CIRANO
Centre interuniversitaire de recherche
en analyse des organisations

Série Scientifique
Scientific Series

Nº 95s-4

FIRM HETEROGENEITY AND WORKER SELF-SELECTION BIAS ESTIMATED RETURNS TO SENIORITY

David N. Margolis

Montréal
Janvier 1995

CIRANO

Le CIRANO est une corporation privée à but non lucratif constituée en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de l'Industrie, du Commerce, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche. La *Série Scientifique* est la réalisation d'une des missions que s'est données le CIRANO, soit de développer l'analyse scientifique des organisations et des comportements stratégiques.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de l'Industrie, du Commerce, de la Science et de la Technologie, and grants and research mandates obtained by its research teams. The Scientific Series *fulfils one of the missions of CIRANO: to develop the scientific analysis of organizations and strategic behaviour.*

Les organisations-partenaires / The Partner Organizations

•Ministère de l'Industrie, du Commerce, de la Science et de la Technologie.
•École des Hautes Études Commerciales.
•École Polytechnique.
•Université de Montréal.
•Université Laval.
•McGill University.
•Université du Québec à Montréal.
•Bell Québec.
•La Caisse de dépôt et de placement du Québec.
•Hydro-Québec.
•Banque Laurentienne du Canada.
•Fédération des caisses populaires de Montréal et de l'Ouest-du-Québec.

ISSN 1198-8177

Firm Heterogeneity and Worker Self-Selection Bias Estimated Returns to Seniority

*David N. Margolis**
CIRANO-CRDE-Université de Montréal

Abstract / Résumé

I develop a model under which workers with different marginal productivities self-select into firms based on the firm's seniority reward policy. I show how this may bias upwards the estimates of returns to seniority in cross-sectional and even some longitudinal studies, when differences in workforce composition are ignored. I develop a new estimator of "true" returns to seniority and empirically test the implications of the model. I show how several previous estimation strategies over-estimate returns to seniority, particularly in firms that offer zero or negative returns to job seniority, using a large longitudinal sample of French firms and workers.

Dans ce papier je décris un modèle d'embauches où les individus avec des productivités marginales hétérogènes se trouvent par autosélection, dans les entreprises avec des politiques de rémunération d'ancienneté différentes. Je montre comment ceci peut induire un biais positif dans les estimateurs de rendement de l'ancienneté basés sur les données en coupe transversales et même certains estimateurs basés sur les données longitudinales. Je décris un nouvel estimateur du "vrai" rendement de l'ancienneté, que j'utilise pour tester les implications du modèle. Je montre comment certaines autres approches surestiment les rendements de l'ancienneté, surtout dans les firmes qui rémunèrent très peu l'ancienneté, en utilisant une grande base de données longitudinales des employeurs et employés français.

---

*This paper was prepared for the conference "The Microeconometrics of Human-Resource Management: Multinational Studies of Firm Practices", Paris, France, December 7-9, 1994. This research was supported by the National Science Foundation (SBR 91-11186 and SBR 93-21053). The author would like to thank Kenneth Burdett and Francis Kramarz for helpful discussions. All remaining errors are mine.

1. Introduction

The empirical relation between job seniority and individual earnings has spawned a large literature[1], the vast majority of which has supposed that all firms reward (or fail to reward) job seniority according to the same policy[2]. However, there has been considerable theoretical work to suggest that firms might have incentives to propose different seniority reward policies, particularly with the goal of inducing self-selection of individuals based on characteristics that are not immediately observable to the firm[3]. If these models are correct and firms do offer different seniority reward policies, then the "accepted" models used to estimate the relation between job seniority and earnings deserve reexamination.

This paper proposes an empirical procedure that allows the econometrician to distinguish "true" returns to seniority, defined as the increase in earnings due exclusively to continued employment by the same firm, from changes in earnings due to the evolution in cohort quality arising from worker heterogeneity and self-selection. A simple theoretical model is derived to demonstrate how the bias is introduced into standard estimators that do not account for worker heterogeneity and self-selection. A new estimator is then derived and applied to a large longitudinal data set that allows the econometrician to follow individuals through time and across enterprises, and that allows the econometrician to study the evolution of *cohort* earnings through time within an enterprise.

The paper is organized as follows. Section 2 derives a simple model where firms compete for workers in a labor market, and workers choose their employer based on the contract offered. Firms offer a contract that is comprised of fixed component, a seniority-reward component, an experience-reward component, and a component reflecting the expected productivity of an individual's cohort. I introduce a simple mechanism similar to that suggested by Weiss and Wang [1990] whereby individuals of different marginal productivities have different survival probabilities. This generates self selection of individuals, where firms that offer higher deferred compensation (as seniority rewards) relative to base compensation attract individuals who perceive their probability of separation from their employer in a given period to be "low", and firms that offer lower deferred returns relative to base

[1] See, for example, Topel [1991], Altonji and Shakotko [1987], Abraham and Farber [1987].

[2] A notable exception is Abowd, Kramarz and Margolis (AKM) [1994].

[3] Examples of this work include Salop and Salop [1976] and, more recently, Weiss and Wang [1990].

compensation attract individuals who perceive their probability of separation in a given period to be "high". I show that if there is uncertainty in an individual's belief about her hazard function, this will lead to estimates of returns to job seniority that are biased upwards[4]. I also show that these biases will be strongest for firms that reward job seniority the least, and that these biases lead to biased estimators of returns to job seniority based on "pooled" data, such as those estimated by Topel [1991].

Section 3 then derives a new estimator of returns to seniority. This new estimator takes into account the potential problems for standard estimators as presented in section 2, and shows how to distinguish econometrically between earnings evolution due to contractual provisions and earnings evolution due to changes in expected cohort quality. After briefly describing the French individual panel data in section 4, section 5 describes the empirical implementation of the model of section 2. I first estimate a standard OLS-type model, an individual fixed-effects model and the model proposed by Topel [1991] on the French data, and compare these estimated returns to seniority to both the employment-weighted and equally weighted averages of those of Abowd, Kramarz and Margolis (AKM) [1994], which used the same data as I use here, albeit a different estimation strategy. I then estimate returns to seniority using the cohort-based estimator proposed in section 3, and compare these results to the results obtained by AKM [1994] in order to evaluate the mean bias present even in estimators that permit heterogeneous returns to seniority. I find that even these estimators contain a significant bias component in the vast majority of firms, and that the firms that propose the most "radical" seniority reward policies (seniority rewards with the highest absolute values) are also the firms that have the smallest employment. This suggests that there may be a relation between offered seniority rewards and recruiting ability, as a self-selection model with risk-averse agents might predict. Section 6 summarizes the results and concludes.

2. Self-Selection and Biases in Estimators of Returns to Seniority

In this section, I use a simple model of self-selection based on ability and returns to seniority to show how previous estimates that do not explicitly account for heterogeneous seniority reward policies and self-selection are

[4]Throughout the development of the model I maintain the assumption that high marginal productivity workers have lower hazard rates. This is done for expositional purposes only, and all of the results hold if the hazard rates are reversed, except that the bias becomes positive instead of negative. The empirical work at the ens of section 5 gives an indication as to which hazard rate is greater.

likely to produce upwardly-biased estimators of returns to seniority. Section 2.1 derives a model in which firms compete for workers in the labor market, and workers choose their employer based on the contract offered. Firms offer a contract that is comprised of fixed differential with respect to an individual's expected productivity, a seniority-reward component, and a component reflecting the evolution of the an individual's expected productivity. This includes both the evolution of observable individual specific characteristics (such as total labor market experience), individual specific characteristics that are observable to the firm but not the econometrician, and the evolution of expected productivity based on changes in the composition of an individual's cohort (characteristics that are observable by neither the firm nor the econometrician). Individuals are observed once per period with a fixed probability[5], and higher productivity individuals are less likely to commit a mistake while being observed that is serious enough to result in termination, while lower productivity individuals are more likely to be fired if observed. This generates self selection of individuals when contracts with different earnings profiles are proposed, with firms that offer higher seniority rewards relative to the fixed component of compensation attracting individuals who perceive their probability of termination in a given period to be "low", and firms that offer lower deferred returns relative to the fixed component of compensation attracting individuals who perceive their probability of termination in a given period to be "high".

Section 2.2 shows how, if there is uncertainty in an individual's belief about her hazard function, estimates of returns to job seniority will be biased upwards. These biases are shown to be strongest for firms that reward job seniority the least. Finally, I show that these biases lead to biased estimators of returns to job seniority based on "pooled" data, such as those estimated by Topel [1991], and even those that allow heterogeneity in returns to seniority but suppose constant cohort quality, such as those estimated by AKM [1994].

2.1. A Theoretical Model of Self-Selection and Seniority Rewards

Suppose that there are two types of workers, H and L, where type-H workers have unobserved characteristics that contribute Q_H to their marginal productivities, type-L workers have unobserved characteristics that contribute Q_L to their marginal productivities, and $Q_H > Q_L$. Furthermore, suppose that firms monitor a share s of their workforce each period, and

[5] In a more general context, one can suppose that individuals have a fixed probability of risking separation from their employer, be it employer-induced separation (firing) or employee-induced separation (quitting).

the probability that a type-H individual will be fired as a result of some problem occurring while being monitored is P_H, while the probability that a type-L individual will be fired during monitoring is P_L, and $P_H < P_L < 1$. Individuals know P_H, P_L, Q_H, Q_L, s and their type (although this assumption will be relaxed in section 2.2 below), although they cannot tell if they have been observed. Firms cannot distinguish type-H from type-L workers directly, however they know P_H, P_L, Q_H, and Q_L.

