matrix equal to identity.

Therefore, estimates $\hat{\eta}_i$ do not necessarily satisfy the constraints:

$$\eta_{i2} > 0 \text{ and } \frac{\eta_{i3}}{\eta_{i2}} \in [-\frac{\beta^R}{1-\beta}, 0].$$

We let $\lambda = \frac{\beta^R}{1-\beta}$ and write these contraints as:

$$\eta_{i2} > 0, \ \eta_{i3} < 0 \text{ and } \eta_{i3} + \lambda\eta_{i2} > 0.$$

As explained in Appendix B.4 we can construct constrained estimates $\hat{\eta}_i^c$ by projecting $\hat{\eta}_i$ on the set of constraints using the distance defined by the (log)-likelihood function criteria. We can also construct the distribution of this distance in the data:

$$d(\hat{\eta}_i^c, \hat{\eta}_i) = (\hat{\eta}_i^c - \hat{\eta}_i)'\hat{B}^{-1/2}(\hat{\eta}_i^c - \hat{\eta}_i).$$

We can also use simulation and construct simulated constrained estimates using the developments in Appendix B.5. We can then get simulated samples $\hat{\eta}_i^{c,s}$ of simulated constrained estimators of η_i.

3.5 Counterfactuals

We want to analyze the impact of a change in the levels of κ_i and ρ_i.

3.5.1 Survival probabilities

As for the first one, we change κ_i in such a way that $\kappa_i - \frac{1}{1-\beta}$ is divided by a factor $1+\alpha$. It corresponds to an increase in the survival probabilities after $t = R$ considering that all other things remain equal (social security benefits in particular). A simple scheme might justify this assumption. Suppose that social security contributions are proportional to earnings until retirement. Then this coefficient of proportionality is an increasing function of the survival probability at the level of the population. It does not affect the variance of log earnings.

As:

$$\frac{\eta_{i3}}{\eta_{i2}} = \beta^R(\kappa_i - \frac{1}{1-\beta}),$$

and as ρ_i and c_i are fixed by assumption so that η_{i2} is fixed, the simple experiment corresponds to the division of η_{i3} by a factor $1+\alpha$. Let $\eta_{i3}^{'} = \eta_{i3}/(1+\alpha), \eta_{i2}' = \eta_{i2}, \eta_{i1}' = \eta_{i1}$, we have that:

$$y_i' = M(\beta)\eta_i^{'} + v_i.$$

So that we can compute the counterfactual variance as:

$$V(y_i') = V(M(\beta)\eta_i^{'} + v_i),$$

and use simple plug in estimators for this variance.

3.5.2 Human capital technology

The construction of the counterfactual for the human capital technology is more involved. We are interested in changing parameter ρ_i into ρ_i' holding constant parameters κ_i and c_i. Given that:

$$\begin{aligned}
\eta_{i2} &= \frac{\rho_i}{c_i}\left(\rho_i\frac{\beta}{1-\beta}-1\right), \\
\frac{\eta_{i3}}{\eta_{i2}} &= \beta^R(\kappa_i-\frac{1}{1-\beta}),
\end{aligned}$$

the counterfactual is constructed using $\rho_i' = (1+\alpha)\rho_i$ and:

$$\begin{aligned}
\eta_{i2}' &= \frac{\rho_i'}{c_i}\left(\rho_i'\frac{\beta}{1-\beta}-1\right), \\
\frac{\eta_{i3}'}{\eta_{i2}'} &= \beta^R(\kappa_i-\frac{1}{1-\beta}) = \frac{\eta_{i3}}{\eta_{i2}}.
\end{aligned}$$

In the first expression we have to replace:

$$\frac{\beta}{1-\beta}\rho_i' = \frac{\beta}{1-\beta}(1+\alpha)\rho_i = \frac{1+\alpha}{2}(1+\sqrt{1+4\frac{\beta}{1-\beta}\eta_{i2}c_i}).$$

We can derive that:

$$\frac{\beta}{1-\beta}\eta_{i2}' = \frac{1+\alpha}{2c_i}\left[\alpha(1+\sqrt{1+4\frac{\beta}{1-\beta}\eta_{i2}c_i}) + 2(1+\alpha)\frac{\beta}{1-\beta}\eta_{i2}c_i\right]$$

There are two complications.

First c_i is unknown and known by interval only i.e. $c_i > c_i^L$. Nevertheless, it is easy to show that $\frac{\beta}{1-\beta}\eta_{i2}'$ is decreasing in c_i and reaches the lowest value $(1+\alpha)^2\frac{\beta}{1-\beta}\eta_{i2}$ when $c_i \to \infty$. Thus:

$$\frac{\beta}{1-\beta}\eta_{i2}' \in \left[(1+\alpha)^2\frac{\beta}{1-\beta}\eta_{i2}, \frac{1+\alpha}{2c_i^L}\left(\alpha(1+\sqrt{1+4\frac{\beta}{1-\beta}\eta_{i2}c_i^L}) + 2(1+\alpha)\frac{\beta}{1-\beta}\eta_{i2}c_i^L\right)\right],$$

and a similar expression obtains for η_{i3}'.

Second, even if as before we have

$$y_i' = T(\beta)\eta_i' + v_i,$$

the computation is now highly non linear and the variance of earnings is not directly related to the estimated quantities at the first stage.

3.5.3 Estimation

We estimate by simulation imposing the constraints on the ηs. As explained above, we have that:

$$\hat{\eta}_i = By = \eta_i + B\varepsilon_i$$

so that:

$$\eta_i = \hat{\eta}_i - B\varepsilon_i.$$

We need a distribution assumption for ε_i to impose the constraints. We assume that ε_i is normally distributed and we compute the counterfactual using the expressions above. The bias can be computed as in Arellano and Bonhomme, 2010. We could also estimate by simulated deconvolution (Mallows, 2007, Arellano and Bonhomme, 2010) or by deconvolution (Schennach, 2004) but we let these extensions for future research.

4 Data

4.1 Sample Selection

The French DADS panel dataset on earnings is an extraction from an administrative source named *Déclarations Annuelles de Données Sociales*. DADS data is collected through a mandatory data requirement (by French law) that all firms that have employees must fill in for social security and tax verification purposes. All employers ought to send to the social security and tax administrations the list of all persons who have been employed in their establishments during the year. For each person, are indicated the total wage earnings firms have paid, as well as the beginning and ending dates and a short description of the job. Each person is identified by a unique individual social security number which makes possible the follow-up of individuals through time.

The French National Statistical Institute (INSEE) has been authorized since 1976 to draw a sample from this dataset at a sampling rate of 4%. The sampling device is such that all individuals who were born in October of even years should be included in this sample. Nevertheless, there are two main reasons why observations can be missing. First, data were not collected in three years (1981, 1983 and 1990) for reasons specific to INSEE. Second, this dataset is restricted to individuals employed in the private sector or in publicly-owned companies only. As a consequence, this analysis is restricted to individuals who have been employed at least one year between 1976 and 2007 in the private sector or in a publicly-owned company. In fact, we further restrict the sample to men entering the labor market in 1977 and working in the private sector in 1982 and 1984. The definition of entry here is the same as in Le Minez and Roux (2002). We consider that an individual has entered the labor market as soon as this individual has occupied the same job for more than 6 months and is still employed the following year, possibly in a different firm. The date of entry defining the cohort to which the individual belongs, we focus on the cohort of entrants in 1977. We also consider full time jobs only and censor information about part-time jobs.

We impose these restrictions in order to concentrate on a relatively homogeneous sample of workers with a long term attachment to the labor market to which private firms have access. Admittedly, it does not represent the full working population. Because of the lack of any 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 factors or idiosyncratic components that we estimate in the following refer to this subpopulation.

The empirical analysis uses daily earnings. It is defined as full earnings divided by the number of days worked. In order to weaken the possible impact of measurement error, we coded as missing the first and last percentiles of each annual earnings distribution. In the empirical analysis we do not condition on individual observable characteristics as in the traditional Mincer's wage equation, since individual characteristics cannot be separated from individual unobserved heterogeneity terms. Few observable characteristics are available apart from age of entry on the labor market and a rough measure of education grouping the first job into three categories. As a measure of skill, we prefer to use a grouping given by the age of entry. The first group includes individuals entering the labor market when they are less than 20 years old, the second group of individuals enter between age 20 and 23 and the last group from age 24.

We analyze log earnings centered with respect to the average log earnings of workers within the same age of entry and education group at each point in time. That is, we compute y_{it} our daily log earnings as :

$$y_{it} = \log(E_{it}) - \overline{\log(E_{it})_{et}}$$

with a the age of entry group, t the time period and e the education group.

4.2 Data description

Table 2 reports descriptive statistics of the sample. The sample size is 7446 observations in 1977 and 4670 in 2007. Age of entry groups defined above are of unequal size, the low skill group being the largest. Attrition follows a somewhat irregular pattern which is partly due in the first years to our sampling design since we required that wage earners be present in 1977, 1978, 1982 and 1984. Some years are also completely missing (1981, 1983 and 1990). 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 and missing data patterns in 1994 are very similar across age of entry groups.

To complete this information, Table 3 gives a dynamic view of attrition. This Table reports the frequencies of reported values by pairs of years. For instance, the column 1977, describes the global features of attrition. Attrition is quite severe in the first normal (after selection) year, 1985 since 15% of individuals exit between 1984 and 1985. This is true in every adjacent years at the beginning of the sample period (other columns for instance in cell 1987, 1988) but it is decreasing over time to reach 7 or 8% at the end of the panel. Year 1994 confirms its exceptional status since attrition between 1994 and 1995 is very low. More generally though, most individuals reenter the panel quickly since the attrition at two year intervals is only marginally larger than the one observed at one year intervals (for instance the two cells in 1977, 1985 and 1986, indicate attrition of 15% and 16.5%) although this varies somewhat over time. Finally, there is a core of observations which are almost always reported in the panel. Looking at the row 2007, we can see that out of the 62.7% of the complete sample of individuals present in this year, it is hard to have less than 80% of this sample which is not present between 1985 and 2006 at the exception of 1994 again.

We report in Figure 1, the increase of average earnings over the period for the three groups defined by age of entry. These are earnings at current prices although the shape of real earnings is hardly different. Inflation, as measured by consumer prices, leads to a substracting factor for current earnings over the whole period which is equal to 1.17. 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. We do not report the evolution of average earnings by groups defined by education and age of entry, the only individual characteristics that are available in the dataset, although these evolutions are parallel to the ones graphed in Figure 1. Nonetheless, as already said, the variance of log earnings that we consider from now on are computed by centering log earnings with respect to averages of all covariates and periods.

The left panel of Figure 2 represents the change in the cross-sectional variance of (log) earnings for the full sample, while the right panel represents the variance by groups defined by age of entry. The first few years witness a strong variability of earnings. From the sixth year, 1982 (respectively the third, 1980), the variance drops for the low skill groups (resp. the others) whereas it increases gradually over the rest of the sample period till 2007 (except in 1994, and 2003 at a lesser degree, which confirm the outlier status of these years). From the right panel one can notice that late entrants on the labor market experience a higher variance level and a larger rate of growth of the variance in their earnings trajectories. The full covariance matrix is reported in Table 4 to give information about correlations although this is easier to use graphs to describe the main features of this matrix. Figure 3 displays for the full sample the autocorrelation with an early year, 1986, and a late year, 2007. This Figure reveals an assymmetric pattern over time which is quite robust to the choice of these years. The correlation between earnings at years $t-k$ and posterior t is quickly disappearing between t and $t-k$ in early years of the panel while it is roughly linear in lags in late years. Figure 4 takes a different view that confirm 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 we move to the end of the period, the larger the autocorrelation coefficients are.

Finally, Table 5 reports the autocorrelation patterns of the first differences in the earning residuals. Contrary to what is found in some papers in the literature (for instance, Meghir and Pistaferri, 2010) we do not find strong evidence that the correlation disappears after taking a two period difference. Some very long difference autocorrelations seem significant and no regular pattern seems to emerge.

5 Results

We first present the estimated parameters of the reduced form earnings equation by random effect ML estimation and we discuss the selection of the ARMA specification. In the next section 5.2 we detail the procedure we implement to estimate individual factors and simulate a sample of observations on the structural parameters (rate of return and the terminal value coefficient). Lastly, in 5.3 we assess the impact of changes in the structural parameters on the variance of earnings by direct estimation or simulation.

5.1 Random effect estimation and reduced form parameters

Firstly, we estimate simultaneously the covariance matrices of the permanent and transitory components of the error as well as their correlation with the initial conditions. The former is composed by three individual unobserved factors $(\eta_{i1}, \eta_{i2}, \eta_{i3})$, while the latter is represented by an ARMA process as explained in the previous section. To solve the identification issue that we face in attributing the magnitudes of variances and covariances of log earnings over time to the effects of the individual specific factors or to the effects of the idiosyncratic error terms, we chose to fix the discount factor β to the value 0.95. Arellano and Bonhomme (2010) shows that along with a finite lag specification assumption about the ARMA process, this assumption is sufficient to get identification.

Table 6 provides the values of the Akaike criterion based on the likelihood values for the different specifications varying the orders of the autoregressive and moving average components by going from an ARMA(1,1) to an ARMA(3,3). Unsurprisingly, increasing the number of AR or MA components strongly increases the value of the sample likelihood function. Nonetheless, increasing the number of AR and MA components beyond 3 faces difficulties in the implementation since it involves a year, 1981, in which obervation is completely missing. This is why we did not pursue further the exploration of higher orders for the ARMA processes. According to the Akaike criterion we would choose the ARMA(3,3) specification, a much more persistent specification than in most studies in the literature. Nevertheless, the estimates of the ARMA(3,3) exhibit some estimates which are very imprecise, specifically the ones describing the correlation between initial conditions and the MA components (Table 7). That is why in the rest of the analysis we will use as a pivot result the ARMA(3,1) model although the robustness of our results to this choice should be checked.

Table 7 details the parameters for the different ARMA processes. Each column reports results for different ARMA(p,q) specifications until $p = q = 3$. In every model, autoregressive coefficients as well as their sum, which can be interpreted as a long term effect, remain largely lower than one. The sum of the AR coefficients reflects a high persistence of shocks though it is far enough from one to reject an unit root. A formal statistical test concludes with no doubt that the process is stationary (see Magnac and Roux, 2009). This result parallels the result of Alvarez and Arellano (2004) on US and Spanish data or of Guvenen (2009). The AR coefficients are ranging from .2 to .02 in the ARMA(3,1) specification and would describe the persistence of shocks due to unemployment spells or mobility for instance while the MA coefficient is negative and might stand for measurement errors.