Firms are perfectly competitive in both product and labor markets. This imposes 2 conditions. First, all firms earn zero expected profits ex ante, i.e. the present discounted value of expected total labor costs equals the present discounted value of expected total output[6]. Second, workers are paid (as a base) their expected marginal product, with firms using fixed differentials with respect to expected marginal productivity and fixed seniority reward policies to incite workers of different types to self-select towards different firms. This implies that, if firms made no attempts to attract workers of particular types, and if the expected share of type-L workers hired by firm j at date T_0 in a cohort of age t is ℓ_{j,T_0+t}, individual i employed in firm j hired at date T_0 and employed for t periods would receive earnings

$$y_{i,T_0+t} = \beta X_{i,T_0+t} + \alpha_i + \ell_{j,T_0+t} Q_L + (1 - \ell_{j,T_0+t}) Q_H,$$

where X_{i,T_0+t} is a vector of measurable characteristics of individual i at date T_0+t, β is a vector reflecting the contribution of each of these characteristics to her marginal revenue product and α_i is an individual specific fixed effect that reflects characteristics that contribute to the individual's marginal revenue product and are observable by the firm (hence compensated) but not by the econometrician.

Firms offer new workers a given contract (ϕ_j, γ_j, s_j), where ϕ_j is a fixed differential with respect to expected marginal productivity that an individual employed in firm j earns, and γ_j is the return to an additional year of seniority[7]. Thus individual i employed by firm j at date T_0 and employed

[6] In this analysis we ignore non-labor costs. Alternatively, we could consider that firms have Leontief-type production functions, and that Q_H and Q_L are net of capital costs.

[7] In this paper I suppose that individuals choose their employer as if they expected to stay with the same employer throughout their career. This reduces the complexity of the individual's decision problem from one of choosing a sequence of contracts with a sequence of potentially stochastic termination dates to one of choosing a single contract with a single stochastic termination date. In this sense, workers do not have "employment histories" that could be used by potential employers to gain more information about the individual's type.

Alternatively, I could allow individuals to accept several contracts sequentially and suppose that firms do not observe each other's s_j and that individuals who are mistaken in their assessment of their own types never learn the truth. In this case, since $P_L < 1$, for t periods earns:

$$y_{j,T_0+t} = \beta X_{i,T_0+t} + \alpha_i + \phi_j + \gamma_j t + \ell_{j,T_0+t} Q_L + (1 - \ell_{j,T_0+t}) Q_H. \quad (2.1)$$

Workers and firms share a common discount rate δ. As in most of the literature on self selection and labor contracts, they make one choice from the set of available contracts $\{(\phi_j, \gamma_j, s_j)\}$, selecting the contract that maximizes expected discounted earnings in the job, conditional on the individual's observable characteristics, perception of her type, and the contract offered. If the employment relation is terminated, the worker takes her reservation utility of 0 for the rest of time.

Let N_{j,T_0+t} denote the size of the cohort of employees employed at the date T_0 by firm j observed at the date $T_0 + t$. The expected number of type-L employees employed by firm j in the cohort t periods old at date $T_0 + t$ can be written

$$L_{j,T_0+t} = \ell_{j,T_0+t} N_{j,T_0+t}$$

and the expected number of type-H employees employed by firm j in the cohort t periods old at date $T_0 + t$ can be written

$$H_{j,T_0+t} = (1 - \ell_{j,T_0+t})\, N_{j,T_0+t}.$$

At the time a cohort is hired, $L_{j,T_0} = \ell_{j,T_0} N_{j,T_0}$ and $H_{j,T_0} = (1 - \ell_{j,T_0}) N_{j,T_0}$.

After one period goes by, a share s_j of the workforce is sampled. Those type-L workers sampled separate with probability P_L and stay with probability $(1 - P_L)$. Similarly, those type-H workers sampled separate with probability P_H and stay with probability $(1 - P_H)$. This implies that the firm's workforce after one period will be composed as follows.

$$L_{j,T_0+1} = L_{j,T_0}\,(1 - s_j P_L)$$

$$H_{j,T_{0+1}} = H_{j,T_0}\,(1 - s_j P_H)$$

$$N_{j,T_0+1} = N_{j,T_0}\,(1 - s_j\,[\ell_{j,T_0} P_L + (1 - \ell_{j,T_0})\, P_H])$$

$$\ell_{j,T_0+1} = \frac{\ell_{j,T_0}}{1 - s_j\left[\ell_{j,T_0} P_L + \left(1 - \ell_{j,T_0}\right) P_H\right]}$$

knowing individual i's seniority at the date of separation for the previous job can not be used to gain information about her type. This does not, however, reduce to complexity of the individual's decision problem.

In Margolis [1994], I derive a two-period model where workers are heterogeneous in the same general manner as this paper and can choose to leave their initial employer after one period or risk being fired after being evaluated. Firms learn about the workers based on their choice of initial contract and their decision to quit or risk being fired.

By induction, it is clear that

$$L_{j,T_0+t} = L_{j,T_0+t-1}(1-s_jP_L) = L_{j,T_0}(1-s_jP_L)^t$$

$$H_{j,T_0+t} = H_{j,T_0+t-1}(1-s_jP_H) = H_{j,T_0}(1-s_jP_H)^t$$

$$N_{j,T_0+t} = N_{j,T_0+t-1}(1-s_j[\ell_{j,T_0+t-1}P_L + (1-\ell_{j,T_0+t-1})P_H])$$

and

$$\begin{aligned} \ell_{j,T_0+t} &= \frac{L_{j,T_0+t}}{L_{j,T_0+t}+H_{j,T_0+t}} \\ &= \frac{\ell_{j,T_0}(1-s_jP_L)^t}{\ell_{j,T_0}(1-s_jP_L)^t+\left(1-\ell_{j,T_0}\right)(1-s_jP_H)^t} \end{aligned} . \tag{2.2}$$

Given equation (2.1), individual i who believes herself to be of type q will have expected discounted earnings from accepting a contract (ϕ_j, γ_j, s_j) at date T_0 of:

$$Y_{i,T_0} \quad | \quad q = \sum_{t=1}^{\infty} \delta^{t-1}(1-s_jP_q)^t \tag{2.3}$$

$$(\beta X_{i,T_0+t} + \alpha_i + \phi_j + \gamma_j t + \ell_{j,T_0+t}Q_L + (1-\ell_{j,T_0+t})Q_H).$$

Equation 2.3 can be rewritten as

$$Y_{i,T_0} \mid q = \frac{(1-s_jP_q)(\alpha_i+\phi_j+Q_H)}{1-\delta(1-s_jP_q)} + \frac{\gamma_j(1-s_jP_q)}{(1-\delta(1-s_jP_q))^2} + A \tag{2.4}$$

where

$$A = \sum_{t=1}^{\infty} \delta^{t-1}(1-s_jP_q)^t(\beta X_{i,T_0+t} + \ell_{j,T_0+t}(Q_L-Q_H)).$$

Using 2.4, it can be shown that the marginal rate of substitution between returns to seniority (γ_j) and the earnings differential with respect to expected marginal productivity (ϕ_j) is

$$MRS_{\gamma,\phi}(q) = 1-\delta(1-s_jP_q). \tag{2.5}$$

This implies that an individual of type q requires γ_j to increase by $1-\delta(1-s_jP_q)$ for every unit of ϕ_j given up if utility is to remain constant[8]. Since $P_L > P_H$, this implies that $MRS_{\gamma,\phi}(L) > MRS_{\gamma,\phi}(H)$. In other words, if

[8]Note that there is no factor in this expression corresponding to $\frac{\partial \ell_{j,T_0}}{\partial \phi}$ or $\frac{\partial \ell_{j,T_0}}{\partial \phi}$. This is due to the presence of only 2 types of individuals. If an individual were to choose a contract (ϕ_1, γ_1, s_1) over a contract (ϕ_2, γ_2, s_2), that implies her utility under contract

an employer proposes two contracts of equal expected total compensation cost over the expected duration of the contract, but one has a high ϕ_j and a low γ_j while the other has a low ϕ_j and a high γ_j, type L workers will prefer the first contract to the second, and type H workers will provide the second to the first, provided

$$\triangle\gamma_j = -\left(1 - \delta(1 - s_j P_{\tilde{q}})\right) \triangle\phi_j, \tag{2.6}$$

when $P_L \geq P_{\tilde{q}} \geq P_H$[9].

In order for this separating equilibrium to exist, it has to be shown that

1. there exist at least two contracts $\{(\phi_1, \gamma_1, s), (\phi_2, \gamma_2, s)\}$ such that each contract provides the employer with zero expected profits over the expected duration of the contract conditional upon who accepts it (ℓ_{j,T_0}) and

2. there exists some $P_{\tilde{q}}$, $P_L \geq P_{\tilde{q}} \geq P_H$, such that, for each contract, $\frac{(\gamma_1 - \gamma_2)}{(\phi_1 - \phi_2)} = -\left(1 - \delta(1 - sP_{\tilde{q}})\right)$.