The estimate of the covariance matrix of the individual factors is quite stable across the different specifications even if it can be identified only under a specific assumption for the dynamics. Their variances are very precisely estimated at around .30 for the fixed level factor, η_1, and the geometric factor, η_3, and at around .03 for the linear trend factor, η_2. The correlation between the linear trend and geometric factors is very strongly negative and equal to -.95 consistently across the specifications. This is to be expected if the structural constraint derived above between η_2 and η_3 ($\eta_3 \in [-\lambda\eta_2, 0]$) is verified. We will analyze this issue more in detail below. The other correlation coefficients are also quite strong being negative and around -0.6 between the geometric and the level factors, η_3 and η_1 and positive and around .4 between the level and linear trend factors. The sign of the latter correlation coefficient is to be expected if the level of human capital at the entry date is positively correlated with the returns to human capital which govern the linear trend factor.

The correlations between initial conditions and these factors are also informative. They are significant and have an economically significant magnitude of around .2 or .3 in absolute value. The estimated correlations between the linear trend and geometric factors η_2 and η_3, and the initial conditions are similar to the estimated correlations between both of them and the level factor. They are respectively significantly positive and negative. More surprisingly, the correlation between η_1 and the initial conditions is also negative. That would indicate that individuals endowed with higher starting human capital stock have more difficulties to acquire immediately the level of earnings that correspond to their skill levels.

Finally, the estimated variance of the idiosyncratic terms is reported in Table 8. Note first that these parameters are identified even in the years 1981, 1983 and 1990 in which there is no information although estimates are imprecise and have a magnitude that can differ widely from the others and across ARMA specifications. Regarding the normal years, they start from a rather high level in the first three years between .20 and .30. They generally decrease over the sample periods albeit very slowly. Between 1984 and 2000 they are quite precisely estimated at a level around .18, except the exceptional year 1994 in which we know that the measurement error is large, and levels off at around .14 after 2000 (except the exceptional year 2003). These estimates certainly pick up the patterns of autocorrelations increasing over time that we spotted in the raw data (see Table 4). Part of it is certainly attributable to measurement errors although another part of it could be attributed to a decreasing impact of shocks along the life cycle.

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 evolution of variances is reproduced but at a level which is higher than the observed level. It confirms that 1994 is an abnormal year. More broadly, we interpret this difference by the fact that the variance parameters σ_t are fitting both variances and autocorrelations and the latter seem to have more influence on the estimation. Indeed, the goodness of fit is very good for the autocorrelation coefficients that are reproduced in Figures 3 and 4. One possibility would be to allow for an additional measurement error term in every period, or in 1994 only, like in Guvenen (2009) in order to reconcile observed and estimated variances during the second part of the period.

5.2 Fixed effect estimation, structural restrictions and structural parameters

Using the previous estimates, it is easy to construct fixed effect estimates of the three individual factors. Appendix B.1 gives the relationship between the estimates and the estimates of the covariance matrices and log earning residuals as well as the way we impute back the earning averages to the individual factors. It is worth recalling that these estimates are not consistent if the number of periods T is fixed (for instance, Arellano and Bonhomme, 2010). Table 9 presents the estimates of the quantiles of their distributions distinguishing observations according to the number of periods we observe them (between 4 and 28). It is possible to notice the bias in $1/T$ since the larger the number of observed periods is, the lower the inter-quartile ratio for all three factors. Overall the median of the coefficient attached to the level factor is of the order of magnitude of the mean earnings at around 2.5 and the range between the 20 and 80% quantile is .5 if the number of periods is maximal ($T = 28$). The median of the coefficient of the linear trend factor which can be interpreted as the return to experience at the initial stage is of the order of 3 or 4% while its 20-80 quantile range is about 6%. Finally, the median of the coefficient of the geometric factor lies around -.17 and its inter-quantile range is .40. This coefficient enters multiplicatively in the curvature of the earnings profiles over time since the second derivative of the latter wrt time is this coefficient multiplied by $(\log \beta)^2 = 2.5.10^{-3}$. This squares well with the usual estimates of earnings equations predicting the maximal value of earnings at a time t close to $\log(\log(\beta)\eta_2/\eta_3)/\log \beta$ which is equal to 31.2 at the median estimates.

With these estimates in hand, we can directly evaluate the relevance of economic restrictions; We have three restrictions, the coefficient of the linear trend should be positive ($\eta_2 > 0$), the coefficient of the geometric factor should be negative ($\eta_3 < 0$) and these two coefficients should be related by $\eta_3 + \lambda\eta_2 > 0$. Parameter $\lambda > 0$ is fixed in the population and a function of β. There is some leeway in the choice of parameter λ depending on what is expected to be the proper horizon of investment. We chose parameter λ in the following as being the value that accomodates the better the economic restrictions while keeping in line with the conditions under which we estimated the model i.e. investments remain positive until the end of the panel.

Table 10 reports the frequencies of restriction violations using the previous estimates and the same presentation regarding the number of observed periods since estimating bias is proportional to their inverse. First, the frequency of rejections decreases with the number of sample periods. This is specifically the case when looking at the restriction that the coefficient of the linear trend factor should be positive and whose violation frequency drops down to 10% when the number of sample periods is maximal (from more than 50% when the observed periods are few). This is also true for the second restriction which is acceptable in 80% of the case (when $T = 28$). This is less true for the last restriction which is the most problematic. More than 50% at least of the observations do not comply with the restriction that the linear combination, $\lambda\eta_2 + \eta_3$, is positive. Nonetheless, it is also true that this linear combination is not very precisely estimated. Table 11 reports the same exercise as before by replacing the point estimates of the factors by the 95% confidence intervals of these estimates as their variances can be estimated. If some points in these intervals satisfy the restrictions, we say that there is no violation of the restrictions which is informative although arguably a rather biased view towards accepting restrictions. In this Table, violations of restrictions become very infrequent and specifically the third one.

Another way of representing those restrictions is brought about by Figure 5. The clouds of points for η_2 and η_3 is scattered around a downward sloping line and this represents the strong negative correlations between the two factors that was found in the covariance matrix estimated by random ML. This is no doubt attributable to the very different asymptotic behaviour of the three factors, one being a linear trend and the other being geometric. Second, points in orange (or light) refer to observations for which the sample periods are few (less than 20) and they are more scattered than the blue or dark points which refer to more continuously observed agents. Finally, the constraints are represented by the triangle in red or dark. This Figure makes clear that the satisfaction of the constraints are very sensitive to two key elements. The position of the origin point (0,0) whose estimation depends on the model we have for average earnings that is described in Appendix B.1 and that leads to the imputation of averages for ηs. Second, the λ parameter which determines the slope of the bottom line of the triangle. As already said, we chose λ in a way that it is maximal under the structural constraints that we have to satisfy.

It is also possible to compute the constrained estimates of the individual factors by projecting the unconstrained factors on the set of constraints using the quadratic metric given by the covariance matrix of the ηs estimated in the first-step random effect estimation. This procedure is explained in Appendix B.4. The quantiles of their distribution function appear in Table 12. They obviously satisfy the restrictions $\eta_2 > 0$ and $\eta_3 < 0$, the third restriction being more difficult to see although it is easy to check in a diagram which is not reported here. The constrained estimates of the median of the coefficient of the level factor which is not involved in any restriction, do not change significantly, in spite of being affected by the projection on the contraint set since the level factor is correlated with the other factors. This is also true for the median of the linear trend factor whereas its 20-80 % quantile range decreases to 4% from 6%. The coefficient of the third factor is more affected since its median decreases by 30% and the 20-80% quantile range decreases by almost 50%.

So as to evaluate the strength of the restrictions, we computed the distance between the unconstrained and the constrained estimates and compare this distance to the distance between the same constrained estimates and simulated unconstrained estimates using normal random draws for the simulations. In all these experiments, we use the covariance matrix of the ηs as a weighting matrix to compute the distance and as the basis for simulating the normal errors. Table 13 reports the quantiles of the distributions of the actual and simulated distances. The two distributions coincide rather well for all quantiles until 60% but the divergence becomes severe afterwards and specifically at the upper end. This can be either due to the rejection of the constraints or to the non normality of the factors which is a standard finding in studies that assess the normality of individual effects in earning functions (Hirano, 2002 for instance).

An interesting question arises as to whether these simulated constrained estimates are able to reproduce the pattern of variances over time. Arellano and Bonhomme (2010) quantifies the difference between random effect estimates and those who would be obtained by this simulation exercise (although with the additional twist of imposing constraints in our case). Figure 6 reports the curves associated to different number of observed periods. First, observations with more than 22 observed periods only were used since the estimates for the others are widely out of line with the random effect estimates. From this Figure, the larger the number of sample periods, the closer is the estimated variance to the random effect estimates.

Finally, these constrained reduced form estimates can be used to construct the structural parameters, at least the ones that are point identified or the estimated intervals for those which are interval identified. Parameter κ_i that governs the magnitude of post-retirement returns in human capital can be easily estimated using the distribution of η_3 and its distribution is bounded between 0 and $1/(1-\beta) = 20$. Figure 7 presents a kernel estimate of its distribution. Parameter ρ_i is only partially identified. The lower bound lies between .05 and .07 while the upper bound range is between .05 and .10. Figure B.6 reports the width of the estimated intervals as a function of the lower bound.

5.3 Counterfactual exercises

In this section, we report results related to the counterfactual exercises we proposed in Section 3.5.

We start by evaluating the counterfactuals using the random effect estimates that were derived in the first step. Figures 8 and 9 report the results of increases and decreases by 1 or 5% the variance of the structural parameters. For the point identified parameter κ, we proceed by changing κ into

$$\frac{1}{1-\beta} + (\kappa - \frac{1}{1-\beta})/(1+\alpha),$$

i.e. making κ closer to the same capitalized discount factor before retirement, $1/(1-\beta)$. It thus stands for instance for an increase in the probability of survival if pension rights are related to earnings and human capital stocks. For the interval estimate for ρ, we only report the lower bound of the impact since the upper bound is widely above the lower bound. In this experiment, it is safer to recognize than the upper bound is too large to be useful. In other words, we estimate the counterfactual impact as an interval which is unbounded from above.

In Figure 8, the left (resp. right) panel reports decreases (resp. increases) the variance of κ_i. Nonetheless, it increases (resp. decreases) the variance of earnings at the end of the sampled period. This seems to be due to the fact that other heterogeneity factors like heterogeneous returns can play more role in human capital investments. Increasing survival rates or making pensions closer to working life earnings tend to increase inequality beforehand through their effects on human capital investments. The impact of increasing the variance of returns to human capital investments is more standard. Figure 9 shows unambiguously that it increases the variance of earnings.

We can also use the simulated structural estimates as derived in the previous section to formulate other estimates for those experiments. Figures 10 and 11 report the impact of changes in κ and ρ on the averages of earnings over time. Both experiments increase earnings. Figures 12 and 13 report the impact of changes in κ and ρ on the variances of earnings over time as in the previous counterfactual experiment. Those confirm what was found in the previous counterfactual exercises.

6 Extensions

We now develop an extended model where investment costs depend on human capital levels. We present a version of the model in which human capital depreciates at a common and exogenous rate $\alpha \in (0;1)$ in the human capital accumulation equation

$$H_i(t+1) = H_i(t)^{\alpha} \exp[\rho_i \tau_i(t) - \lambda_i(t)].$$

This is equivalent to make human capital investments more and more costly when human capital levels increase.

Individuals maximize the present discounted value of their earnings streams, and their objective function is given by

$$V_t(H_i(t), \tau_i(t)) = \delta(t) + \log H_i(t) - \gamma_i \left(\tau_i(t) + c_i \frac{\tau_i(t)^2}{2} \right) + \beta E_t \left[W_{t+1}(H_i(t+1)) \right]$$

The first order condition of the maximization problem for $t < T$ is

$$-\gamma_i \left[1 + c\tau_i(t)\right] + \beta \rho_i H_i(t+1) E_t \left[\frac{\partial W_{t+1}}{\partial H_i(t+1)} \right] = 0. \tag{23}$$

The marginal value of human capital is the derivative of the Bellman equation so that by the envelope theorem:

$$\frac{\partial W_t}{\partial H_i(t)} = \frac{1}{H_i(t)} + \alpha \beta E_t \left[\frac{\partial W_{t+1}}{\partial H_i(t+1)} \right] \frac{H_i(t+1)}{H_i(t)} \tag{24}$$

Introducing condition (23) into condition (24) we obtain

$$\frac{\partial W_t}{\partial H_i(t)} = \frac{1}{H_i(t)} + \frac{\alpha \gamma_i}{\rho_i} \frac{\left[1 + c_i \tau_i(t)\right]}{H_i(t)}.$$

Inserting this condition at lead $t+1$ in condition (23), we obtain the Euler equation for $\tau_i(\cdot)$

$$\gamma_i \left(1 + c_i \tau_i(t)\right) = \beta \left[\rho_i + \alpha E_t \gamma_i \left(1 + c\tau_i(t+1)\right)\right],$$

which can written, denoting $m_i(t) = \gamma_i \left(1 + c_i \tau_i(t)\right)$, as:

$$m_i(t) = \beta \left[\rho_i + \alpha E_t m_i(t+1)\right]. \tag{25}$$

For $t = T$, condition (24) writes more simply as:

$$\frac{\partial W_T}{\partial H_i(T)} = \frac{\kappa}{H_i(T)},$$

so that condition (25) at time $T-1$ becomes:

$$m_i(T-1) = \beta \rho_i \kappa.$$

We can solve forward equation (25):

$$m_i(t) = \beta \rho_i \sum_{j=1}^{T-t-1} (\alpha\beta)^j + \frac{(\alpha\beta)^{T-t}}{\alpha} \rho_i \kappa$$

so that:

$$\gamma_i(1 + c_i \tau_i(t)) = \rho_i a_t \tag{26}$$

with

$$a_t = \frac{(\alpha\beta)^{T-t}}{\alpha}\kappa + \beta \frac{1-(\alpha\beta)^{T-t-1}}{1-\alpha\beta}$$

and therefore:

$$\tau_i(t) = \frac{1}{c_i}\{\frac{\rho_i}{\gamma} a_t - 1\} \quad \forall t < T \tag{27}$$

Moreover, the stock of human capital in period t depends on previous investment choices. Using lower case letter to denote log variables (ie: $h_i(t) = \log H_i(t)$):

$$\begin{aligned}
h_i(t+1) &= \alpha^{t-s} h_i(s) + \rho_i \sum_{l=s}^{t} \alpha^{t-l} \tau_i(l) - \sum_{l=s}^{t} \alpha^{t-l} \lambda_i(l) \quad \text{for } t > s. \\
&= \alpha^{t-s} h_i(s) + \rho_i \sum_{l=s}^{t} \alpha^{t-l} \left[\frac{1}{c_i} \left(\frac{\rho_i}{\gamma_i} \{ \frac{(\alpha\beta)^{T-l}}{\alpha}\kappa + \beta \frac{1-(\alpha\beta)^{T-l-1}}{1-\alpha\beta} \} - 1 \right) \lambda_i(l) \right] \\
&= \alpha^{t-s} h_i(s) - \sum_{l=s}^{t} \alpha^{t-l} \lambda_i(l) + \frac{\rho_i^2}{c_i} \sum_{l=s}^{t} \alpha^{t-l} \left[\frac{(\alpha\beta)^{T-l}}{\alpha}\kappa + \beta \frac{1-(\alpha\beta)^{T-l-1}}{1-\alpha\beta}) \right] \\
&- \frac{\rho_i}{c_i} \left(\frac{\alpha^{s-t-1} - 1}{\alpha - 1} \right) \alpha^{t+1-s}
\end{aligned}$$

Since $\log y_i(t) = \delta(t) + h_i(t)$ we have

$$\begin{aligned}
\log y_i(t) = \delta_t - \sum_{l=s}^{t} \alpha^{t-l} \lambda_i(l) + \alpha^{t-s} h_i(s) + \frac{\rho_i}{c_i} \left(\frac{\alpha^{s-t-1} - 1}{\alpha - 1} \right) \alpha^{t+1-s} \\
+ \frac{\rho_i^2}{c_i} \sum_{l=s}^{t} \alpha^{t-l} \left[\frac{(\alpha\beta)^{T-l}}{\alpha}\kappa + \beta \frac{1-(\alpha\beta)^{T-l-1}}{1-\alpha\beta}) \right]
\end{aligned}$$

7 Conclusion

In this paper, we proposed a structural model of human capital investments that leads to a three factor model describing unobserved heterogeneity components of an earning equation. Using a long panel on a single cohort of wage earners in France from 1977 to 2007, we estimated the reduced form parameters by random effect maximum likelihood methods that deliver the covariance matrix of the random effects. Some direct counterfactual experiments can be evaluated using these estimations. Moreover, we also constructed estimates of the factors which are biased by an order of $1/T$ and assess their degree of accuracy. This is what allows us to evaluate the relevance of structural restrictions and construct constrained estimators. We can then derive the estimates of the structural components in the original model in terms of returns and post-retirement returns to investments. This allows us to compute richer counterfactuals than the ones that are directly available through variances.