Start by noting that the structure of the contracts described in equation (2.1) is such that each worker receives her expected marginal productivity as compensation at each point in time, plus the ϕ_j and $\gamma_j t$ components, which can be used to shift the dates of arrival of compensation and offset current-period compensation from current-period expected productivity. Next, consider the contract $(0, 0, s)$, where $0 \leq s \leq 1$. This contract

1 is greater than or equal to the utility she would receive under contract 2.

When

$$U(\phi_1, \gamma_1, s_1) = U(\phi_2, \gamma_2, s_2),$$

the individual may change the contract chosen for a marginal change in ϕ or γ. But such a case suggests that the (unique) contract on offer is a pooling equilibrium. Rothschild and Stiglitz [1976] showed that a pooling contract can never be a Nash equilibrium, and thus we need only be concerned with the case of $U(\phi_1, \gamma_1, s_1) > U(\phi_2, \gamma_2, s_2)$.

Since I am implicitly assuming that an individual's utility is a continuously differentiable function of her present discounted value of earnings, a marginal change in either ϕ or γ will not change her choice of contracts when

$$U(\phi_1, \gamma_1, s_1) > U(\phi_2, \gamma_2, s_2).$$

This is because the function that maps contracts into present discounted values is continuously differentiable in both ϕ and γ, and thus the composite function that maps from contract into utilities is also continuously differentiable in ϕ and γ.

[9] In what follows, I assume that the share of individuals risking separation each period is the same under the 2 contracts, i.e. $s_i = s_j \;\forall\, i, j$. Since the observation rate just has the effect of shifting the failure probabilities in one firm relative to another (albeit under a particular functional form), this restriction implies that the probability that an individual of a given type will be fired in any period is the same across all firms. I make this assumption to simplify the exposition; allowing s to vary across firms only adds another dimension of flexibility to the set of available contracts and does not change the theoretical results in any substantial manner.

satisfies condition 1, that the offering firm earns zero profits, since the wage bill at each point in time for each worker is equal to that worker's expected marginal product[10].

Next, consider the contract $(-\phi_j, [1-\delta(1-s[P_H+\varepsilon])]\phi_j, s)$, where $\phi_j > 0$. By construction, the condition for a separating equilibrium given in 2 is satisfied. Furthermore, since this contract will attract type-H workers while type-L workers will prefer contract 1, $\ell_{j,T_0} = 0$. Given the structure of contracts as described in equation 2.4, it can be shown that, as $\varepsilon \to 0$, this contract also generates zero expected profits over its lifetime and thereby satisfies condition 1. Thus we have shown, using logic similar to that of Salop and Salop [1976], that there exists a separating Nash equilibrium, and that it is not necessarily optimal for all firms to offer the same seniority reward policy.

2.2. The Bias Induced by Self-Selection

Now suppose that individuals have some uncertainty as to their hazard function, i.e. type-H workers consider themselves type-L with probability $f_{L|H}$, while type-L workers consider themselves type-H with probability $f_{H|L}$. In the context of the model described above in section 2.1, this implies that type-L individuals will choose the "wrong" contract with probability $f_{H|L}$, while type-H workers choose the "wrong" contract with probability $f_{L|H}$. In this case, a separating equilibrium will generate

$$\begin{aligned} \ell_{1,T_0} &= \frac{\Lambda(1-f_{H|L})}{\Lambda(1-f_{H|L})+(1-\Lambda)f_{L|H}} \\ \ell_{2,T_0} &= \frac{\Lambda f_{H|L}}{\Lambda f_{H|L}+(1-\Lambda)(1-f_{L|H})} \end{aligned} \tag{2.7}$$

and the zero expected profit condition implies that contract 2 will have to be of the form suggested in footnote 10, namely

$$\left(-\phi_2, \quad \begin{array}{c} \phi_2 \quad [1-\delta(1-sP_L)][1-\delta(1-sP_H)] \\ \left(\frac{\ell_{2,T_0}(1-sP_L)(1-\delta(1-sP_H))+\left(1-\ell_{2,T_0}\right)(1-sP_H)(1-\delta(1-sP_L))}{\ell_{2,T_0}(1-sP_L)(1-\delta(1-sP_H))^2+\left(1-\ell_{2,T_0}\right)(1-sP_H)(1-\delta(1-sP_L))^2} \right) \end{array}, s \right)$$

---

[10] Of course, this is not the only contract that generates zero profits for the offering firm. Any contract that satisfies

$$\gamma_j = -\phi_j \quad \begin{array}{c} [1-\delta(1-sP_L)][1-\delta(1-sP_H)] \\ \left(\frac{\ell_{j,T_0}(1-sP_L)(1-\delta(1-sP_H))+\left(1-\ell_{j,T_0}\right)(1-sP_H)(1-\delta(1-sP_L))}{\ell_{j,T_0}(1-sP_L)(1-\delta(1-sP_H))^2+\left(1-\ell_{j,T_0}\right)(1-sP_H)(1-\delta(1-sP_L))^2} \right) \end{array}$$

will generate zero profits for the firm that proposes it. In the case where employees know their type with certainty, $\ell_{j,T_0} = 1$ for the firms that attract type-L workers and $\ell_{j,T_0} = 0$ for the firms that attract type-H workers. This reduces to the condition that $\gamma_j = -(1-\delta(1-s_jP_q))\phi_j$, where q corresponds to the type of worker that the firm attracts.

where $\phi_2 > 0$ and Λ is the proportion of type-L workers in the labor force. Note that this contract also satisfies the condition for a separating equilibrium (condition 2 above), with

$$P_{\tilde{q}} = \frac{\left[\ell_{2,T_0}(1-sP_L)(1-\delta(1-sP_H))^2\right]P_L+\left[\left(1-\ell_{2,T_0}\right)(1-sP_H)(1-\delta(1-sP_L))^2\right]P_H}{\ell_{2,T_0}(1-sP_L)(1-\delta(1-sP_H))^2+\left(1-\ell_{2,T_0}\right)(1-sP_H)(1-\delta(1-sP_L))^2}$$

which satisfies $P_L > P_{\tilde{q}} > P_H$.

This has important implications for most estimators of returns to experience. When individuals are certain of their type, it is clear from equation 2.2 that $\ell_{j,T_0+t} = \ell_{j,T_0}$ for any t, meaning that expected productivity of the cohort was not evolving. Thus any estimator that consistently estimates returns to tenure under the assumption that $\gamma_j = \gamma_k$ for all (j,k) would consistently estimate the employment-weighted mean of γ_j in the sample population.

On the other hand, once $\ell_{j,T_0} \notin \{0,1\}$, the expected productivity of the workforce will evolve over time, as workers of different types hazard out of their cohorts at different rates. Furthermore, it can be shown (after some messy algebra) that

$$\frac{\partial(\ell_{j,T_0+t+1}-\ell_{j,T_0+t})}{\partial\ell_{j,T_0}} = \begin{array}{l} \frac{(1-sP_L)^t(1-sP_H)^{t+1}}{\left(\ell_{j,T_0}(1-sP_L)^{t+1}+(1-\ell_{j,T_0})(1-sP_H)^{t+1}\right)^2} \\ -\frac{(1-sP_L)^t(1-sP_H)^t}{\left(\ell_{j,T_0}(1-sP_L)^t+(1-\ell_{j,T_0})(1-sP_H)^t\right)^2} \end{array} > 0.$$

In other words, the larger the initial share of type-L workers in the cohort, the faster the expected productivity will be seen to rise.

This fact has implications for both homogeneous seniority returns estimators and heterogeneous returns-homogeneous workforce estimators. In the first case, this suggests that not only will estimated returns to seniority in low "true" returns (and thus high ℓ_{j,T_0}) firms be overestimated, but that the larger the share of low-γ_j firms in the sample, the larger the overestimate is likely to be[11]. In the case of estimators that allow for firm-level heterogeneity in γ_j but not in ℓ_{j,T_0}, the estimated returns to seniority will be biased upwards by a larger amount in the firms that reward seniority relatively poorly than in those firms that reward seniority relatively well. Whether or not this bias is will be sufficiently large to overcome the difference in true returns to seniority (and thus provide the opposite impression; namely that the low actual γ_j firms will have estimated returns higher than the high actual γ_j firms) depends on many factors, including $\gamma_1 - \gamma_2$, Λ, $f_{L/H}$ and $f_{H/L}$, and is not predictable a priori.

[11] This condition poses particular problems for models of labor markets that suppose an "ocean of small firms" in which only the large employers reward seniority. In these models the small firms, which are more numerous, will all appear to be rewarding seniority positively, which might invite the econometrician to reject the "ocean of small firms" hypothesis incorrectly.

2.2.1. Topel's 2-step Estimator

Consider the two-step estimator of returns to job seniority proposed by Topel [1991]. In this model, a first-step equation of the form

$$\begin{aligned} \triangle y_{i,T_0+t} &= (\gamma_1+\beta_1)\triangle t_{i,j,T_0+t} + \gamma_2 \triangle t^2_{i,j,T_0+t} + \gamma_3 \triangle t^3_{i,j,T_0+t} \\ &+ \gamma_4 \triangle t^4_{i,j,T_0+t} + \beta_2 \triangle Exp^2_{i,T_0+t} + \beta_3 \triangle Exp^3_{i,T_0+t} \\ &+ \beta_4 \triangle Exp^4_{i,T_0+t} + \triangle \varepsilon_{i,T_0+t} \end{aligned} \tag{2.8}$$

is used to recover estimates of higher-order terms in returns to seniority and returns to experience ($\triangle t_{i,j,T_0+t}$ is the change in the number of years of seniority of individual i in firm j between dates T_0+t and T_0+t-1, $\triangle Exp_{i,T_0+t}$ is the change in the level of total labor market experience of individual i between dates T_0+t and T_0+t-1, and $\triangle\varepsilon_{i,T_0+t}$ is the difference in stochastic components of individual i's earnings between dates T_0+t and T_0+t-1). Note that the coefficients γ_1 through γ_4 do not have j subscripts, implying that they are shared by all firms in the economy. The coefficient on the difference in level seniority is interpreted as the sum of returns to seniority (in levels) and returns to experience (in levels). Since the year-on-year difference in seniority is 1, as is the year-on-year difference in experience, for an individual who does not change employers during the course of the year[12], this is essentially the intercept in a first-differenced regression[13]. A second-step equation of the form

$$\tilde{y}_{i,T_0+t} = \beta_1 Exp_{i,T_0} + \theta F_{i,T_0+t} + e_{i,T_0+t}$$

where

$$\begin{aligned} \tilde{y}_{i,T_0+t} = y_{i,T_0+t} &- \widehat{(\gamma_1+\beta_1)} t_{i,j,T_0+t} - \hat{\gamma}_2 t^2_{i,j,T_0+t} - \hat{\gamma}_3 t^3_{i,j,T_0+t} \\ &- \hat{\gamma}_4 t^4_{i,j,T_0+t} - \hat{\beta}_2 Exp^2_{i,T_0+t} \\ &- \hat{\beta}_3 Exp^3_{i,T_0+t} - \hat{\beta}_4 Exp^4_{i,T_0+t} \end{aligned}$$

is then used to decompose the estimated intercept $\widehat{(\gamma_1+\beta_1)}$ into the part γ_1 that should be attributed to returns to seniority and the part β_1 that should be attributed to returns to experience. Topel argues that, in the presence of individual effects α_i and firm effects ϕ_j, the first step of his two step procedure allows one to recover a consistent estimate of $(\gamma_1+\beta_1)$, as well as consistent estimates of γ_1 through γ_4 and β_1 through β_4. The

[12] For the higher-order terms, such as t^2_{i,j,T_0+t} or Exp^2_{i,T_0+t}, the year-on-year differences will be different for all employers except the first employer encountered in the worker's career.

[13] If the individual stayed with the same employer during the entire year but suffered temporary layoffs during the course of the year, $\Delta t_{i,j,T_0+t}$ would be less than one, although the coefficient would still confound returns to seniority and returns to experience.

second step allows one to obtain an upper bound on β_1, and thus a lower bound on γ_1, the (level) returns to seniority[14].

The first problem with this procedure is that, by constraining all firms to reward seniority in the same manner, it precludes the possibility that firms might strategically use different levels of initial earnings and rates of earnings growth of the career to incite individuals with particular combinations of productive capacities (unobservable by the firm) to join their firms. Thus the firms do not try to take advantage of the heterogeneity of the pool of available workers, which is equivalent to imposing that employers do not behave rationally when choosing their compensation policies[15].

This approach could be applied on a firm-by-firm basis, but that imposes an entirely different set of difficulties. First, if rewards to total labor market experience are meant to reflect the market's perception of the rate of accumulation of general human capital, then it would be incorrect to allow β_1 through β_4 to vary by firm. Imposing this restriction across regression equations can be complicated econometrically for data sets with large numbers of employers and individuals[16]. However, once one can separately determine the firm-specific returns to seniority γ_{1j} through γ_{4j} in Topel's original model, one can directly control for individual- and firm-specific effects α_i as in Abowd, Kramarz and Margolis [1994] (see below), and thus the two-step procedure becomes unnecessary. Furthermore, without a theoretical justification for the presence of the higher order terms in seniority in either the initial earnings equation or the first-differenced one estimated in step one, there is no clear interpretation of the coefficients on seniority squared, cubed or to the fourth power.

On the other hand, the model suggested in section 2.1 above provides such a justification. The coefficients on the higher order terms can be interpreted as the second- third- and fourth-order terms in a fourth-order Taylor expansion of ℓ_{j,T_0+t} around T_0. Unfortunately, every term in the fourth-order expansion of ℓ_{j,T_0+t} is different from the corresponding term

---

[14] This estimator of β_1 is an upper bound due to anticipated covariance between Exp_{i,T_0} and e_{i,T_0+t}. The idea that "good jobs last longer" implies that jobs that pay more will have been held longer, and thus a lower initial level of experience is likely to be found in jobs with above-average earnings conditional on all observables. This is a testable assumption, and although Topel does not reject it for the U.S. data, it is rejected on the French data used here.

[15] This could nevertheless be a fairly accurate representation of certain unionized labor markets, where the employer is not completely free to choose intercept and slope of the earnings profiles of its workers. On the other hand, if employers are free to choose their compensation policies, this specification will only be optimal if attempting to induce the sort of selection described above would be too costly for all firms, in which case all firms would optimally choose the same compensation policy.

[16] See Abowd, Kramarz and Margolis (1994), and below, for an econometric procedure designed to handle this sort of computational problem.

in the fourth-order expansion of $\ell_{j,T_0+t} - \ell_{j,T_0+t-1}$. In addition, the first-order term will represent a combination of "real" returns to seniority (the γ_j in equation (2.1) above) and returns to changes in cohort quality[17]. This suggests that $\tilde{y}_{i,T_0+t}$ used as the dependent variable in the second-step regression is inappropriate, since the coefficients on the tenure terms of the Taylor expansion from the first step equation will not, in general, be identical to those from the non-differenced earnings equation.

2.2.2. Abowd, Kramarz and Margolis' Heterogeneous Returns Estimator

AKM [1994] used the same data set I use here to estimate firm-specific intercepts (ϕ_j) and firm-specific coefficients on seniority (which might be interpreted as "returns to seniority" (γ_j)). Their initial statistical model is of the form

$$y_{i,T_0+t} = x_{i,T_0+t}\beta + u_i\eta + \alpha_i + \phi_j + \gamma_{1j}t + \gamma_{2j}\,\mathrm{T}_1(t-10) + \varepsilon_{i,T_0+t} \quad (2.9)$$

which is similar to the form in equation (2.1) except that they explicitly decompose the X_{i,T_0+t} from equation (2.1) into a time-varying components (x_{it}) and a non-time-varying component (u_i), and they model nonlinearities in the returns to seniority with a linear spline function having a break at 10 years. Their model is essentially descriptive, and thus they do not attempt to explicitly model the source of the nonlinearities in returns to seniority, as is done in section 2.1 above. The econometric procedure involves projecting the firm-effect onto a vector of firm and person characteristics constructed so as to allow the desired correlation among the individual-effects and the firm-effects. This permits consistent estimation of the components of β from equation (2.1) that correspond to time-varying individual specific characteristics. A second step uses least squares to decompose the non-time-varying, individual specific component of earnings (which they call θ_i) into the observable components (representing the rest of the X_i terms in equation (2.1)) and the unobservable component α_i. A third step uses the resulting estimates to generate consistent estimates of the firm-effects ϕ_j and γ_j.

This procedure provides a good starting point for discriminating between true returns to seniority and returns to evolution in cohort quality. However, the estimated coefficient γ_{1j} does not provide enough information to be able to recover the true γ_j, and even considering γ_{1j} and γ_{2j} together does not provide enough information. The basic problem is that

[17] Section 2.2.2 below discusses this point more thoroughly.

$$\begin{aligned}\gamma_{1j} &= E\left[\frac{\partial y_{i,T_0+t}}{\partial t} \mid t<10\right] \\ &= \gamma_j + (Q_L - Q_H)\, E\left[\frac{\partial \ell_{j,T_0+t}}{\partial t} \mid t<10\right]\end{aligned} \tag{2.10}$$

and

$$\begin{aligned}\gamma_{2j} = &\ E\left[\tfrac{\partial y_{i,T_0+t}}{\partial t} \mid t \geq 10\right] - \gamma_{1j} \\ = &\ (Q_L - Q_H)\left\{E\left[\tfrac{\partial \ell_{j,T_0+t}}{\partial t} \mid t\geq 10\right] - E\left[\tfrac{\partial \ell_{j,T_0+t}}{\partial t} \mid t<10\right]\right\}.\end{aligned}$$

Clearly, given the fact that ℓ_{j,T_0+t} is nonlinear in t in a rather complicated way, these expressions do not allow one to recover γ_j directly.