REFERENCES

Abowd, J. and Card, D. (1989). "On the covariance structure of earnings and hours changes", *Econometrica* 57(2): 411-445.

Altonji, J., Smith, A. and Vidangos, I. (2009). "Modeling earnings dynamics", Cowles Foundation, 2009-08.

Alvarez, J. and M., Arellano, 2003, "The Time Series and Cross Section Asymptotics of Dynamic Panel Data Estimators", *Econometrica*, 71:1121-1159.

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

Alvarez, J., Browning, M. and Ejrnaes, M. (2010). "Modelling income processes with lots of heterogeneity", forthcoming *Review of Economic Studies*

Arellano M., and S., Bonhomme, 2010, "Identifying Distributional Characteristics in Random Coefficient Panel Data Models", unpublished manuscript.

Arellano M., and O., Bover, 1995, "Another Look at the Instrumental-Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29-51.

Bai, J., 2009, Likelihood Approach to Small T Dynamic Panel Models with Interactive Effects", unpublished manuscript

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

Baker, M. and Solon, G. (2003). Earnings dynamics and inequality among Canadian men, 1976-1992: Evidence fom lngitudinal income tax records, Journal of Labor Economics 21(2): 267-288.

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.

Blundell, R., and S., Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics*, 87, 115-143.

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.

Bonhomme, S., and J.M., Robin, 2010, "Generalized Nonparametric Deconvolution with an Application to Earnings Dynamics," *Review of Economic Studies,* 77(2), 491-533.

Cappellari, L. (2004). The dynamics and inequality of italian men's earnings: Long-term changes or transitory fluctuations?, *The Journal of Human Resources,* 39(2):475-499.

Card., D., (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica,* 69(5):1127-60.

Cunha,F., J.J. Heckman and S.M., Schennach, 2010, "Estimating the Technology of Cognitive and Noncognitive Skill Formation", *Econometrica,* 78(3):883-931.

Dhaene, G. and K., Jochmans, 2009, "An Adjusted Profile Likelihood for Nonstationary Panel Data Models with Fixed Effects", unpublished manuscript.

Dickens, R. (2000). "The evolution of individual male earnings in Great-Britain: 1975-95", *The Economic Journal* 110(January): 27-49.

Farber H.S. and R., Gibbons, 1996, **"**Learning and Wage Dynamics", *The Quarterly Journal of Economics,* 111(4):1007-1047

Geweke, J. and Keane, M. (2000). "An empirical analysis of earnings dynamics among men in the PSID: 1968-1989", *Journal of Econometrics,* 96: 293-356.

Gouriéroux, C., and A. Monfort (1995): *Statistics and Econometric Models*, Volumes I and II, Cambridge University Press, Cambridge.

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.

Hahn, J., and G., Kuersteiner, (2002). Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T Are Large, *Econometrica*, 70(4), 1639–1657

Han, C., P. C.B. Phillips and D. Sul, 2010, "X-Differencing and Dynamic Panel Model Estimation", Cowles Foundation WP 1747.

Hause, J. (1980). "The fine structure of earnings and the on-the-job training hypothesis", *Econometrica,* 48(4):1013-1029.

Heckman, J. J., Lochner, L. L. and Todd, P. E. (2006). Earnings functions, rates of return and treatment e ects: the mincer equation and beyond, in E. A. Hanushek and F. Welch (eds), *Handbook of Economics of Education*, Elsevier Science, Amsterdam.

Hirano, K. (2002). "Semiparametric bayesian inference in autoregressive panel data models", *Econometrica* 70(2): 35-45.

Hoffman, F. (2010). "An empirical model of life cycle earnings and mobility dynamics". WP Toronto.

Horowitz, J. and M., Markatou (1996). "Semiparametric estimation of regression models for panel data", *Review of Economic Studies,* 63:145-68.

Hospido, L. (2010). "Modelling heterogeneity and dynamics in the volatility in individual wages", forthcoming *Journal of Applied Econometrics.*

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002). "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods", *Journal of Econometrics*, 109, 107–150

Kalwij, A. s. and Alessie, R. (2007). "Permanent and transitory wages of British men 1975-2001: Year, age and cohort effects", *Journal of Applied Econometrics,* 22: 1063-1093.

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.

Lillard, L. A. and Weiss, Y. (1979). "Components of variation in panel earnings data : American scientists 1960-70", *Econometrica* 47(2): 437-454.

Lillard, L. A. and Willis, R. J. (1978). Dynamic aspects of earnings mobility, *Econometrica* 46(5):985-1012.

MaCurdy, T. (1982). "The use of time series processes to model the error structure of earnings in a longitudinal data analysis", *Journal of Econometrics* 18: 83-114.

Mallows, C. (2007): "Deconvolution by Simulation," in: Liu, R., Strawderman, W., and C.H. Zhang (Eds.), *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*, Beachwood, Ohio, USA: Institute of Mathematical Statistics.

Magnac, T. and Roux, S. (2009). "Dynamique des salaires dans une cohorte", *Economie et prévision,* 187:1-24.

Meghir, C. and Pistaferri, L. (2004). "Income variance dynamics and heterogeneity", *Econometrica,* 72:1-32.

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.

Mincer, J. (1974). *Schooling experience and earnings,* NBER.

Moffitt, R. A. and Gottschalk, P. (1995). "Trends in the autocovariance structure of earnings in the U.S.: 1969-1987". WP Brown University.

Moffitt, R. A. and Gottschalk, P. (2002). "Trends in the transitory variance of earnings in the united states", *The Economic Journal* 112 (March): C68-C73.

Moffitt, R. A. and Gottschalk, P. (2008). "Trends in the transitory variance of male earnings in the us, 1970-2004", unpublished manuscript.

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

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.

Press, W.H., S.A., Teukolsky, W.T., Vetterling and B.P., Flannery (1992), *Numerical Recipes in Fortan 77: The Art of Scientific Computing,* Cambridge UP.

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.

Schennach, S., 2004, "Estimation of Nonlinear Models with Measurement Error", *Econometrica,* 72:33-75.

Sologon, D. M. and O'Donoghue, C. (2009). "Earnings dynamics and inequality among men across 14 EU countries 1994-2001: Evidence from ECHP", unpublished manuscript.

APPENDICES

A 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 time variation. Equation (22) writes

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

where $y_i^{[1,T]} = (y_{i1}, ..., y_{iT})'$, $v_i^{[1,T]} = (v_{i1}, ..., v_{iT})'$, $\eta_i = (\eta_{i1}, \eta_{i2}, \eta_{i3})$ 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 some initial conditions, the number of which depends on the order of the autoregressive process. Denote p this order and write the initial conditions as:

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

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

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

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 is directly put on earnings y_{it} and in the absence of covariates, this is equivalent to specifying it in the residual part v_{it} because there is a single individual effect. This equivalence still holds when another heterogeneity factor interacted with a linear ttrend is present. Nevertheless, our specification includes a third factor interatced with a geometric term and this breaks the equivalence. To circumvent this problem, we posit that v_{it} is a (time heteroskedastic) ARMA process whose innovations are independent of the individual heterogeneity terms, η_i. As a consequence, our variable of interest, y_{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 supposed to be independent between them although they are both correlated with the initial conditions, $y_i^{[1-p,0]}$.

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

A.1 Time heteroskedastic ARMA specification

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

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

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

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

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

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

As $v_i^{[1-p,T]} = \left(v_{i(1-p)}, ..., v_{iT} \right)$, 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:

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

To compute the covariance of $v_i^{[1-p,T]}$, we derive the covariance matrix of $\begin{pmatrix} v_i^{[1-p,0]} & \zeta_i^{[1-q,T]} \end{pmatrix}$. 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 East corner is assumed to be an unrestricted covariance matrix $Vy_i^{[1-p,0]} = \Gamma_{00}$. Assuming as usual that $E(y_{i\tau}\zeta_{it}) = 0$ for any $\tau < t$, we have that $E(v_i^{[1-p,0]}.(\zeta_i^{[1,T]})') = 0$. Only $E(y_i^{[1-p,0]}.(\zeta_i^{[1-q,0]})')$ remains to be defined:

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

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_i^s is drawn after r and is supposed to be not correlated with y_i^r.

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

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

A.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 from the whole innovation process $\zeta_i^{[1-q,T]}$. As for the covariance structure between initial conditions and those factors, we assume that:

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

Consider the covariance matrix of initrial conditions Σ :

$$\Sigma = V \left(\begin{array}{c} v_i^{[1-p,0]} \\ \eta_i \\ \zeta_i^{[1-q,0]} \end{array} \right) = \left(\begin{array}{ccc} \Gamma_{00} & \Gamma_{0\eta} & \Omega \\ \Gamma_{0\eta}' & \Sigma_\eta & 0 \\ \Omega & 0 & I_q \end{array} \right).$$

and define,

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

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

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

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

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

and

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

Hence,

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

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 does not appear in the correlation between the two processes since innovations are supposed to be independent with individual heterogeneity factors. Initial conditions are given by $\zeta_i^{[1-q,0]}$, η and $v_i^{[1-p,0]}$.

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

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

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

B Construction of Counterfactuals

B.1 Estimates of individual factors given observed wages

The main equation is:

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

where η_i and $v_i^{[1-p,T]}$ are supposed to be centered by construction and where a row of $M(\beta)$ is defined as $M(\beta)^{[t]} = (1, t, 1/\beta^t)$. Later on, we shall reintroduce the averages 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:

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

We continue by looking at the decomposition of the residuals only:

$$u_i^{[1-p,T]} = y_i^{[1-p,T]} - \bar{y}_{g\ni i}^{[1-p,T]} = M(\beta)^{[0,T+p]}\eta_i + v_i^{[1-p,T]}$$

We consider first the case with no missing values and extend it to the case with missing values. We finally analyze how to deal in the simulations with constraints on η_i.

B.2 No missing values

To deal with the correlation between η_i and v_i, we can always write:

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

where $E(\eta_i' w_i^{[1-p,T]}) = 0$. We strengthen the restriction into:

$$w_i^{[1-p,T]} \mid \eta_i \underset{n\to\infty}{\overset{d}{\rightsquigarrow}} N(0, \Omega_w),$$

which is independent of η_i. Given the absence of correlation between η_i and $w_i^{[1-p,T]}$ we get:

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

and:

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

Writing:

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

define the conditional likelihood function as:

$$L(u_i^{[1-p,T]} \mid \eta_i) = \frac{1}{(2\pi)^{T/2} \det \Omega_v^{1/2}} \exp\left(-\frac{1}{2}(u_i^{[1-p,T]} - D\eta_i)'\Omega_w^{-1}(u_i^{[1-p,T]} - D\eta_i)\right).$$

We are seeking the conditional distribution of η_i conditional on the observed $u_i^{[1-p,T]}$ which can be expressed by Bayes law, using a prior for η_i, $L_0(\eta_i)$ as:

$$L(\eta_i \mid u_i^{[1-p,T]}) = \frac{L(u_i^{[1-p,T]} \mid \eta_i)L_0(\eta_i)}{\int L(u_i^{[1-p,T]} \mid \eta_i)L_0(\eta_i)d\eta_i}.$$

Consequently, the distribution function $L(\eta_i \mid u_i^{[1-p,T]})$ is equal to:

$$H(u_i^{[1-p,T]}).\exp\left(-\frac{1}{2}(\eta_i - Bu_i^{[1-p,T]})'\Omega_\eta^{-1}(\eta_i - Bu_i^{[1-p,T]})\right) L_0(\eta_i)$$

where the constant of integration is derived by setting to one the integral over η_i. In the case of a diffuse prior i.e. $L_0(\eta_i) = 1$, the constant of integration is no longer dependent on $u_i^{[1-p,T]}$ and is equal to the usual reciprocal of $(2\pi)^{3/2}\det\Omega_\eta^{1/2}$.