3. A New Estimator of Returns to Job Seniority

Neither the estimator presented in section 2.2.1 nor in section 2.2.2 explicitly allowed for self-selection of individuals based on the offered contract, and thus neither gives a clean interpretation of the nonlinearity in returns to seniority. However, the expression for γ_{1j} in equation (2.10) above suggests an econometric approach to directly estimating γ_j.

The ideal situation would be to be able to estimate separately, for each starting year ($\bar{T}_0$) and each level of seniority ($\bar{t}$), the equation

$$y_{i,T_0+t} = x_{i,T_0+t}\beta + u_i\eta + \alpha_i + \phi_j + \gamma_j t + \varepsilon_{i,T_0+t} \tag{3.1}$$

using the method suggested in AKM [1994]. According to equation (2.10), the estimated $\hat{\gamma}_j$ could be interpreted as $\gamma_j + (Q_L - Q_H)\frac{\partial \ell_{j,\bar{T}_0+\bar{t}}}{\partial t}$. The expression for $\frac{\partial \ell_{j,\bar{T}_0+\bar{t}}}{\partial t}$, although rather messy, is calculable. We could look across levels of seniority for the same cohort and, given at least 6 years with estimated $\hat{\gamma}_j$, recover the parameters $\gamma_j, Q_L, Q_H, s_jP_L, s_jP_H$, and ℓ_{j,T_0}. More than 6 years of data would provide overidentifying restrictions Furthermore, if we had 2 cohorts and 8 estimated $\hat{\gamma}_j$, or 3 cohorts and 9 estimated $\hat{\gamma}_j$, etc..., we would be in the same situation. And additional estimated $\hat{\gamma}_j$ would provide additional overidentifying restrictions. If we were willing to suppose that Q_L, Q_H, P_L, and P_H were constant across firms, we could also use cross-firm restrictions to recover s_j and to test hypotheses about Q_L, Q_H, P_L, and P_H.

Unfortunately, regardless of the number of individuals in the cohort, one can not hope to estimate $\hat{\gamma}_j$ unless there is some variation in t, which we would be precluding if we were to run regressions by cohort-seniority level.

On the other hand, if we just ran the regression (3.1) for all observations having the same entering year ($\bar{T}_0$) for all seniority levels, the interpretation of $\hat{\gamma}_j$ would now be

$$\hat{\gamma}_{j,T_0} = E\left[\frac{\partial y_{i,T_0+t}}{\partial t} \mid T_0 = \bar{T}_0\right] = \gamma_j + (Q_L - Q_H)\, E\left[\frac{\partial \ell_{j,T_0+t}}{\partial t} \mid T_0 = \bar{T}_0\right]. \tag{3.2}$$

Note that although the expectation of the derivative is no longer a constant (it will vary with the distribution of observations within the cohort across tenure levels, which will in turn be a function of P_H, P_L, s, and ℓ_{j,T_0}), by considering only workers starting at $T_0 = \bar{T}_0$, we are effectively controlling for ℓ_{j,T_0}.

We run this regression for each entering cohort within the firm[18] and assign the estimated $\hat{\gamma}_{j,T_0}$ for each cohort to all observations corresponding to individuals hired at T_0. Denote the vector of estimated $\hat{\gamma}_{j,T_0}$ corresponding to all of the observations in all of the cohorts employed by firm j by $,_j$, and regress $,_j$ on tenure and an intercept.

$$,_j = \psi_{1j} + \psi_{2j} t + \varsigma_j \tag{3.3}$$

The coefficients of this secondary regression have clear interpretations. First and foremost, $\hat{\psi}_{1j} = \gamma_j$. In other words, the intercept from the regression of estimated returns to seniority on seniority is the "true" returns to seniority. The estimate of the coefficient on seniority is $\hat{\psi}_{2j} = E\left[\frac{\partial^2 \ell_{j,T_0+t}}{\partial t^2}\right]$. We would like to be able to use this estimate, in conjunction with the estimated $\hat{\gamma}_{j,T_0}$, to recover the remaining parameters $(Q_L, Q_H, s_j P_L, s_j P_H, \ell_{j,T_0^1}, ..., \ell_{j,T_0^N})$, where T_0^N refers to the cohort entering at date N. Unfortunately this will not be possible, since there will be as many ℓ_{j,T_0^n} terms as there are cohorts, which is one less than the number of "data" points we have available from which to recover the parameters. If we are willing to impose restrictions on the ℓ_{j,T_0^n} terms (such as $\ell_{j,T_0^n} = \ell_{j,T_0}$ for all n), we can try to solve the nonlinear system of equations to recover the other parameters of interest. And if we are willing to impose other restrictions on the other parameters (such as equality across firms as mentioned above), we should be able to recover s_j and test hypotheses on the other estimated parameters.

[18] As mentioned below, the data requirements necessary to implement this procedure are fairly rigorous. For all individuals that are members of cohorts with fewer than 3 observations, or for which no individual survived into the second year, we will be unable to recover a separate $\hat{\gamma}_j$. Thus, in addition to the cohort-specific $\hat{\gamma}_j$ estimates, one may be forced to pool the observations from the remaining cohorts. In this case, one must assume that these smaller, or less successful, cohorts all had the same initial compostion ℓ_{j,T_0}.

Unfortunately, the requirements placed on the data are slightly more onerous than in AKM [1994]. First, the procedure requires that we have at least 3 observations per cohort on which we run the third-step regression. Among the observations in a given firm cohort, we must have more than one year's worth of data for at least one individual. In addition, in order to recover the "true" γ_j, we must have at least 3 cohorts in the same firm which meet these requirements. Fortunately, the data used here allows us to apply this procedure to firms covering the vast majority of observations in our sample.

4. The French Longitudinal Data

The data used in the empirical work described in section 5 below are the same as those described in AKM [1994]. These data are a panel of individual-firm matched observations with identifying information sufficient to follow both firms and individuals over time, collected by INSEE (the French National Statistics Institute). The data cover the period 1976-1987, although observations from 1981 and 1983 were not made available by INSEE. They cover all workers employed in France born in October of an even-numbered year, thus ensuring a random initial sample of individuals. The data made available concern all such people working in enterprises with 10 employees or more, and French government employees are also excluded (although employees of government-owned enterprises are present in the data).

These data were cleaned[19] and compiled into a database containing initially 4,784,284 observations, of which 3,099,056 were from men and 1,685,228 were from women. A simple regression of the form

$$LFRAISRE_{i,T_0+t} = X_{i,T_0+t}\beta + \varepsilon_{i,T_0+t}$$

was run on these data, where $LFRAISRE_{i,T_0+t}$ represents the log of the real annualized total compensation cost, and the vector X_{i,T_0+t} includes the variables male, Paris region, experience through experience to the fourth power, 7 education-level indicators, and 10 year indicators. All observations more than 5 standard deviations away from their predicted values were considered outliers. These observations were eliminated from the data set on which the Topel, AKM and cohort-based estimations were generated, but section 5.1 below presents the results of models run on both the full and outliers-eliminated data. Table 4.1 provides some basic sample statistics for both of the data sets used.

[19] See the data appendix of AKM [1994] for details on the preliminary treatment of the data.

Table 4.1: Descriptive Statistics for Data used in OLS and Fixed-Effects Estimation

Variable Name	All Observations		Outliers Eliminated	
	Mean	Std. Dev.	Mean	Std. Dev.
LFRAISRE	4.251	0.556	4.257	0.519
Seniority	6.425	7.111	6.437	7.112
$\text{Seniority}^2/100$	0.918	1.475	0.920	1.476
$\text{Seniority}^3/1000$	1.584	3.266	1.587	3.268
$\text{Seniority}^4/10000$	3.018	7.672	3.024	7.677
Experience	16.612	11.920	16.612	11.915
$\text{Experience}^2/100$	4.180	4.934	4.179	4.930
$\text{Experience}^3/1000$	12.603	19.533	12.596	19.516
$\text{Experience}^4/10000$	41.808	78.875	41.772	78.778
Elementary School	0.168	0.157	0.168	0.157
Junior High School	0.067	0.088	0.067	0.088
High School	0.059	0.084	0.059	0.083
Basic Vocational-Technical School	0.239	0.178	0.239	0.178
Advanced Vocational-Technical School	0.064	0.087	0.064	0.087
Technical/Undergraduate University	0.060	0.097	0.060	0.097
Graduate School Degree	0.039	0.085	0.039	0.085
Male	0.648	0.478	0.648	0.478
Paris Region	0.270	0.444	0.269	0.444

5. Empirical Examination of the Ability-Seniority-Earnings Relation

In this section, I use the French individual panel data to estimate returns to job seniority using an OLS estimator, an individual fixed-effects panel estimator and Topel's [1991] two-step panel estimator, which I then compare to AKM's [1994] correlated effects panel estimator. Finally, I estimate the cohort-based panel estimator proposed in section 3 and use it to show what part of estimated returns to seniority in the AKM [1994] correlated effects panel estimator is due to "true" seniority returns and what part is due to evolution in cohort quality.