As all terms in η_i and $u_i^{[1-p,T]}$ are quadratic, we can derive the unknown matrices B and Ω_η by solving:

$$(u_i^{[1-p,T]} - D\eta_i)'\Omega_w^{-1}(u_i^{[1-p,T]} - D\eta_i) = (\eta_i - Bu_i^{[1-p,T]})'\Omega_\eta^{-1}(\eta_i - Bu_i^{[1-p,T]}) + u_i^{[1-p,T]\prime}Au_i^{[1-p,T]}.$$

By identifying quadratic terms in (η_i, η_i), $(u_i^{[1-p,T]}, \eta_i)$ and $(u_i^{[1-p,T]}, u_i^{[1-p,T]})$, we obtain three equations:

$$\begin{cases} D'\Omega_w^{-1}D &= \Omega_\eta^{-1}, \\ -D'\Omega_w^{-1} &= -\Omega_\eta^{-1}B, \\ \Omega_w^{-1} &= B'\Omega_\eta^{-1}B + A, \end{cases}$$

so that, as $D'\Omega_w^{-1}D$ is invertible:

$$\begin{cases} \Omega_\eta &= (D'\Omega_w^{-1}D)^{-1}, \\ B &= (D'\Omega_w^{-1}D)^{-1}D'\Omega_w^{-1}, \\ A &= \Omega_w^{-1} - \Omega_w^{-1}D(D'\Omega_w^{-1}D)^{-1}D'\Omega_w^{-1}. \end{cases}$$

The unconstrained estimator for the individual fixed effects, by reinclusion of the estimated averages, are:

$$\hat{\eta}_i = \bar{\eta}_g + Bu_i^{[1-p,T]} = \bar{\eta}_g + B(D\eta_i + w_i^{[1-p,T]}) = \bar{\eta}_{g\ni i} + \eta_i + Bw_i^{[1-p,T]}.$$

They are such that:

$$\begin{aligned} V(\hat{\eta}_i - \bar{\eta}_{g\ni i}) \quad &= \quad EV(\hat{\eta}_i - \bar{\eta}_{g\ni i} \mid \eta_i) + VE(\hat{\eta}_i - \bar{\eta}_{g\ni i} \mid \eta_i) \\ &\Longrightarrow \quad V(\hat{\eta}_i - \bar{\eta}_{g\ni i}) = B\Omega_w B^T + V\eta_i = \Omega_\eta + V\eta_i. \end{aligned}$$

The term Ω_η goes to zero at least at the rate $1/T$ since matrix D is determined by different factors. Some are going to zero faster than T but they are dominated by the simple factors.

We now analyse the case with missing values.

B.3 Missing values

Suppose that $u_i^{[1-p,T]}$ is not observable, only $M_i u_i^{[1-p,T]}$ is where M_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, the distribution function $L(\eta_i \mid M_i u_i^{[1-p,T]})$ becomes:

$$H_i(M_i u_i^{[1-p,T]}).\exp\left(-\frac{1}{2}(\eta_i - B_i M_i u_i^{[1-p,T]})'\Omega_{\eta i}^{-1}(\eta_i - B_i M_i u_i^{[1-p,T]})\right) L_0(\eta_i),$$

where by simple analogy to the results of the previous section:

$$\begin{cases} \Omega_{\eta i} & = (D'M_i'(M_i\Omega_w M_i')^{-1}M_iD)^{-1}, \\ B_i & = (D'M_i'(M_i\Omega_w M_i')^{-1}M_iD)^{-1}D'M_i'(M_i\Omega_w M_i')^{-1}. \end{cases}$$

In all cases, denote C the Choleski decomposition of the permutation of matrix Ω_η (or $\Omega_{\eta i}$ in the case of missing values) such that:

$$CC' = \Omega_\eta$$

so that we can write, assuming that the expectation of each η_{ij} is $\alpha_{j.}$:

$$\begin{cases} \eta_2 = & c_{11}\xi_1, \\ \eta_3 = & c_{21}\xi_1 + c_{22}\xi_2, \\ \eta_1 = & c_{31}\xi_1 + c_{32}\xi_2 + c_{33}\xi_3. \end{cases}$$

B.4 Constrained estimator

Using that the likelihood function $L(\eta_i \mid u_i^{[1-p,T]})$ is proportional to:

$$\exp\left(-\frac{1}{2}(\eta_i - \hat{\eta}_i)'\Omega_\eta^{-1}(\eta_i - \hat{\eta}_i)\right) L_0(\eta_i)$$

where $\hat{\eta}_i$ is the unconstrained estimator, we solve the following program to compute the constrained estimator of η

$$\min_{\eta_i}(\eta_i - \hat{\eta}_i)'\Omega_\eta^{-1}(\eta_i - \hat{\eta}_i)$$

under the constraints:

$$\eta_{i2} > 0, \eta_{i3} < 0, \eta_{i3} > -\lambda\eta_{i2}.$$

Denote μ_1, μ_2 and μ_3 the Lagrange multipliers associated to each constraint and write the Lagrangian as:

$$L(\eta_i) = (\eta_i - \hat{\eta}_i)'\Omega_\eta^{-1}(\eta_i - \hat{\eta}_i) - \mu_1\eta_{i2} + \mu_2\eta_{i3} - \mu_3(\eta_{i3} + \lambda\eta_{i2}).$$

Taking derivatives yields:

$$2\Omega_\eta^{-1}(\tilde{\eta}_i - \hat{\eta}_i) - \begin{pmatrix} 0 \\ \mu_1 + \lambda\mu_3 \\ \mu_3 - \mu_2 \end{pmatrix} = 0.$$

We immediately have that:

1. If $\mu_2 > 0, \mu_1 = 0$ then $\tilde{\eta}_{i3} = 0$ and $\tilde{\eta}_{i2} > 0$, and this implies that $\lambda\tilde{\eta}_{i2} + \tilde{\eta}_{i3} > 0$ so that $\mu_3 = 0$. Therefore:

$$\begin{pmatrix} \tilde{\eta}_{i1} - \hat{\eta}_{i1} \\ \tilde{\eta}_{i2} - \hat{\eta}_{i2} \\ -\hat{\eta}_{i3} \end{pmatrix} + \frac{\Omega_\eta}{2}\begin{pmatrix} 0 \\ 0 \\ \mu_2 \end{pmatrix} = 0 \Longrightarrow \mu_2 e_3^T \frac{\Omega_\eta}{2} e_3 = \hat{\eta}_{i3},$$

where $e_3 = (0,0,1)^T$. This is compatible if $\mu_2 = \frac{\hat{\eta}_{i3}}{e^T \frac{\Omega_\eta}{2} e} > 0$ and therefore if $\hat{\eta}_{i3} > 0$ since Ω_η is definite positive. Denoting $e_2 = (0,1,0)^T$, we also have:

$$\tilde{\eta}_{i2} - \hat{\eta}_{i2} = -\mu_2 . e_2^T \frac{\Omega_\eta}{2} e_3.$$

This satisfies the condition $\mu_1 = 0$ iff $\tilde{\eta}_{i2} > 0$.

2. If $\mu_3 > 0, \mu_1 = 0$ then $\tilde{\eta}_{i3} = -\lambda \tilde{\eta}_{i2}$ and $\tilde{\eta}_{i2} > 0$, and this implies that $\tilde{\eta}_{i3} < 0$ so that $\mu_2 = 0$. We have:

$$2\Omega_\eta^{-1}(\tilde{\eta}_i - \hat{\eta}_i) - \begin{pmatrix} 0 \\ \lambda \\ 1 \end{pmatrix} \mu_3 = 0 \Longrightarrow (\tilde{\eta}_i - \hat{\eta}_i) = \mu_3 \frac{\Omega_\eta}{2} v_\lambda$$

denoting $v_\lambda = (0, \lambda, 1)^T$. Given that $v_\lambda^T \tilde{\eta}_i = \tilde{\eta}_{i3} + \lambda \tilde{\eta}_{i2} = 0$, this implies that :

$$\mu_3 = -\frac{v_\lambda^T \hat{\eta}_i}{v_\lambda^T \frac{\Omega_\eta}{2} v_\lambda} > 0,$$

if $v_\lambda^T \hat{\eta}_i = \hat{\eta}_{i3} + \lambda \hat{\eta}_{i2} < 0$. This yields the constrained estimators, $\tilde{\eta}_{i2}$ and $\tilde{\eta}_{i3}$:

$$(\tilde{\eta}_i - \hat{\eta}_i) = \mu_3 \frac{\Omega_\eta}{2} v_\lambda^T$$

which satisfy the constraint $\mu_1 = 0$ iff $\tilde{\eta}_{i2} > 0$.

3. If $\mu_1 > 0$ then $\tilde{\eta}_{i2} = 0$ and thus the constraints $\lambda \tilde{\eta}_{i2} +$ $\tilde{\eta}_{i3} \geq 0$ and $\tilde{\eta}_{i3} \leq 0$ imply that $\tilde{\eta}_{i3} = 0$, that $\mu_2 \mu_3 = 0$ and that one of them is positive.

Summarizing:

- If $\hat{\eta}_{i3} > 0, \hat{\eta}_{i3} + \lambda \hat{\eta}_{i2} > 0$ case 1 applies if $\tilde{\eta}_{i2} > 0$.
- If $\hat{\eta}_{i3} + \lambda \hat{\eta}_{i2} < 0$, $\hat{\eta}_{i3} < 0$ case 2 applies if $\tilde{\eta}_{i2} > 0$.
- In all other cases, $\tilde{\eta}_{i2} = \tilde{\eta}_{i3} = 0$. In this case:

$$\tilde{\eta}_i - \hat{\eta}_i = \begin{pmatrix} \tilde{\eta}_{i1} - \hat{\eta}_{i1} \\ -\hat{\eta}_{i2} \\ -\hat{\eta}_{i3} \end{pmatrix} = \frac{\Omega_\eta}{2} \begin{pmatrix} e_2 & e_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

where v_j are unknown. They are obtain using:

$$\begin{pmatrix} e_2^T \\ e_3^T \end{pmatrix} (\tilde{\eta}_i - \hat{\eta}_i) = \begin{pmatrix} e_2^T \\ e_3^T \end{pmatrix} \begin{pmatrix} 0 \\ -\hat{\eta}_{i2} \\ -\hat{\eta}_{i3} \end{pmatrix} = \begin{pmatrix} e_2^T \\ e_3^T \end{pmatrix} \frac{\Omega_\eta}{2} \begin{pmatrix} e_2 & e_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

Denoting $I_c^T = \begin{pmatrix} e_2^T \\ e_3^T \end{pmatrix}$ so that:

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \left[I_c^T \frac{\Omega_\eta}{2} I_c \right]^{-1} I_c^T \begin{pmatrix} 0 \\ -\hat{\eta}_{i2} \\ -\hat{\eta}_{i3} \end{pmatrix}$$

so that we get the vector:

$$\tilde{\eta}_i - \hat{\eta}_i = \Omega_\eta I_c \left[I_c^T \Omega_\eta I_c \right]^{-1} I_c^T \begin{pmatrix} 0 \\ -\hat{\eta}_{i2} \\ -\hat{\eta}_{i3} \end{pmatrix}.$$

B.5 Imposing constraints on simulations

When the prior distribution is complicated, some algorithms like the Metropolis-Hastings algorithm do not require knowledge of the constant of integration (Gouriéroux and Monfort, 1996, Simulation Based Econometric Methods, Oxford, page 58). If the prior distribution is simple i.e. for instance, it imposes inequality constraints on η_i only, we can use Gibbs sampling as explained below.

Known means Assume that we want to impose the constraints that $\eta_{i2}^u > 0$ and that $\eta_{i3}^u < 0$ and $\eta_{i3}^u > -\lambda\eta_{i2}$ where the exponent u stands for uncentered. We assume that α_j is the unconditional expectation of η_{ij}^u. Drawing in a multivariate normal distribution with multiple constraints is not as easy as with a single constraint. We use efficient Gibbs sampling as proposed by Rodriguez-Yam, Davis and Scharf (2004).

It starts from the remark that it is easy to draw in univariate truncated normal distributions conditional to the other variates. For instance, $f(\eta_1^u \mid \eta_2^u, \eta_3^u, \eta_2^u \leq 0, \eta_3^u \in [-\lambda\eta_2^u, 0])$. Second, drawing repetitively in the conditional distributions to construct a Markov chain yields drawings that are distributed according to the joint distribution we are looking for. Furthermore, Rodriguez-Yam, Davis and Scharf (2004) recommends drawing the independent errors ξ_1, ξ_2 and ξ_3 instead of the original variables. For this, we have to rewrite the constraints as (using c_{11}, c_{22} and c_{33} are positive):

$$\begin{array}{r} \xi_1 > -\frac{\alpha_2}{c_{11}}, \\ \xi_2 + \frac{c_{21}}{c_{22}}\xi_1 < -\frac{\alpha_3}{c_{22}}, \\ \xi_2 + \frac{c_{21}+\lambda c_{11}}{c_{22}}\xi_1 > -\frac{\alpha_3+\lambda\alpha_2}{c_{22}}. \end{array} \tag{B.2}$$

The algorithm proceeds by considering initial values (η_2^0, η_3^0) whose construction we detail below. Then from (η_2^k, η_3^k), we construct $(\eta_2^{k+1}, \eta_3^{k+1})$ using:

1. Draw ξ_2^{k+1} in a truncated normal variable, truncated by the bounds $[-\frac{\alpha_3+\lambda\alpha_2}{c_{22}} - \frac{c_{21}+\lambda c_{11}}{c_{22}}\xi_1^k, -\frac{\alpha_3}{c_{22}} - \frac{c_{21}}{c_{22}}\xi_1^k]$ (a non empty interval because of the constraint $\xi_1 > -\frac{\alpha_2}{c_{11}}$).

2. Draw ξ_1^{k+1} in a truncated normal variable, truncated by the bounds $[L_1, L_2]$. There are five cases:

   - If $c_{21} > 0$: $L_1 = \max(-\frac{\alpha_2}{c_{11}}, -\frac{c_{22}}{c_{21}+\lambda c_{11}}(\frac{\alpha_3+\lambda\alpha_2}{c_{22}} + \xi_2^{k+1})); U_1 = -\frac{c_{22}}{c_{21}}(\frac{\alpha_3}{c_{22}} + \xi_2^{k+1})$
   - If $c_{21} = 0: L_1 = \max(-\frac{\alpha_2}{c_{11}}, -\frac{c_{22}}{c_{21}+\lambda c_{11}}(\frac{\alpha_3+\lambda\alpha_2}{c_{22}} + \xi_2^{k+1})), U_1 = +\infty$
   - If $c_{21} \in (-\lambda c_{11}, 0) : L_1 = \max(-\frac{\alpha_2}{c_{11}}, -\frac{c_{22}}{c_{21}}(\frac{\alpha_3}{c_{22}} + \xi_2^{k+1}), -\frac{c_{22}}{c_{21}+\lambda c_{11}}(\frac{\alpha_3+\lambda\alpha_2}{c_{22}} + \xi_2^{k+1})), U_1 = +\infty$

- If $c_{21} = -\lambda c_{11}$: $L_1 = \max(-\frac{\alpha_2}{c_{11}}, -\frac{c_{22}}{c_{21}}(\frac{\alpha_3}{c_{22}} + \xi_2^{k+1})), U_1 = +\infty$
- If $c_{21} < -\lambda c_{11}$: $L_1 = \max(-\frac{\alpha_2}{c_{11}}, -\frac{c_{22}}{c_{21}}(\frac{\alpha_3}{c_{22}} + \xi_2^{k+1}))$, $U_1 = -\frac{c_{22}}{c_{21}+\lambda c_{11}}(\frac{\alpha_3+\lambda\alpha_2}{c_{22}} + \xi_2^{k+1}))$.

Then construct .

When the algorithm is said to have converged to $(\xi_1^\infty, \xi_2^\infty)$ then finish by drawing ξ_3 in a N(0,1) variate since no constraints are binding on η_1. Construct the final values $\eta_2^{k+1} = \alpha_2 + c_{11}\xi_1^\infty$, $\eta_3^{k+1} = \alpha_3 + c_{21}\xi_1^\infty + c_{22}\xi_2^\infty$, $\eta_1^{k+1} = \alpha_1 + c_{31}\xi_1^\infty + c_{32}\xi_2^\infty + c_{33}\xi_3$.

Initial conditions The initial conditions are constructed by neglecting the multivariate aspects of constraints:

- Draw ξ_1^0 in a truncated normal distribution, truncated by the bound $\xi_1^0 > -\frac{\alpha_2}{c_{11}}$. Construct $\eta_2^0 = \alpha_2 + c_{11}\xi_1^0$.
- Draw ξ_2^0 in a truncated normal distribution, truncated by the bound $[-\frac{\alpha_3+\lambda\alpha_2}{c_{22}} - \frac{c_{21}+\lambda c_{11}}{c_{22}}\xi_1^0, -\frac{\alpha_3}{c_{22}} - \frac{c_{21}}{c_{22}}\xi_1^0]$. Construct $\eta_3^0 = \alpha_3 + c_{21}\xi_1^0 + c_{22}\xi_2^0$.
- Draw ξ_3^0 in a normal distribution and construct $\eta_1^0 = \alpha_1 + c_{31}\xi_1^0 + c_{32}\xi_2^0 + c_{33}\xi_3^0$.

These draws satisfy the constraints but they are NOT truncated normally distributed.

Unknown means We now assume that $(\alpha_1, \alpha_2, \alpha_3)$ are unknown. Denote the constraint $\eta_{i2}^u > 0, \eta_{i3}^u < 0, \eta_{i3}^u > -\lambda\eta_{i2}^u$ as $\eta_i \in \mathcal{D}_\alpha$.

Notice that the averages in each group, $\bar{\eta}_{gj}$ for $j = 1, ., 3$ estimate the following quantities, different from the previous α_1, α_2 and α_3:

$$E(\alpha_1 + \eta_{i1} \mid \eta_i \in \mathcal{D}_\alpha), E(\alpha_2 + \eta_{i2} \mid \eta_i \in \mathcal{D}_\alpha) \text{ and } E(\alpha_3 + \eta_{i3} \mid \eta_i \in \mathcal{D}_\alpha).$$

As the distributions of η are unknown we cannot directly compute these quantities.

Nevertheless, we can use the following device:

$$E(\eta_i \mid \eta_i \in \mathcal{D}_\alpha) = E(E(\eta_i \mid u_i^{[1-p,T]}, \eta_i \in \mathcal{D}_\alpha)) = E(E(Bu_i^{[1-p,T]} + \tilde{C}\xi_i \mid \xi_i \in \mathcal{D}_\alpha^\xi) = 0$$

where ξ_i the Choleski factors of η_i and $\mathcal{D}_\alpha^\xi$ are defined above and where $\tilde{C}$ is obtained by permuting C. We can adapt this case to missing values by replacing $u_i^{[1-p,T]}$ by $Mu_i^{[1-p,T]}$.

We can thus solve the sytem of equations:

$$\alpha_g + E(Bu_i^{[1-p,T]}) + \tilde{C}E(\xi_i \mid \xi_i \in \mathcal{D}_{\alpha_g}^\xi) = \eta_g$$

in which $E(Bu_i^{[1-p,T]})$ is equal to zero by construction. We estimate $\hat{\alpha}_g$ using:

$$\hat{\alpha}_g + \frac{1}{n_g}\sum_{i\in g} \tilde{C}E(\xi_i \mid \xi_i \in \mathcal{D}_{\hat{\alpha}_g}^\xi) = \bar{\eta}_g$$

The initial conditions for αs are computed as before by neglecting the multiple constraints on the parameters. Using density and cumulative distribuiton functions for the normal $N(0,1)$, we compute:

$$E(\xi_1 \mid \xi_1 > -\frac{\alpha_2}{c_{11}}) = \frac{\phi(\frac{\alpha_2}{c_{11}})}{\Phi(\frac{\alpha_2}{c_{11}})},$$

and we compute the mean over simulations of ξ_1^0 of:

$$E(\xi_2 \mid \xi_2 \in [-\frac{\alpha_3 + \lambda\alpha_2}{c_{22}} - \frac{c_{21} + \lambda c_{11}}{c_{22}}\xi_1^0, -\frac{\alpha_3}{c_{22}} - \frac{c_{21}}{c_{22}}\xi_1^0]) = \frac{\phi(\frac{\alpha_3}{c_{22}} + \frac{c_{21}}{c_{22}}\xi_1^0) - \phi(\frac{\alpha_3+\lambda\alpha_2}{c_{22}} + \frac{c_{21}+\lambda c_{11}}{c_{22}}\xi_1^0)}{\Phi(\frac{\alpha_3+\lambda\alpha_2}{c_{22}} + \frac{c_{21}+\lambda c_{11}}{c_{22}}\xi_1^0) - \Phi(\frac{\alpha_3}{c_{22}} + \frac{c_{21}}{c_{22}}\xi_1^0)}.$$

B.6 Mallows estimates

Suppose that we know estimates $\hat{\alpha}_g$ of the unconditional distribution of η_i^us as developed above so that we can write:

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

Multiply by $\hat{B}^{-1/2}$ and define ξ_i and ζ_i as:

$$\xi_i = \hat{B}^{-1/2}\eta_i, \hat{\xi}_i = \hat{B}^{-1/2}(\bar{\eta}_g + \hat{\eta}_i - \hat{\alpha}_g),$$

so that:

$$\hat{\xi}_i = \xi_i + w_i$$

where w_i is distributed as $N(0,I)$. We write the constraints on ξ_i as they were written in the system of equations B.2 above.

We then proceed by decomposing Mallows estimation in three steps:

- We draw a normal variate w_{i1} in a $N(0,1)$ and look for rearrangements of $\hat{\xi}_{i1}$ so that $\hat{\xi}_{i1} - w_{i1}$ is approximately orthogonal to w_{i1}.
- We draw a normal variate w_{i2} in a $N(0,1)$ and look for rearrangements of $\hat{\xi}_{i2}$ so that $\hat{\xi}_{i2} - w_{i2} > 0$ and is approximately orthogonal to w_{i2}.
- We draw a normal variate w_{i3} in a $N(0,1)$ and look for rearrangements of $\hat{\xi}_{i3}$ so that $\hat{\xi}_{i3} - w_{i3} < 0$, $\lambda(\hat{\xi}_{i2} - w_{i2}) + \hat{\xi}_{i3} - w_{i3} > 0$ and $\hat{\xi}_{i3} - w_{i3}$ is approximately orthogonal to w_{i3}.

Table 1: Previous literature: Some empirical specifications

	Permanent Component	Transitory Component	Time / Cohort Effect	Note: change in components
Lillard Willis (1978)	Indiv. Effect	AR(1)	N	Permanent compo: 73.1% total variance
Lillard Weiss (1979)	RG	AR(1)	N	Heterogeneity in slope and level
Hause (1980)	RG	AR(1)	N	No decomposition is presented
Macurdy (1982)	RG RW	ARMA(1,2)	N	Random Walk component not rejected
Abowd and Card (1989)	RW	MA(2) diff.	Time effect	Compatible with unit root in level
Lollivier Payen (1990)	RG+quadratic	AR(1)	N	Permanent Component: 80%
Moffit Gottschalk (1995)	RW et RG	ARMA(1,1)	N/A	50% increase in permanent comp.
Baker (1997)	RW et RG	ARMA(1,2)	N	Favors random growth
Moffit Gottschalk (1998)	RW	ARMA(1,1)	Time effect	50% increase in permanent comp.
Lillard Reville (1999)	RG+quadaratic	ARMA(1,2)	N	Favors random growth
Dickens (2000)	RW	ARMA(1,2)	Time effect	50% increase in permanent comp.
Haider (2001)	RG	ARMA(1,1)	Time effect	50% increase in permanent comp.
Moffit Gottschalk (2002)	RW	ARMA(1,1)	Time effect	Increase in the permanent in $70's$
Baker Solon (2003)	RW RG	AR(1)	Time effect	Increase in permanent and transitory comp.
Cappellari (2004)	RG	AR(1)	Time +cohort effect	Increase in permanent comp.
Ramos (2003)	RW et RG	ARMA(1,1)	Time effect	Increase in transitory comp.
Meghir Pistaferri (2004)	ARCH	ARCH	Time effect	Reject presence of unit root
Kawji Alessie (2007)	RW	ARMA(1,1)	Time+cohort effect	Increase in transitory comp.
Moffit Gottschalk (2008)	RW	ARMA(1,1)	Time effect	Increase in transitory comp. in $80's$
Sologon (2009)	RG	ARMA(1,1)	Time+cohort effect	Decrease in transitory comp. in many countr

RG stands for random growth model: $y_{it} = \eta_{i1} + \eta_{i2}t + v_{it}$ RW stands for randow walk model $yit = \eta_{it} + v_{it}$ with $\eta_{it} = \eta_{it-1} + \epsilon_{it}$. N/A: not available.

Table 2: Sample size

	Age of Entry			
	Below 20	20 - 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 3: 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 relative to the full sample

Table 4: 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

Table 5: Autocorrelation matrix of earnings residuals in differences

	1978	1979	1980	1985	1986	1987	1988	1989	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
1979	-.400																						
1980	-.009	-.277																					
1985	-.018	-.016	-.084																				
1986	.003	-.031	.090	-.434																			
1987	.043	.093	-.013	-.058	-.345																		
1988	.004	.035	.011	-.055	-.046	-.299																	
1989	.041	-.036	-.008	.028	-.054	-.020	-.323																
1992	-.053	.060	-.055	-.014	-.006	-.074	-.003	-.039															
1993	-.021	.015	-.019	.007	.013	.048	-.072	.000	-.351														
1994	.018	-.013	.024	.000	-.021	-.013	.003	-.037	-.108	-.385													
1995	.021	-.001	.017	-.027	.029	.038	.001	.032	.043	-.070	-.519												
1996	.012	-.013	-.034	.008	-.020	.000	.036	.046	.029	-.021	.026	-.440											
1997	-.052	.032	-.047	.026	-.046	.006	-.022	-.058	-.007	-.005	-.004	-.019	-.520										
1998	.010	-.010	.052	-.047	.049	-.040	.004	.000	.009	.015	-.031	.036	-.015	-.391									
1999	.056	-.017	-.017	.013	.006	-.013	.040	-.004	.014	-.067	.004	-.020	.003	-.010	-.244								
2000	-.087	.085	-.059	.008	.016	-.014	-.006	-.023	.041	.023	.005	-.042	.022	-.003	-.047	-.420							
2001	.024	-.051	.051	.009	-.082	.044	-.028	.052	-.046	-.018	.032	-.009	-.062	.051	.044	-.013	-.539						
2002	.008	.001	-.037	.027	.010	-.090	.046	-.025	-.019	-.002	-.043	.013	.031	.024	-.028	.005	-.010	-.298					
2003	.005	-.050	.001	.041	-.040	-.108	.001	-.015	.061	-.028	.062	-.025	-.049	.052	-.006	.025	.027	-.010	-.247				
2004	-.036	.068	.008	-.061	.057	.144	-.005	.004	-.047	.013	-.031	.012	.025	-.043	.005	-.024	-.025	.014	-.157	-.705			
2005	.073	-.011	.001	-.021	-.017	.026	-.011	-.010	-.019	.020	.005	.004	-.001	-.009	-.013	.056	.014	-.043	.002	.012	-.227		
2006	-.031	.063	-.042	.009	.035	-.025	.021	-.031	.055	-.014	.034	-.023	-.002	-.031	-.013	-.015	.013	-.028	-.002	.039	-.069	-.375	
2007	-.002	-.022	-.010	-.042	-.003	-.026	.026	-.036	-.016	.079	-.070	.022	.015	-.015	-.035	.035	-.015	.020	.030	-.028	-.006	.053	-.254

Table 6: 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 7: Estimated parameters

	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 8: 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)

Nb of Periods ↓	Quantiles ⟶ ηs ↓	0.05	0.2	0.35	0.5	0.65	0.8	0.95
4,η_1		2.5	2.86	2.98	2.98	3.00	3.03	3.08
4,η_2		(1e-10)	(1e-10)	(0.00619)	(0.0155)	(0.0193)	(0.0519)	(0.142)
4,η_3		[-0.0618]	[-0.0502]	[-0.0278]	[-1.5e-10]	[-1.5e-10]	[-1.2e-10]	[-3e-11]
8		-4.99	-0.994	0.469	2	3.09	5.69	9.78
		(-0.594)	(-0.261)	(-0.122)	(-0.00392)	(0.118)	(0.297)	(0.689)
		[-7.91]	[-3.48]	[-0.912]	[0.345]	[2.07]	[3.81]	[7.94]
12		-1.20	1.05	1.72	2.40	2.85	3.5	5.33
		(-0.377)	(-0.153)	(-0.068)	(0.00608)	(0.078)	(0.165)	(0.368)
		[-3.86]	[-1.66]	[-0.443]	[0.028]	[0.85]	[1.68]	[4.64]
16		1.20	1.97	2.26	2.46	2.85	3.06	4.00
		(-0.161)	(-0.0557)	(0.00112)	(0.0328)	(0.075)	(0.115)	(0.25)
		[-2.29]	[-0.881]	[-0.473]	[-0.160]	[0.143]	[0.589]	[1.63]
20		1.73	2.12	2.32	2.52	2.69	2.99	4.03
		(-0.104)	(-0.0222)	(0.0134)	(0.0361)	(0.0602)	(0.105)	(0.223)
		[-1.6]	[-0.644]	[-0.339]	[-0.175]	[0.051]	[0.334]	[0.989]
24		2.01	2.27	2.37	2.51	2.67	2.88	3.59
		(-0.0522)	(-0.00851)	(0.0175)	(0.0376)	(0.0604)	(0.0876)	(0.169)
		[-1.08]	[-0.493]	[-0.282]	[-0.149]	[-0.0204]	[0.142]	[0.58]
28		2.19	2.37	2.48	2.58	2.69	2.86	3.27
		(-0.0140)	(0.0163)	(0.0301)	(0.0429)	(0.057)	(0.0757)	(0.117)
		[-0.65]	[-0.375]	[-0.263]	[-0.171]	[-0.0894]	[0.00895]	[0.215]

Note: No brackets: η_1, brackets: η_2, square brackets: η_3.