5.1. Simple Estimators: OLS and Individual Fixed Effects

I first estimate returns to job seniority using a simple specification of the form

$$\log\left(y_{i,T_0+t}\right) = \beta X_{i,T_0+t} + \varepsilon_{i,T_0+t} \tag{5.1}$$

where the matrix X_{i,T_0+t} contains seniority through seniority to the fourth power (rescaled), experience through experience to the fourth power (rescaled), seven indicators for educational attainment (highest degree terminated), an indicator variable for a job in the Paris metropolitan area, and an indicator for the sex of the individual. The dependent variable was measured as the log of real annual total compensation cost ($LFRAISRE$). The regressions were run on two data sets, the first containing all observations and the second being identical to the first with the outliers eliminated[20]. For each data set, I estimated returns to seniority under two different specifications. In the first (model 1), I included only the level of job seniority (along with all of the other elements of the X_{i,T_0+t} matrix), while in the second (model 2) I included job seniority, seniority squared, cubed and to the fourth power in addition to all of the variables used in the first regression. Table 5.1 presents the results of each of the two OLS regressions on each of the two data sets.

I next estimate returns to job seniority using a specification that allows for individual-specific heterogeneity of the form

$$\log\left(y_{i,T_0+t}\right) = \beta X_{i,T_0+t} + \alpha_i + \varepsilon_{i,T_0+t}$$

where all of the variables are as defined above. In order to control for the individual-specific effects, I project the X_{i,T_0+t} and $\log\left(y_{i,T_0+t}\right)$ onto their

[20] An outlier was defined as any observation for which the log of the real total annual compensation cost was more than 5 standard deviations away from its predicted value. See section 4 and AKM [1994] for details.

Table 5.1: OLS Results (Std. Dev. in Parentheses)

Variable	All Observations		Outliers Eliminated	
	Model 1	Model 2	Model 1	Model 2
Intercept	3.386	3.394	3.397	3.405
	(1.17E-3)	(1.17E-3)	(1.06E-3)	(1.06E-3)
Seniority	0.012	0.071	0.011	0.062
	(3.70E-5)	(3.38E-4)	(3.36E-5)	(3.06E-4)
$\text{Seniority}^2/100$	0	-1.132	0	-1.017
		(6.31E-4)		(5.74E-3)
$\text{Seniority}^3/1000$	0	0.643	0	0.583
		(3.93E-3)		(3.57E-3)
$\text{Seniority}^4/10000$	0	-0.109	0	-0.099
		(7.69E-4)		(6.99E-4)
Experience	0.057	0.045	0.056	0.046
	(2.97E-4)	(3.05E-4)	(2.70E-4)	(2.78E-4)
$\text{Experience}^2/100$	-0.249	-0.124	-0.244	-0.132
	(2.72E-3)	(2.78E-3)	(2.47E-3)	(2.54E-3)
$\text{Experience}^3/1000$	0.056	0.011	0.055	0.014
	(9.20E-4)	(9.40E-4)	(8.38E-4)	(8.56E-4)
$\text{Experience}^4/10000$	-5.13E-3	-2.99E-4	-5.02E-3	-5.61E-4
	(1.03E-4)	(1.05E-4)	(9.34E-5)	(9.52E-5)

Table 5.1: OLS Results (Continued)

Variable	All Observations		Outliers Eliminated	
	Model 1	Model 2	Model 1	Model 2
Elementary School	-0.043 (1.81E-3)	-0.046 (1.80E-3)	-0.040 (1.65E-3)	-0.041 (1.64E-3)
Junior High School	0.418 (2.62E-3)	0.396 (2.61E-3)	0.421 (2.38E-3)	0.400 (2.37E-3)
High School Grad	0.608 (3.00E-3)	0.584 (2.99E-3)	0.616 (2.73E-3)	0.592 (2.72E-3)
Basic Vo-Tech	0.236 (1.54E-3)	0.209 (1.54E-3)	0.233 (1.40E-3)	0.209 (1.40E-3)
Advanced Vo-Tech	0.571 (2.65E-3)	0.545 (2.64E-3)	0.573 (2.41E-3)	0.547 (2.40E-3)
Tech U / Undergrad	0.565 (2.56E-3)	0.557 (2.55E-3)	0.568 (2.33E-3)	0.559 (2.32E-3)
Graduate School	1.310 (2.93E-3)	1.294 (2.92E-3)	1.335 (2.68E-3)	1.319 (2.66E-3)
Male	0.201 (4.91E-4)	0.197 (4.90E-4)	0.199 (4.47E-4)	0.194 (4.46E-4)
Paris Region	0.138 (5.06E-4)	0.142 (5.03E-4)	0.141 (4.60E-4)	0.144 (4.57E-4)
	$R^2 = 0.277$	$R^2 = 0.284$	$R^2 = 0.316$	$R^2 = 0.323$
	$n = 4,784,284$		$n = 4,765,997$	

individual-specific means. I thus estimate an equation of the form

$$\log\left(\widetilde{y_{i,T_0+t}}\right) = \beta \widetilde{X_{i,T_0+t}} + \widetilde{\varepsilon_{i,T_0+t}} \tag{5.2}$$

where

$$\begin{aligned} \log\left(\widetilde{y_{i,T_0+t}}\right) &= \log\left(y_{i,T_0+t}\right) - \overline{\log\left(y_i\right)} \\ \widetilde{X_{i,T_0+t}} &= X_{i,T_0+t} - \overline{X_i} \\ \widetilde{\varepsilon_{i,T_0+t}} &= \varepsilon_{i,T_0+t} - \overline{\varepsilon_i} \end{aligned}$$

and the terms $\overline{\log(y_i)}$, $\overline{X_i}$, and $\overline{\varepsilon_i}$ refer to means of the relevant variables over all observations corresponding to individual i[21].I use the same data as used in estimating equation (5.1), thus descriptive statistics can be found once again in table 4.1. The results of the estimations of each of the two OLS regressions on each of the two data sets are found in table 5.2.

As has been noted by others, the estimated returns to seniority that one finds in the OLS estimates (coefficients on level seniority of 0.071 on the whole sample, 0.062 for the sample with outliers eliminated) are greatly reduced when one includes individual specific effects (coefficients on level seniority of 0.029 on the whole sample and 0.026 on the outliers-eliminated sample). However, these estimators do not make any allowance for firm specific heterogeneity in compensation, which the following three estimators (Topel, AKM and the cohort-based estimator) explicitly model.

5.2. The Topel Estimator and Its Relation to the AKM Estimators

I next calculated returns to job seniority using the estimator suggested in Topel [1991]. In the first step of the estimation process, I calculated first differences within individual-firm pairs for the outliers-eliminated data. Sample statistics for this data set appear in table 5.3 below[22]. As described in section 2.2.1 above, an equation of the form

$$\begin{aligned} \triangle \log\left(y_{i,T_0+t}\right) &= (\gamma_1 + \beta_1)\triangle t_{i,j,T_0+t} + \gamma_2 \triangle t^2_{i,j,T_0+t} + \gamma_3 \triangle t^3_{i,j,T_0+t} \\ &+ \gamma_4 \triangle t^4_{i,j,T_0+t} + \beta_2 \triangle Exp^2_{i,T_0+t} + \beta_3 \triangle Exp^3_{i,T_0+t} \\ &+ \beta_4 \triangle Exp^4_{i,T_0+t} + \triangle \varepsilon_{i,T_0+t} \end{aligned}$$

[21] I am assuming that ε_{i,T_0+t} is distributed such that not only is $E\left(\varepsilon_{i,T_0+t}\right) = 0$, but also $E\left(\varepsilon_{I,T_0+t} \mid I = i\right) = 0$ for all i.

[22] Note that the coefficient on differenced seniority is not exactly 1. This is because the data allowed us to determine what fraction of the year and individual was employed with the same firm. Thus factors that might allow contiued employment with the same firm in which seniority and experience increase by less that 1 each year (temporary layoffs, for example) might be the source of the difference. See Topel [1991] for a discussion of the impact that this might have on estimated returns to seniority.

Table 5.2: Individual Fixed-Effect Results (Std. Dev. in Parentheses)

Variable	All Observations		Outliers Eliminated	
	Model 1	Model 2	Model 1	Model 2
Seniority	2.35E-3 (4.72E-5)	0.029 (2.99E-4)	2.05E-3 (4.05E-5)	0.026 (2.57E-4)
$Seniority^2/100$	0	-0.460 (5.59E-3)	0	-0.416 (4.79E-3)
$Seniority^3/1000$	0	0.254 (3.43E-3)	0	0.231 (2.94E-3)
$Seniority^4/10000$	0	-0.044 (6.60E-4)	0	-0.040 (5.66E-4)
Experience	0.084 (3.01E-4)	0.078 (3.12E-4)	0.084 (2.59E-4)	0.078 (2.68E-4)
$Experience^2/100$	-0.364 (2.88E-3)	-0.301 (2.99E-3)	-0.359 (2.47E-3)	-0.303 (2.56E-3)
$Experience^3$ $/1000$	0.084 (1.01E-3)	0.062 (1.05E-3)	0.083 (8.71E-4)	0.064 (9.01E-4)
$Experience^4$ $/10000$	-7.27E-3 (1.18E-4)	-4.97E-3 (1.21E-4)	-7.18E-3 (1.01E-4)	-5.12E-3 (1.04E-4)
Paris Region	0.073 (1.15E-3)	0.074 (1.15E-3)	0.081 (9.93E-4)	0.081 (9.91E-4)
	$R^2 = 0.749$	$R^2 = 0.750$	$R^2 = 0.789$	$R^2 = 0.789$
	$n = 4,784,284$		$n = 4,765,997$	

Table 5.3: Descriptive Statistics for Data used in the First Step Topel Regression

Variable Name	Differenced Data	
	Mean	Std. Dev.
$\Delta LFRAISRE$	0.035	0.270
ΔSeniority/Experience	0.963	0.149
ΔSeniority^2/100	0.164	0.143
ΔSeniority^3/1000	0.355	0.471
ΔSeniority^4/10000	0.815	1.411
ΔExperience^2/100	0.361	0.233
ΔExperience^3/1000	1.382	1.488
ΔExperience^4/10000	5.556	7.897

was then estimated on the differenced data. This equation was estimated in two forms: with the parameter restriction $\gamma_2 = \gamma_3 = \gamma_4 = 0$ (only level seniority included, denoted Model T1A below) and without any parameter restrictions (Model T1B below). The results of these first-step models are shown in table 5.4 below.