Table 9: Quantiles of the distribution of individual effects: unconstrained estimates

Number of observed periods	$\eta_2 < 0$	$\eta_3 > 0$	$\eta_3 + \lambda\eta_2 < 0$
4	0.6	0.6	0.8
5	0.452	0.595	0.786
6	0.493	0.5	0.824
7	0.471	0.507	0.814
8	0.51	0.551	0.925
9	0.447	0.474	0.855
10	0.478	0.511	0.945
11	0.413	0.446	0.918
12	0.462	0.517	0.86
13	0.351	0.358	0.9
14	0.425	0.473	0.892
15	0.354	0.435	0.826
16	0.337	0.417	0.859
17	0.428	0.482	0.825
18	0.337	0.421	0.8
19	0.388	0.461	0.825
20	0.279	0.369	0.742
21	0.280	0.339	0.732
22	0.183	0.309	0.694
23	0.207	0.323	0.685
24	0.227	0.316	0.662
25	0.182	0.325	0.617
26	0.149	0.246	0.642
27	0.135	0.246	0.589
28	0.100	0.211	0.554

Table 10: Frequencies of violations of restrictions: Unconstrained estimates

Number of observed periods	$\eta_2 < 0$	$\eta_3 > 0$	$\eta_3 + \lambda\eta_2 < 0$
4	0	0	0
5	0	0	0
6	0.0211	0.0141	0.0211
7	0.0143	0.0214	0.0214
8	0.0612	0.0612	0.0068
9	0.0526	0.0658	0.0132
10	0.0824	0.088	0.0165
11	0.0326	0.0380	0.0109
12	0.091	0.105	0.007
13	0.0132	0.0199	0.00662
14	0.0359	0.0359	0.00599
15	0.0559	0.0745	0.00621
16	0.0491	0.0552	0
17	0.0361	0.0663	0
18	0.0316	0.0526	0
19	0.0583	0.0922	0
20	0.0343	0.0472	0
21	0.0354	0.0433	0
22	0.00917	0.0183	0
23	0.0225	0.0275	0
24	0.0173	0.0247	0
25	0.0147	0.022	0
26	0.0123	0.0185	0
27	0.00380	0.0114	0
28	0.00186	0.00372	0

Note: A violation occurs if the interval estimate (95%) does not intersect with the restriction.

Table 11: Frequencies of violations of restrictions: (Favorable) bounds of the unconstrained confidence interval estimates

Nb of Periods ↓	Quantiles ⟶ / ηs ↓	0.05	0.2	0.35	0.5	0.65	0.8	0.95
4,η_1		0.363	1.64	2.07	2.07	2.15	2.26	2.41
4,η_2		(0)	(0)	(0)	(0)	(0.116)	(0.264)	(0.479)
4,η_3		[-0.948]	[-0.857]	[-0.496]	[0]	[0]	[0]	[0]
8		1.34	2.01	2.20	2.35	2.51	2.69	3.1
		(0)	(0)	(0.00803)	(0.0328)	(0.0503)	(0.101)	(0.202)
		[-0.629]	[-0.191]	[0]	[0]	[0]	[0]	[0]
12		1.64	2.05	2.22	2.37	2.48	2.62	3.04
		(0)	(0)	(0.0138)	(0.0253)	(0.041)	(0.0663)	(0.098)
		[-0.39]	[-0.135]	[-0.00481]	[0]	[0]	[0]	[0]
16		1.91	2.21	2.32	2.41	2.57	2.74	3.14
		(0)	(0.00797)	(0.0169)	(0.0263)	(0.0372)	(0.0548)	(0.0899)
		[-0.357]	[-0.195]	[-0.0976]	[0]	[0]	[0]	[0]
20		1.95	2.22	2.34	2.47	2.58	2.77	3.2
		(0)	(0.00867)	(0.0195)	(0.0295)	(0.04)	(0.052)	(0.084)
		[-0.348]	[-0.209]	[-0.128]	[-0.0300]	[0]	[0]	[0]
24		2.1	2.27	2.37	2.48	2.6	2.83	3.37
		(0)	(0.0119)	(0.0204)	(0.0303)	(0.0402)	(0.0535)	(0.0829)
		[-0.343]	[-0.214]	[-0.150]	[-0.0746]	[-0.0013]	[0]	[0]
28		2.2	2.36	2.47	2.56	2.66	2.82	3.17
		(0.00746)	(0.0199)	(0.0275)	(0.0358)	(0.0454)	(0.0583)	(0.084)
		[-0.349]	[-0.235]	[-0.172]	[-0.126]	[-0.066]	[0]	[0]

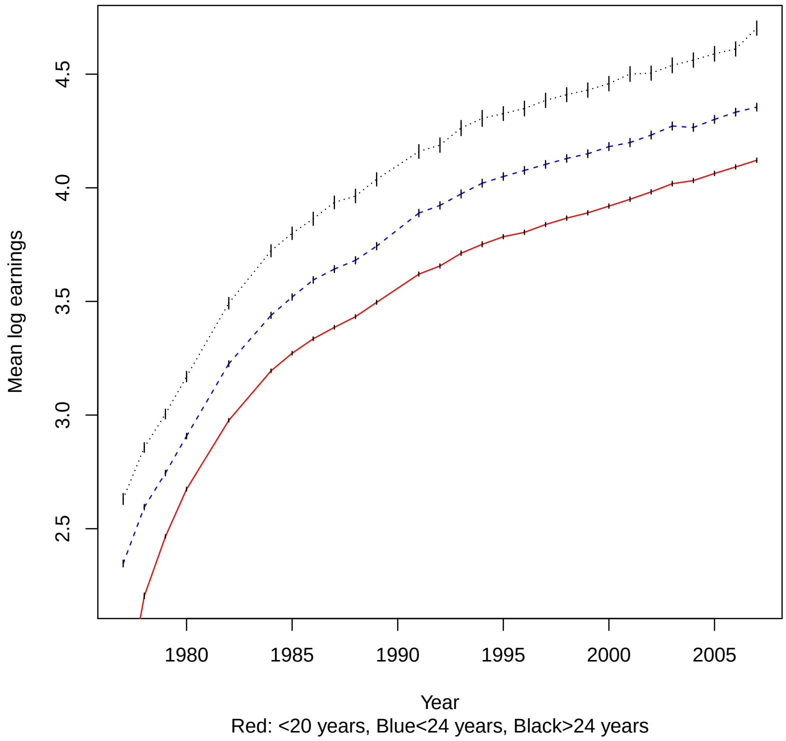
Note: No brackets: η_1, brackets: η_2, square brackets: η_3.

Table 12: Quantiles of the distribution of individual effects: constrained estimates

Quantiles	Observed distance	Simulated distance
0.175	0	0
0.225	0.0021	0.00180
0.275	0.0141	0.0132
0.325	0.0370	0.0391
0.375	0.0763	0.0761
0.425	0.126	0.125
0.475	0.194	0.194
0.525	0.276	0.282
0.575	0.401	0.395
0.625	0.568	0.531
0.675	0.763	0.714
0.725	1.04	0.945
0.775	1.48	1.21
0.825	2.14	1.57
0.875	3.17	2.10
0.925	5.32	2.93
0.975	12.7	4.74

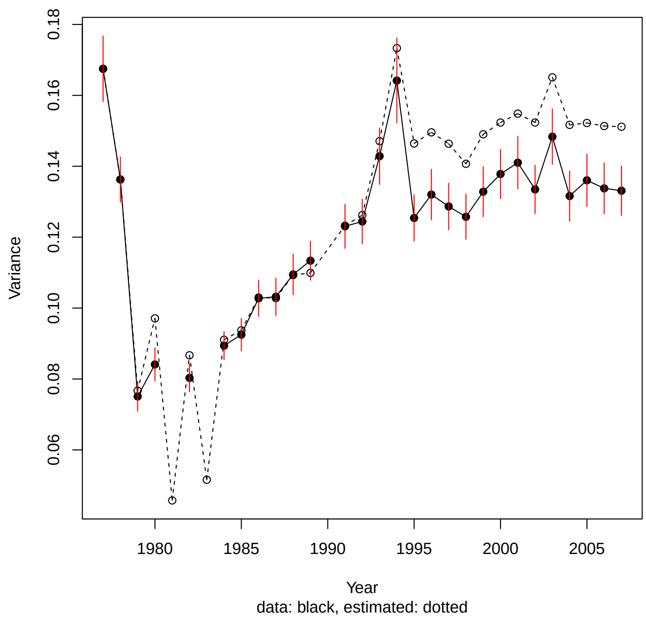
Notes: Distances use as a metric the inverse covariance matrix of ηs. Simulations are performed by adding to the constrained estimates a normal noise and by reprojecting on the constrained set.

Table 13: Distances between unconstrained and constrained estimates for observations and simulations



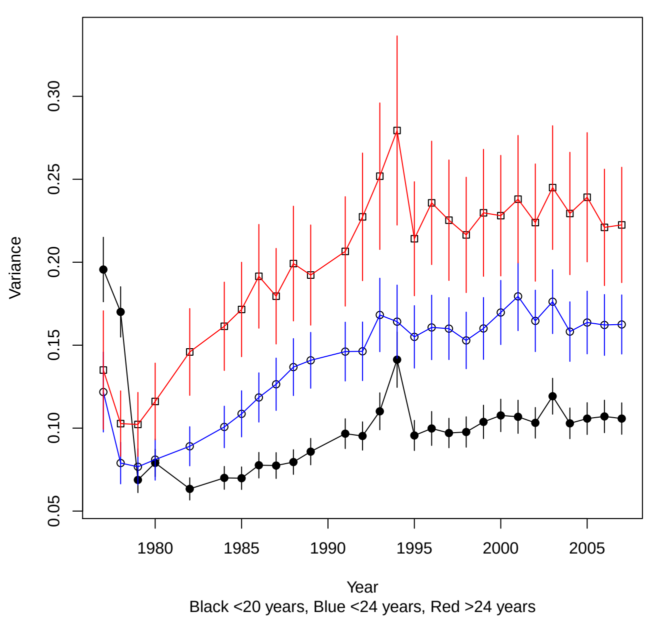
Note: The small lines represent the 95% confidence intervals.

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



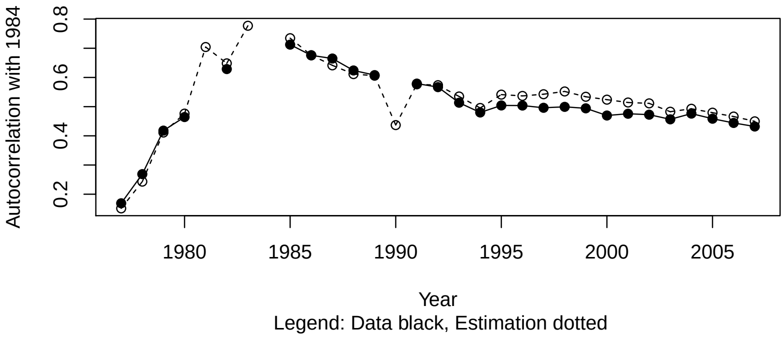
Note: The small lines represent the 95% confidence intervals.

(A) full sample



(B) by age group

Figure 2: Cross-sectional variance of earnings: 1977-2007



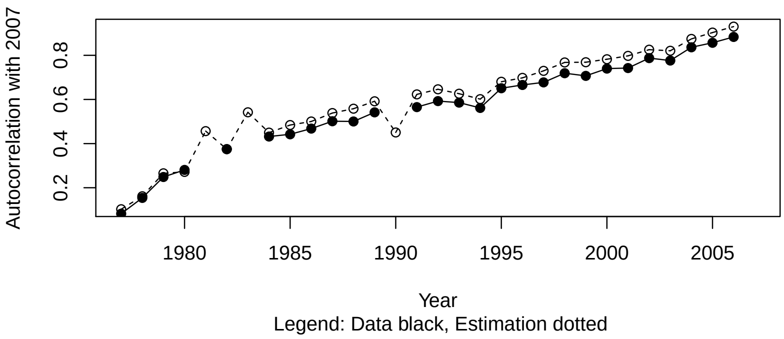
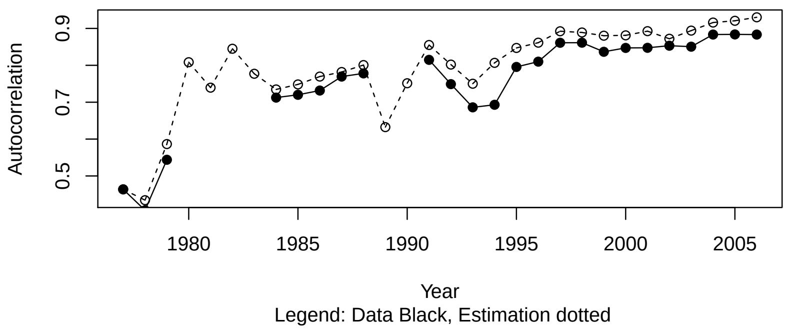


Figure 3: Autocorrelations with 1986 and 2007



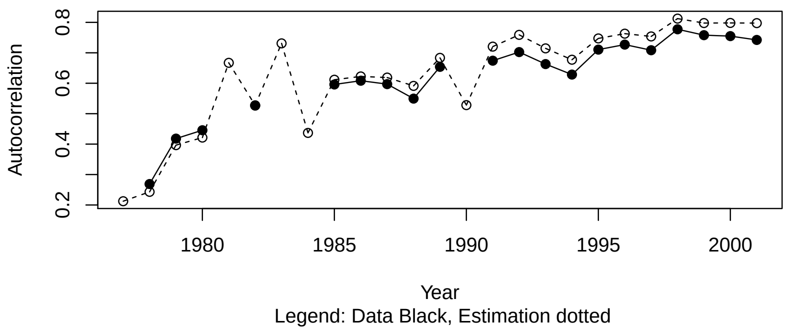


Figure 4: Autocorrelations of order 1 and of order 6

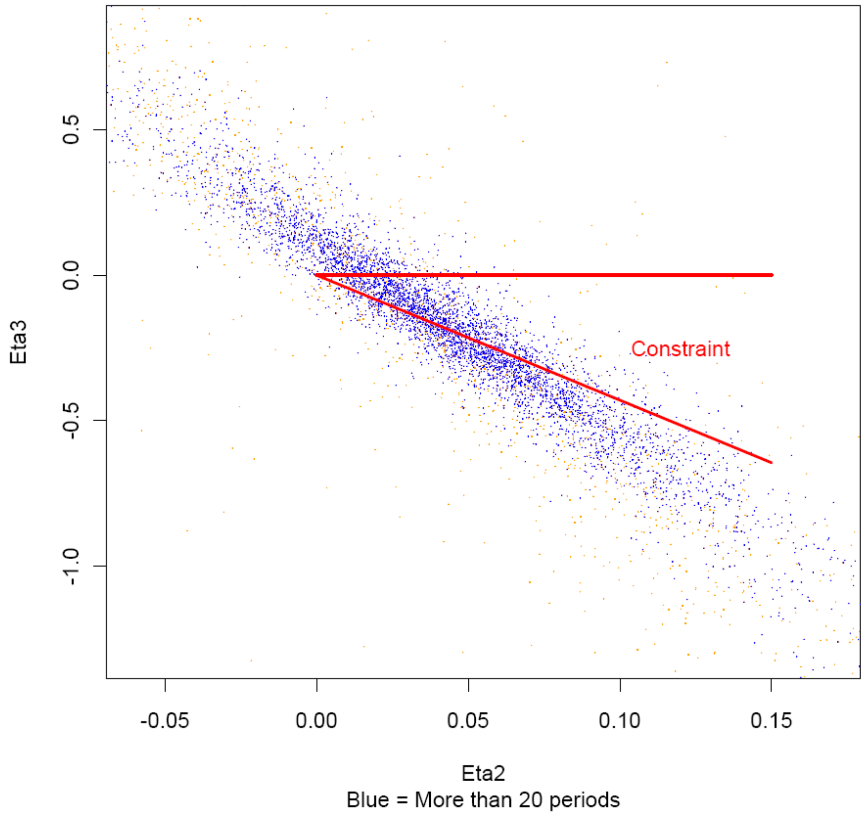


Figure 5: Scatter plot of η_2 and η_3 and the constraint area

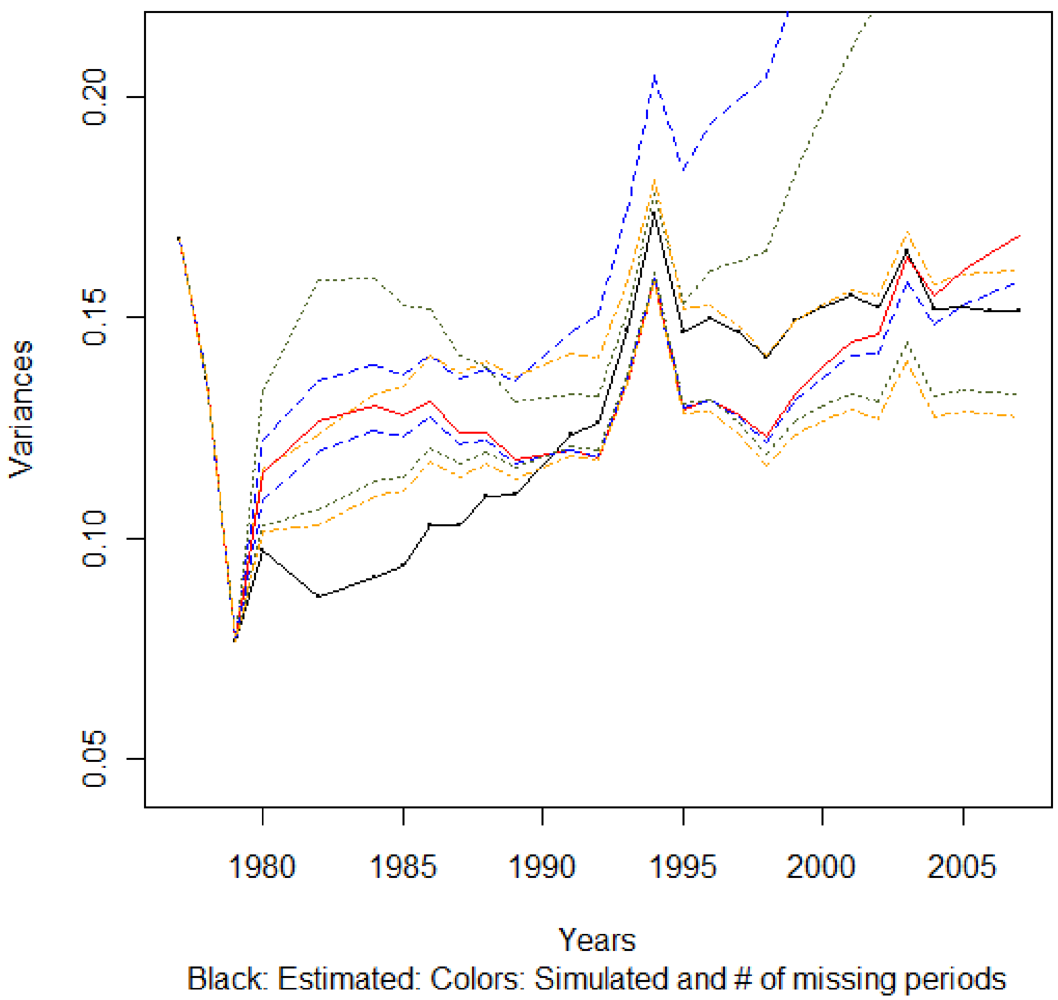


Figure 6: Estimated and simulated variances according to the number of periods of non missing data

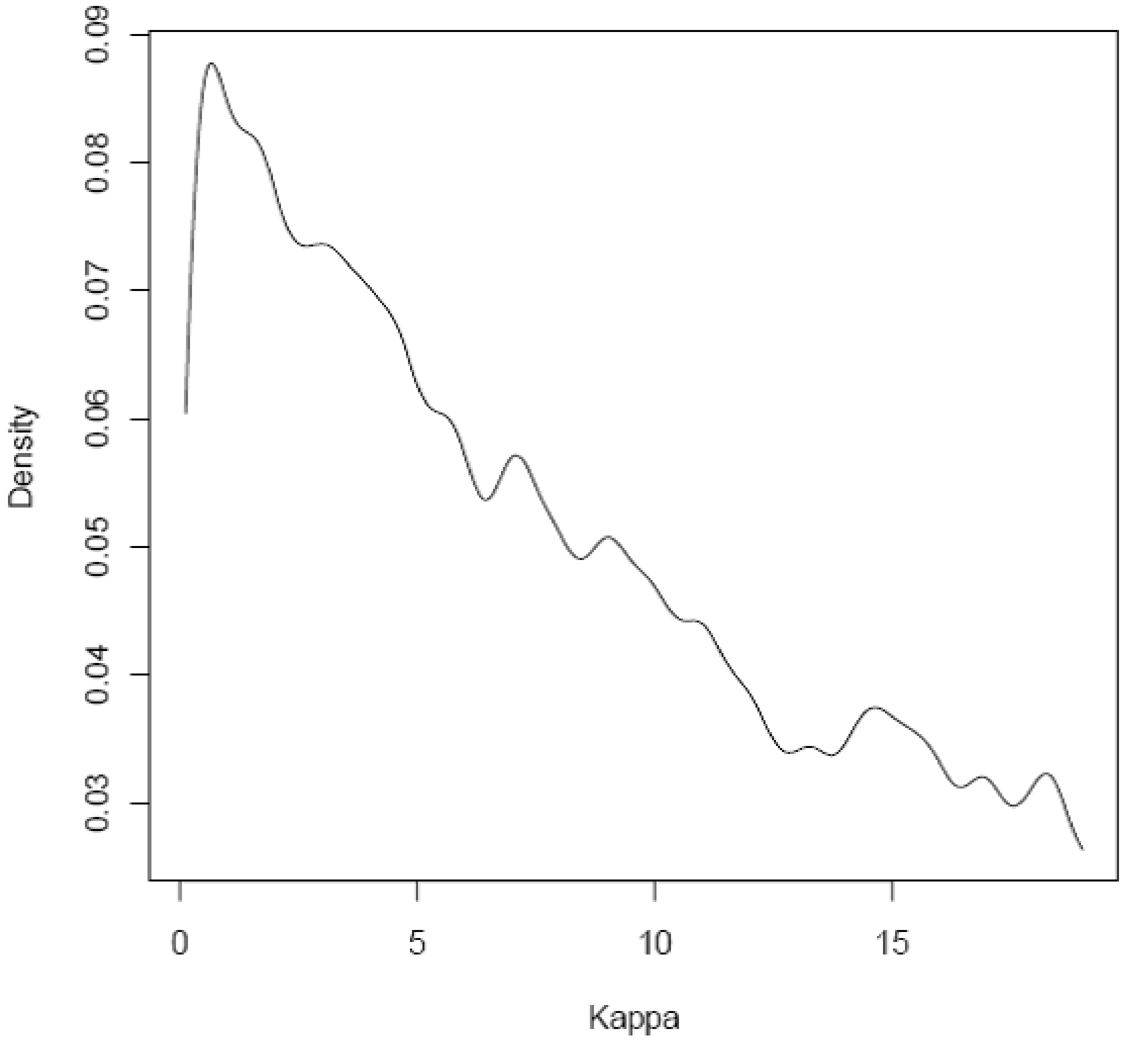


Figure 7: Density function of κ

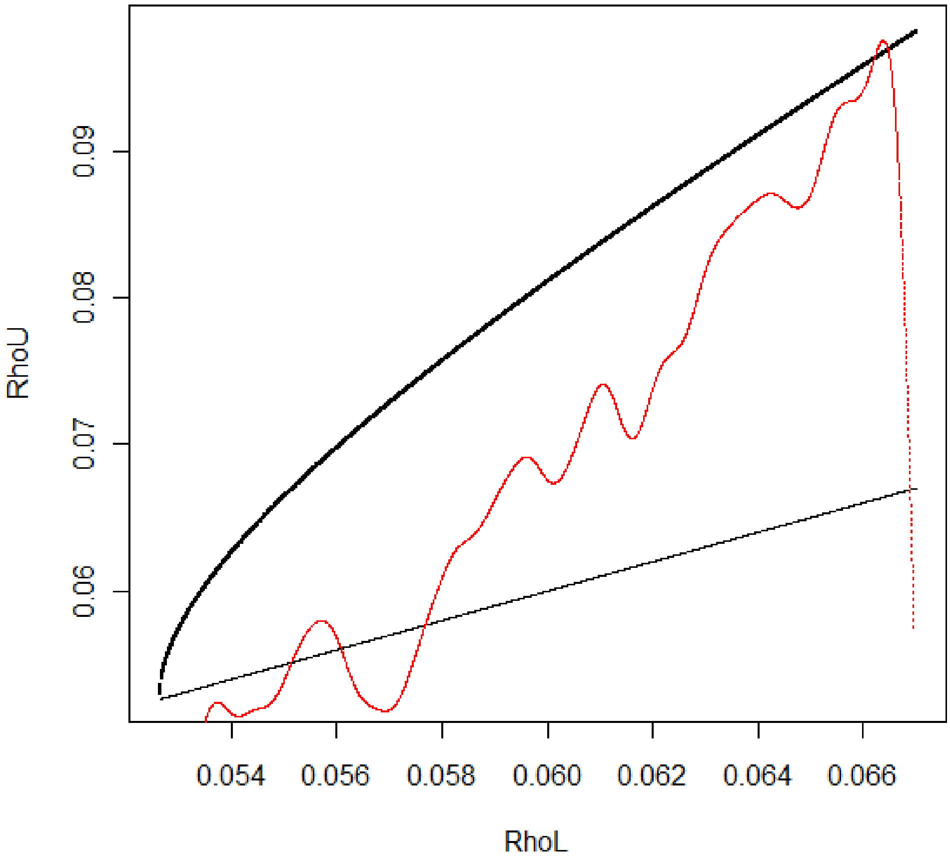
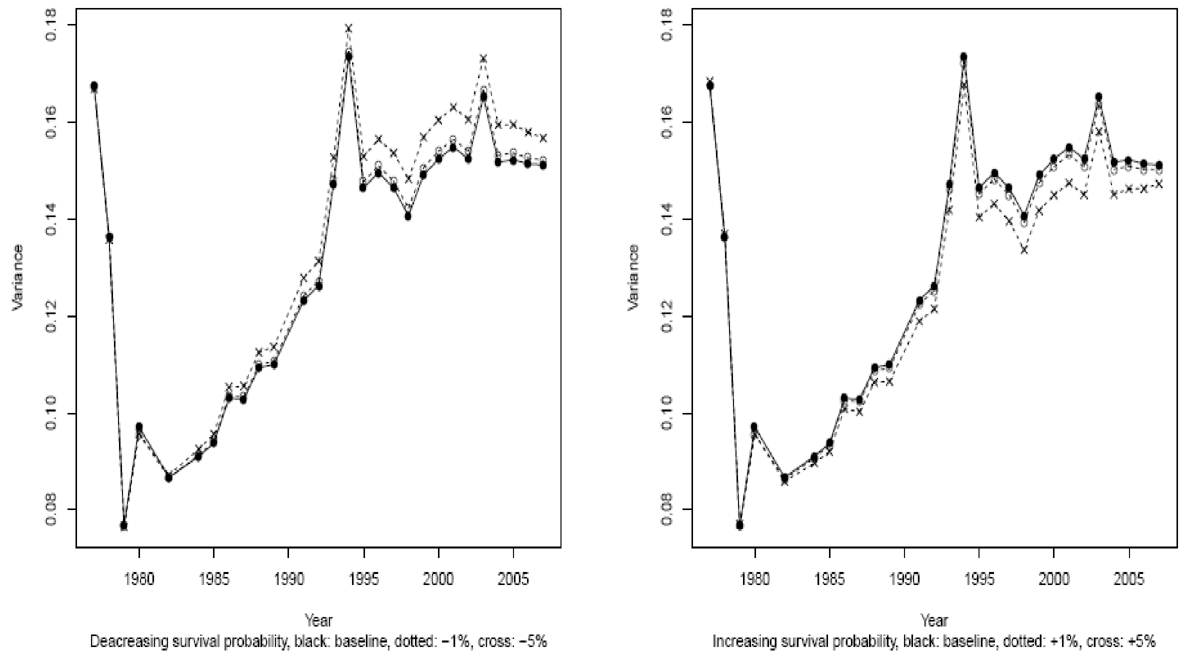
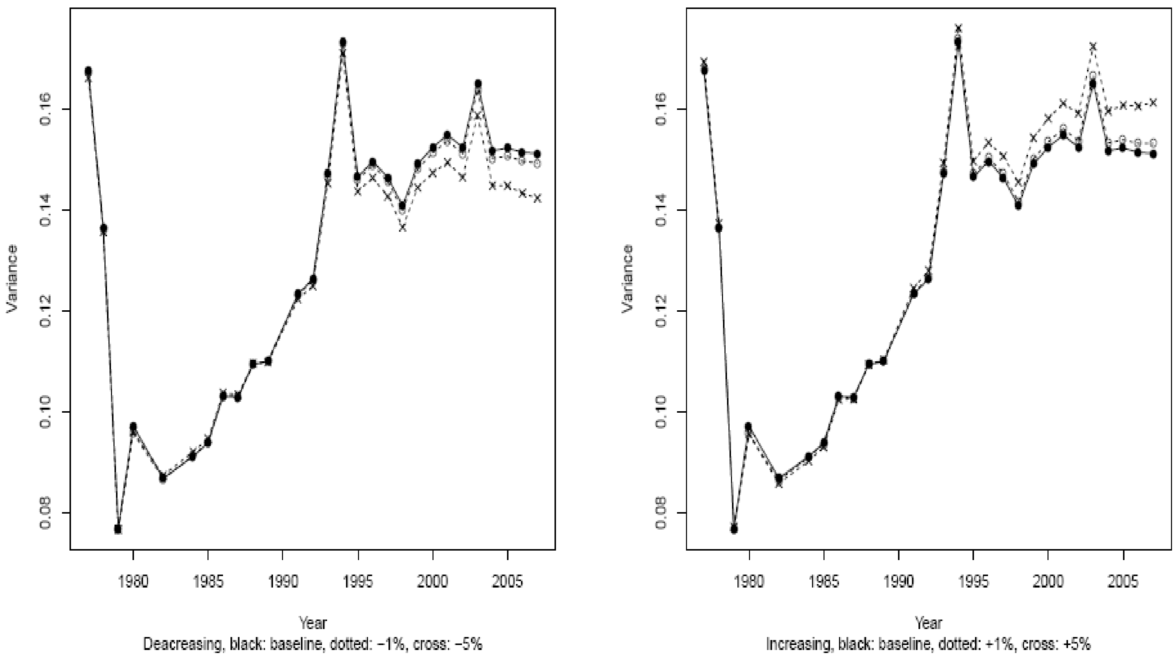


Figure 8: The width of the partially identified set as a function of ρ



(A) Decreasing the variance κ (B) Increasing the variance κ

Figure 9: Counterfactual variance of log earnings: The value of η_3 is divided or multiplied by 1.01 or 1.05.