Exploiting the identity $X_{i,T_0+t} = X_{i,T_0} + t$, Topel [1991] showed that one can separate the estimated coefficient on seniority into a component that can be attributed to seniority returns and a component that can be attributed to experience returns. This involves regressing the difference between current log compensation and predicted log compensation (according the results from the first step estimation) on total labor market experience at the date the job was started (Exp_{i,T_0}) and a vector of other individual specific variables that might be correlated with earnings. I thus estimated the model

$$\log\left(\widetilde{y_{i,T_0+t}}\right) = \beta_1 Exp_{i,T_0} + \theta F_{i,T_0+t} + e_{i,T_0+t}$$

where

$$\begin{aligned}\log\left(\widetilde{y_{i,T_0+t}}\right) = \log\left(y_{i,T_0+t}\right) &- \widehat{(\gamma_1 + \beta_1)} t_{i,j,T_0+t} - \hat{\gamma}_2 t^2_{i,j,T_0+t} - \hat{\gamma}_3 t^3_{i,j,T_0+t} \\ &- \hat{\gamma}_4 t^4_{i,j,T_0+t} - \hat{\beta}_2 Exp^2_{i,T_0+t} - \hat{\beta}_3 Exp^3_{i,T_0+t} \\ &- \hat{\beta}_4 Exp^4_{i,T_0+t}\end{aligned}$$

and F_{i,T_0+t} corresponds to the other individual-specific components. The results of estimations of this model based on the restricted (Model T2A below) and unrestricted (Model T2B below) first- step estimates appear in table 5.5 below.

Table 5.4: First Step Topel Model Results (Std. Dev. in Parentheses)

Variable	Differenced Data	
	Model T1A	Model T1B
Seniority + Experience	0.087	0.106
	(6.83E-4)	(7.44E-4)
$\text{Seniority}^2/100$	0	-0.551
		(9.32E-3)
$\text{Seniority}^3/1000$	0	0.300
		(6.21E-3)
$\text{Seniority}^4/10000$	0	-0.052
		(1.25E-3)
$\text{Experience}^2/100$	-0.297	-0.231
	(6.46E-3)	(6.60E-3)
$\text{Experience}^3/1000$	0.057	0.040
	(2.28E-3)	(2.34E-3)
$\text{Experience}^4/10000$	-3.78E-3	-2.26E-3
	(2.61E-4)	(2.66E-4)
	$R^2 = 0.020$	$R^2 = 0.022$
	$n = 2,517,026$	

Table 5.5: Second Step Topel Model Results (Std. Dev. in Parentheses)

Variable	All Observations (Outliers Removed)	
	Model T2A	Model T2B
Initial Experience	0.074	0.070
	(4.31E-5)	(4.17E-5)
Elementary School	3.225	3.242
	(3.20E-3)	(3.10E-3)
Junior High School	4.605	4.6036
	(4.97E-3)	(4.81E-3)
High School Grad	3.934	3.949
	(6.02E-3)	(5.82E-3)
Basic Vo-Tech	3.797	3.807
	(2.30E-3)	(2.22E-3)
Advanced Vo-Tech	2.618	2.669
	(5.34E-3)	(5.16E-3)
Tech U/Undergrad	3.548	3.607
	(5.01E-3)	(4.84E-3)
Graduate School	2.084	2.164
	(6.07E-3)	(5.87E-3)
Male	0.729	0.736
	(9.86E-4)	(9.52E-4)
Paris Region	0.408	0.395
	(1.06E-3)	(1.06E-3)
	$R^2 = 0.950$	$R^2 = 0.958$
	$n = 4,765,997$	

Topel [1991] noted that his estimators of joint returns to seniority and returns to initial experience would be biased as long as job changing was not exogenously generated. He suggested a procedure designed to indicate the direction of the bias. Table 5.6 below shows how the estimates of the joint returns to seniority and experience (in levels shown in table 5.4) generated by each of the two models (Model A with the parameter restrictions and Model B unrestricted) are broken down into seniority and experience components, as well as estimates of one of the components of the bias. A positive value for the bias term suggests that true returns to level experience are lower than estimated, and thus returns to seniority are higher than estimated.

Table 5.6: Estimated Returns Using the Topel Two-Step Method

Model Type	Total Returns	Seniority	Experience	Wage Growth Bias
Model A	0.087	0.013	0.074	-0.071
Model B	0.106	0.036	0.070	-0.058

These results suggest that, consistent with the OLS and the fixed effects estimators, including the higher order terms in seniority has a dramatic effect on estimated returns to seniority. The estimated coefficients suggest profiles that rise more or less rapidly, according to the estimation procedure, and later flatten out as seniority increases. The largest estimated level returns to seniority come from the OLS procedure (6.2 percent per year on the outliers-eliminated sample), while simple individual fixed effects causes this estimate to drop to 2.6 percent per year. Controlling simultaneously for individual and firm effects (as Topel's procedure does) causes seniority returns to rise slightly (to 3.6 percent per year), although the wage growth bias estimates suggest that this might be illusory.

All of these procedures, however, have found significant (however small) returns to seniority in the fully saturated specifications on the outliers-eliminated data. All of these procedures share a common problem, however; they assume that all firms reward seniority in the same manner. If this assumption were to be violated, the estimated returns to seniority would be a sort of weighted average of the firm-specific returns. Thus even if only a few firms in the sample rewarded seniority and these firms contributed a high enough number of observations to the data set, we would run the risk of estimating positive economy-wide returns, and we might incorrectly assume that the majority of firms in the economy provided positive returns to job seniority.

AKM [1994] show that this is false. Their estimation procedure allows them to account for individual- and firm- specific effects on earnings, as well as estimating firm-specific returns to seniority. Using the same data as I use here (the outliers-eliminated sample) and using a linear spline specification, they find no significant returns to seniority (-3.37E-5 for men and 8.28E-4 for women across individuals). However, when looking across firms, they find a standard error in returns to seniority that is relatively large (0.077 relative to a mean across firms of 2.7E-3) and encompasses all of the above estimated returns within one standard deviation of the mean estimated returns to seniority. This implies that there is substantial heterogeneity in estimated returns to seniority. If this estimated heterogeneity corresponds to actual differences in compensation policies, the self-selection problem discussed in section 2 arises, suggesting that even these estimators might be biased (see section 2.2.2).

5.3. The Cohort-Based Estimator

With the knowledge that there is substantial across-firm heterogeneity in estimated returns to job seniority, I applied the procedure described in section 3 to calculate the cohort-based estimator of "true" returns to job seniority. Because the procedure is particularly demanding on the data, I was not able to obtain large enough sample sizes for every cohort in every firm in the data. Thus each observation was assigned to one of three subgroups: observations for which there was sufficient information to calculate cohort-firm specific $\hat{\gamma}_{j,T_0}$ terms for use in the regression described by equation (3.3), observations for which there was not enough information to calculate a cohort-firm specific $\hat{\gamma}_{j,T_0}$ terms but for which one could calculate a firm specific $\hat{\gamma}_j$, and observations for which I could calculate neither firm nor cohort-firm specific $\hat{\gamma}_{j,T_0}$ terms, but for whom I could calculate cohort specific (but not firm specific) $\hat{\gamma}_{T_0}$ terms for use in the regression described in equation (3.3)[23]. The cohort-firm group contained 3,471,425 observations, the firm only group 701,892 observations and the cohort only group 1,131,791 observations. This distribution of observations across groups is encouraging in two respects. First, for 79 percent of the observations I am able to estimate at least a firm specific $\hat{\gamma}_j$, and 83 percent of these obser-

[23] Clearly, observations corresponding to individuals employed in larger firms are more likely to be found in the first and second groups, whil observations corresponding to smaller employers are more likely to be found in the third group. Unfortunately there is no way to recover cohort-firm specific $\hat{\gamma}_{j,T_0}$ when there is not enough data present. Thus the estimation strategy implicitly imposes a sort of "ocean of small, identical firms" constraint, in which small firms are assumed to reward seniority identically, although this offered rate of return still competes with that of the larger firms in the eyes of potential employees for the purposes of self-selection.

vations can be used to recover the relevant cohort-firm specific $\hat{\gamma}_{j,T_0}$ terms needed to estimate the true γ_j. Second, for all of the observations in the 21 percent for which neither $\hat{\gamma}_j$ nor $\hat{\gamma}_{j,T_0}$ is directly estimable, I am able to recover cohort specific $\hat{\gamma}_{T_0}$, although the estimated $\hat{\gamma}_{T_0}$ will be a weighted average of the component firms' $\hat{\gamma}_{T_0}$ terms[24].