(A) Decreasing the variance ρ (B) Increasing the variance ρ

Figure 10: Counterfactual variance of log earnings: The value of η_2 is divided or multiplied by 1.01 or 1.05.

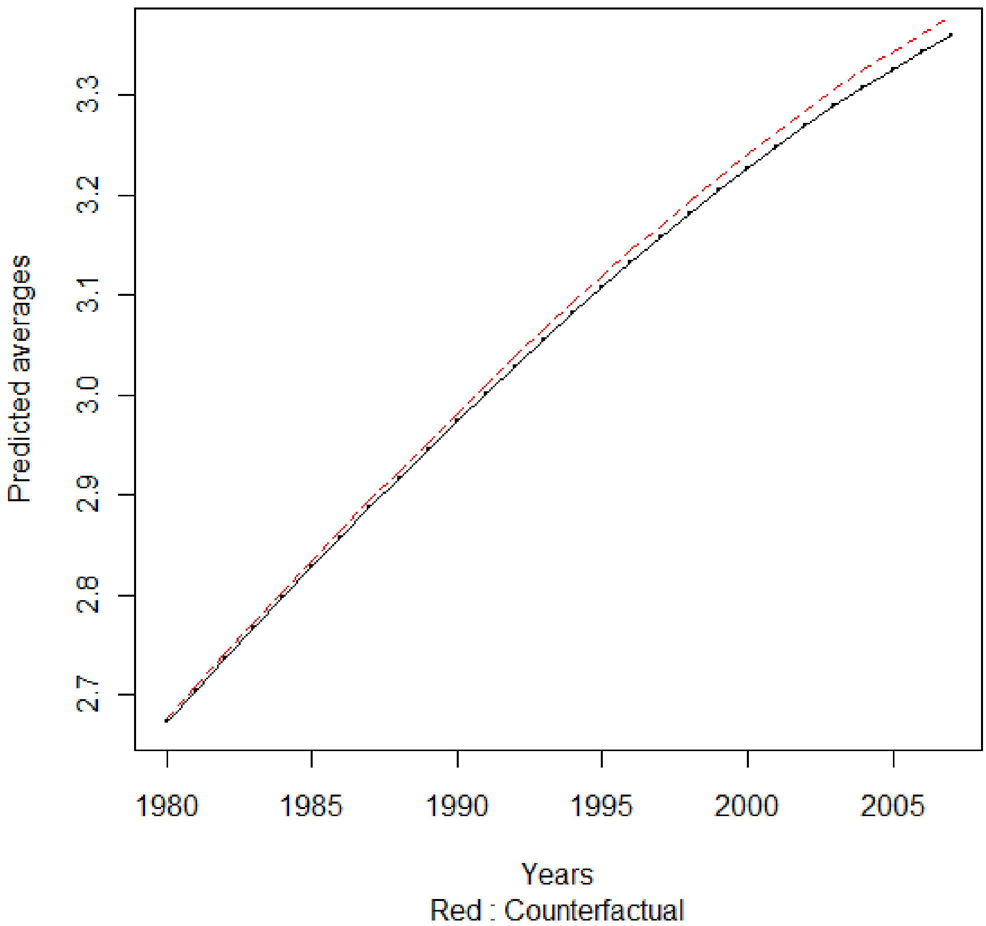


Figure 11: Counterfactual average log earnings (Note: the value of η_3 is divided by 1.05)

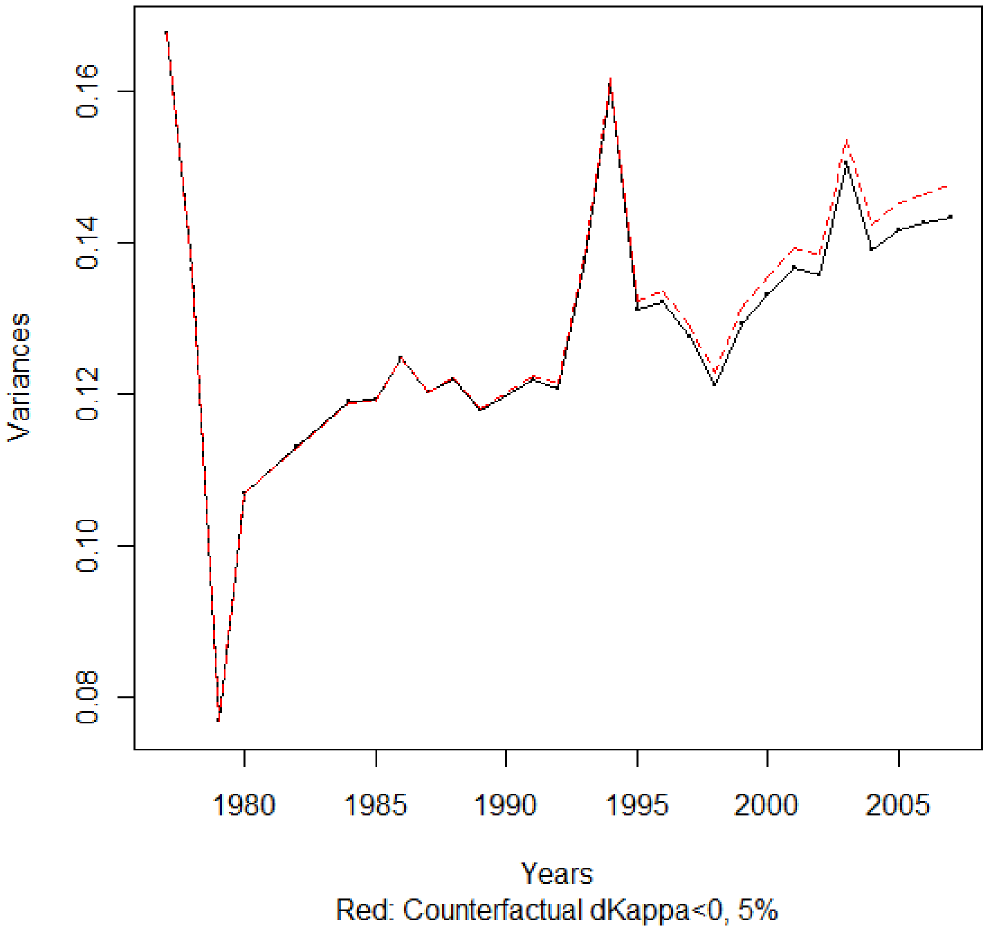


Figure 12: Counterfactual variance of log earnings

Note: The value of η_3 is divided by 1.05 and only observations with more than 20 periods of observations are used.

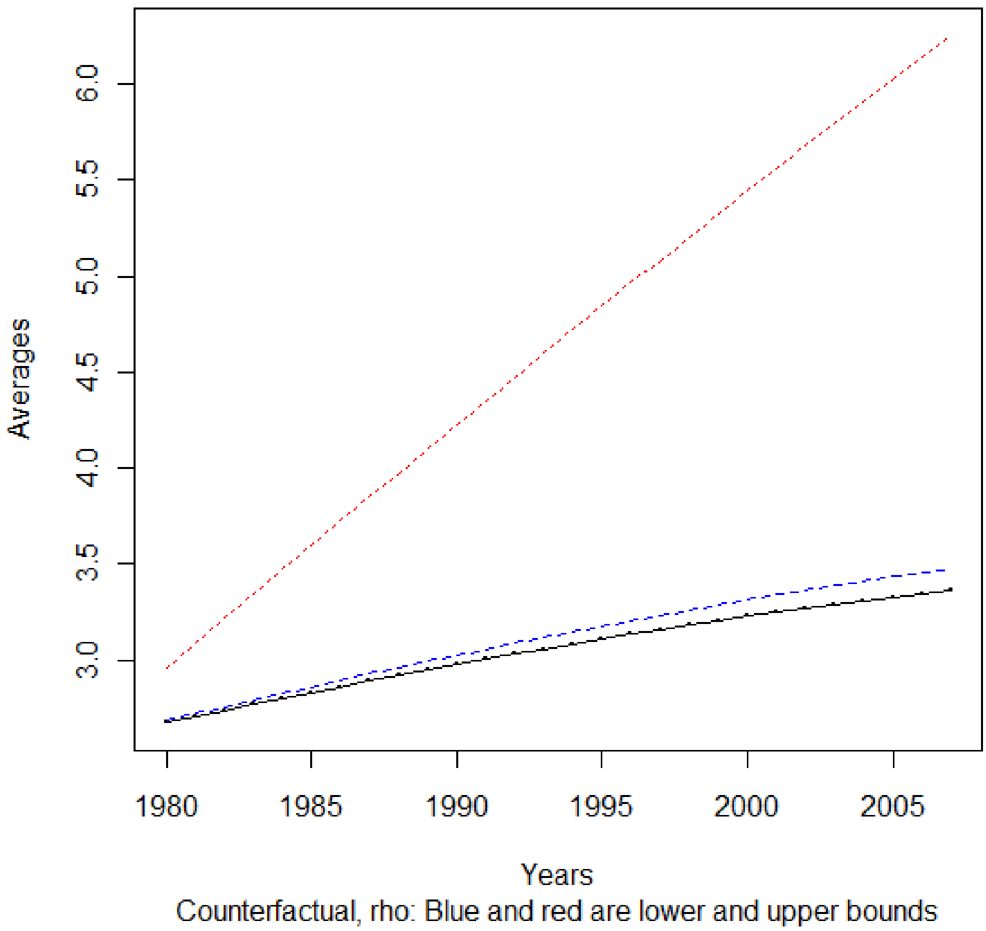


Figure 13: Counterfactual average log earnings: Estimated lower and upper bounds (Note: The value of η_2 is multiplied by 1.05)

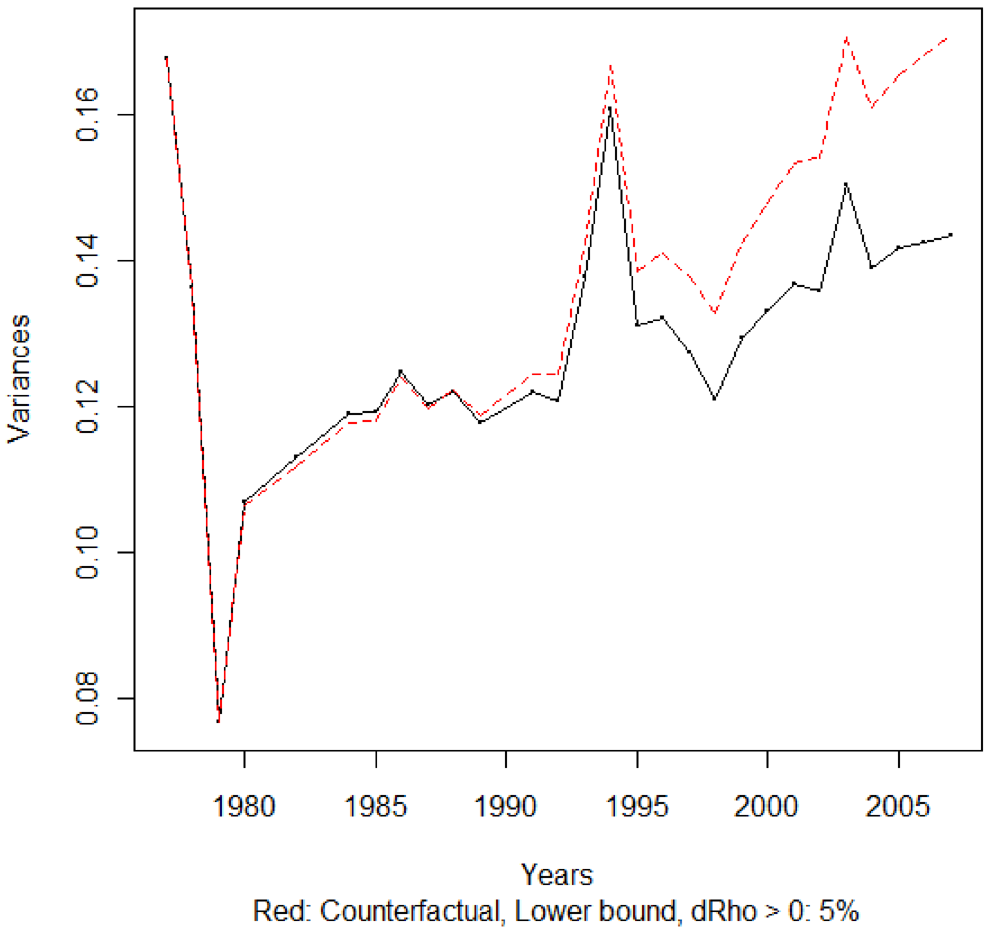


Figure 14: Counterfactual variance of log earnings: Estimated lower bound (Note: The value of η_2 is multiplied by 1.05 and observations with more than 20 observed periods are used)