The AKM [1994] procedure was followed up to the point where they estimate firm-specific intercepts and seniority splines. Thus the estimated effects of education, total labor market experience, sex and region on the log of total real annual compensation are unchanged. I then estimated the first part of the cohort estimator (equation (3.1) in section 3). For the cohort-firm group (group 3 in table 5.7), I estimated equation (3.1) entering cohort by entering cohort, and for the firm only group (group 2 in table 5.7) I pooled all observations within the same firm but retained the firm specific estimate[25]. The estimated coefficients were merged back in with the seniority data and equation (3.3) was estimated firm by firm. For the cohort only group (group 1 in table 5.7), I estimated equation (3.1) by pooling all observations within the same cohort, independent of firm, and running the regression cohort by cohort. These estimated coefficients were merged back in with the seniority data and equation (3.3) was estimated over all observations in the cohort only group.

Using the cohort-based procedure suggested in section 3, I find that the equally-weighted mean of the "true" returns to seniority across all firms in the sample is 6.31E-3, although the standard deviation of estimated true returns across all firms in the sample is 0.120. This suggests that there is a substantial amount of heterogeneity in the returns to seniority that firms propose to their workers. Looking at the subsets of firms for which I was able to recover firm specific true returns, similar results become apparent. Over the entire set of 94,453 firms for which true returns were firm specific, the equally-weighted mean estimated true returns are -1.84E-4 with a standard deviation of 0.282. On the subset of 83,424 firms for which the estimated true returns involved the use of cohort-firm specific $\hat{\gamma}_{j,T_0}$ (the subset most closely meeting the requirements of the econometric procedure), the equally-weighted mean was -7.89E-4 and the standard deviation was 0.289. This suggests that the pooling techniques used to calculate $\hat{\gamma}_j$ and $\hat{\gamma}_{T_0}$ generate a positive, but very small, bias in estimates of true returns to seniority, although the relatively large standard errors render even such

[24] Although I was able to recover firm specific $\hat{\gamma}_{j,T_0}$ or $\hat{\gamma}_j$ for 79 percent of the observations, these observations covered only 22 percent of the firms in the sample (94,603 out of 521,182). This is not surprising since the sampling scheme was designed to be random with respect to individuals and not firms, and thus the distribution of workers in the sample across employers should be representative of the distribution of employment, and not necessarily employers, in the economy as a whole.

[25] This is essentially what is done in AKM [1994].

Table 5.7: Results Using the Cohort-Based Estimator

Variable	All Firms	Groups 1^a	$2^b \& 3^c$	2^b	3^c
True $\hat{\gamma}_j$					
Mean $(\hat{\gamma}_j)$	6.31E-3	7.75E-3	-1.84E-4	4.39E-3	-7.89E-4
Std. Dev. $(\hat{\gamma}_j)$	(0.120)	0	(0.282)	(0.226)	(0.289)
Mean (Var $(\hat{\gamma}_j)$)	3.09E-4	6.16E11	1.70E-3	0	1.93E-3
Std. Dev. (Var $(\hat{\gamma}_j)$)	(0.158)	0	(0.372)	0	(0.396)
Difference $\hat{\gamma}_{1j} - \hat{\gamma}_j$					
Mean $(\hat{\gamma}_{1j} - \hat{\gamma}_j)$	-3.32E-3	-4.47E-3	1.86E-3	-8.30E-4	2.22E-3
Std. Dev. $(\hat{\gamma}_{1j}$-$\hat{\gamma}_j)$	0.119	0	0.279	0.223	0.286
Pct. $(\hat{\gamma}_{1j} - \hat{\gamma}_j) > 0$	9.8%	0%	53.8%	53.8%	53.8%
Correlations					
$\hat{\gamma}_j$ with $\hat{\gamma}_{1j}$	0.147	-	0.146	0.182	0.143
$(\hat{\gamma}_{1j} - \hat{\gamma}_j)$ with $\hat{\gamma}_j$	-0.984	-	-0.984	-0.972	-0.985
Pct. Reject:					
T-test $\hat{\gamma}_j = \hat{\gamma}_{1j}$	88.1%	100%	34.5%	51.9%	32.2%
Nb. of Firms	521,032	426,579	94,453	11,029	83,424

[a] Group 1 refers to cohort-only estimates of true γ_j. AKM[1994] assigned one value to all firms in this group, and the cohort-based estimator did likewise.
[b] Group 2 refers to firm-only estimates of the true γ_j. Differences between true γ_j and $\hat{\gamma}_{1j}$ among group 2 firms are due exclusively to different critical levels (AKM [1994] used 10 observations, I use 4).
[c] Group 3 refers to cohort-firm estimates of the true γ_j.

weak conclusions suspect.

Furthermore, comparisons with the results obtained by AKM [1994] are instruc-tive[26]. I find that, for 88.1 percent of the firms in the data set, the cohort based estimator generates results significantly different from AKM's results for level earnings. Perhaps surprisingly, the cohort-based estimator is usually larger than the AKM level estimator. This suggests that, instead of the more productive workers leaving later, they may hazard out of their current employers at a faster rate than do the less productive workers. However, a closer examination suggests that this effect relies heavily on the pooled estimates, since 53.8 percent of the firms with cohort-firm estimated $\hat{\gamma}_j$ show signs of positive bias.

Some interesting results come out of analyzing the correlation between the estimated returns to seniority under both methods. First, the estimated correlation between the cohort-based estimates and the AKM estimates, although positive, is small (0.147 in the whole sample, 0.143 for the cohort-firm estated enterprises). This is to be expected if the estimated returns in firms with smaller true returns are more heavily biased than those in firms with with larger true returns. This prediction of the theoretical model is independent of the relative hazard rates (provided they are not identical). It is also supported by the data, in that the difference in estimated returns decreases as true returns increases (correlation coefficient of -0.984 overall and -0.985 for the cohort-firm based estimates). Not only does this imply that variance in cohort-based estimator far dominates variance in the AKM estimator (even though the estimated variances at the firm-by-firm level are not very different), but also that the two estimators are farthest apart when true returns are lowest, which is consistent with the idea that the firms that reward seniority the least will have the largest bias in their estimated returns according to an econometric technique that does not account for worker selection.

6. Conclusion

This paper has shown that one cannot ignore firm heterogeneity and worker self-selectionwhen estimating returns to seniority. After developing a theoretical framework from which to explain the source of a potential bias in estimators of returns to seniority that to not take into account worker self-selection based on heterogeneity in compensation policies and subsequent learning, a new estimator of "true" returns to seniority was developed. The results generated by this estimator were compared to results generated by

[26] In table 5.7, estimates using the cohort-based estimator are reffered to as $\hat{\gamma}_j$, while estimates using the AKM [1994] estimator are referred to as $\hat{\gamma}_{1j}$.

an estimator proposed by Abowd, Kramarz and Margolis (AKM) [1994], as well as to estimators that do not allow for heterogeneous seniority reward policies, such as OLS, individual fixed effects and Topel's [1991] two step estimator. The homogeneous returns estimators all signalled significantly positive returns to seniority, although the level of these returns was rather variable. The new cohort-based estimator, on the other hand, showed (like AKM's estimator) mean zero returns across firms, but significant firm-levelheterogeneity. As predicted by the theoretical model, the bias in the AKM estimator was found to be largest for firms for whom the "true" returns to seniority were smalles, suggesting that any estimator, even one that takes into account across-firm heterogeneity in seniority reward policies, will be biased unless one also accounts for work self-selection to take advantage of the diversity of compensation policies proposed by the market.

References

Abowd, John, Francis Kramarz and David Margolis (1994). "High Wage Workers and High Wage Firms," *Cornell University Working Paper*, March.

Abraham, Katherine and Henry Farber (1987). "Job Duration, Seniority, and Earnings," *American Economic Review*, 77, June.

Altonji, Joseph and Robert Shakotko (1987). "Do Wages Rise with Job Seniority?" *Review of Economic Studies*, 54, July.

Lazear, Edward P (1979). "Why Is There Mandatory Retirement?" *Journal of Political Economy*, 87:1261-1284, December.

Margolis, David (1994). "Labor Market Signalling, Self Selection and Returns to Seniority," in progress.

Rothschild, Michael and Joseph Stiglitz (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90, November.

Salop, Joanne and Steven Salop (1976). "Self Selection and Turnover in the Labor Market," *Quarterly Journal of Economics*, 90, November.

Topel, Robert (1991). "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," *Journal of Political Economy*, 99:1.

Weiss, Andrew and Ruqu Wang (1990). "A Sorting Model of Labor Contracts: Implications for Layoffs and Wage-Tenure Profiles," *NBER Working Paper No. 3448*, September.