The Evolution of Earnings in an Internal Labour Market: A Counterfactual Analysis

Tim Barmby[1] and Barbara Eberth (University of Aberdeen)

September 2004

PRELIMINARY – Please do not quote

Abstract: This paper uses the DiNardo, Fortin and Lemieux (1996) methodology to analyse the evolution of the earnings of a cohort of workers within an Internal Labour Market. It finds evidence of downward bias in the returns to tenure suggested by Stephens (2003) and also suggests a possible explanation of some of the findings of Medoff and Abraham (1981)

JEL classification J24,J31

Keywords: Earnings, Tenure, Counterfactuals

I. Motivation

Studying the way earnings distributions evolve within internal labour markets (ILM's) can, potentially, reveal a number of things about the way organisations seek to reward and give incentives to their workers. We would expect to observe earnings growth based on the idea that workers accumulate specific human capital, Becker (1993), and that firms will pay for this increased productivity. Firms' may also choose to construct wage tenure profiles which are steeper than the growth in productivity, Lazear (1979), and this will also have implications for observed distributions, but the

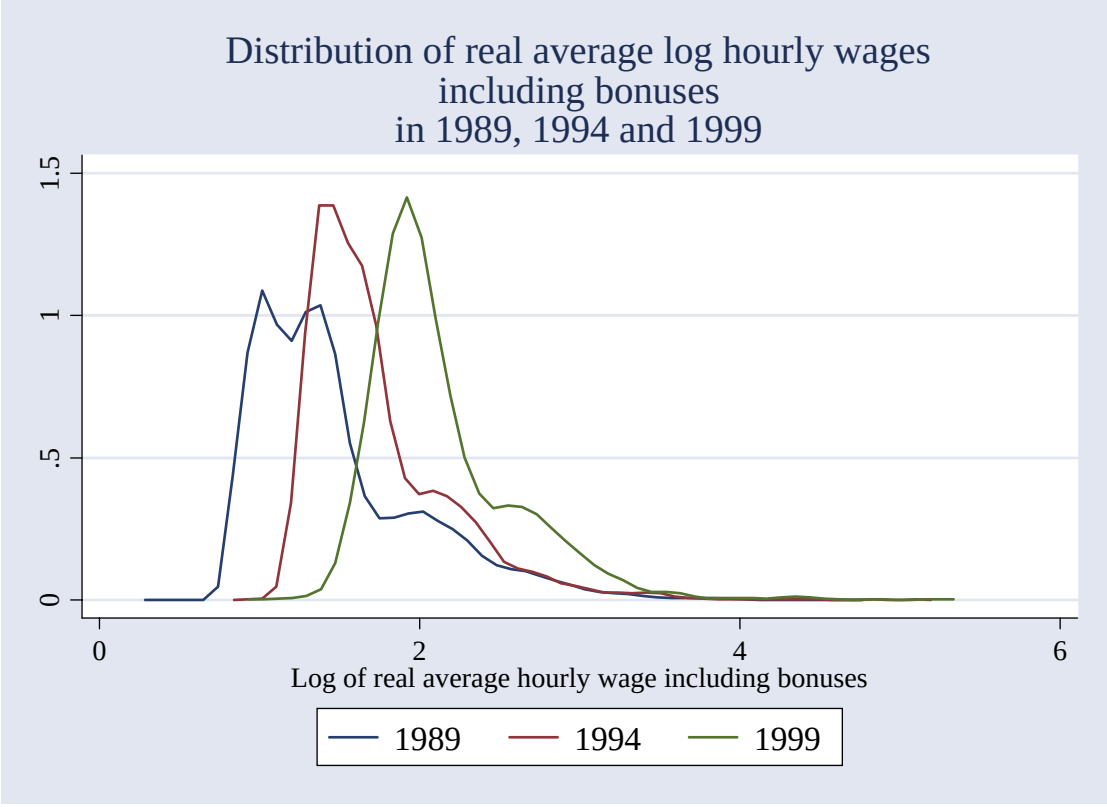
[1] The authors would like to acknowledge helpful comments from Donal O'Neill. Address for correspondence:- Tim Barmby, Centre for European Labour Market Research, Department of Economics, University of Aberdeen, Old Aberdeen, Scotland AB24 3SQ tim.barmby@abdn.ac.uk, b.eberth@abdn.ac.uk

underlying link will be to the flow of productivity of the worker. Also reflected in the earnings distributions that we observe for workers within firms will be aspects of worker's behaviour in terms of their decisions to stay with or leave the organisation. This is the focus of the present paper, and in particular, in examining the distribution of cohorts of workers, especially if workers exit their cohort in a non-random fashion then this will impact on the subsequent distribution of earnings for that cohort.

II. Some Distributions

Our paper will examine the evolution of earnings for a particular cohort within the ILM of a large financial sector firm. The data will be described in more detail in the next section, but to illustrate the basic idea of this paper consider the following distributions:-

Figure 1



What we are looking at here is the initial distribution of the natural log of hourly earnings (plus bonuses) of the cohort of workers who entered the organisation in 1989, and then the same distribution of those who remained in the organisation 5 years later in 1994 and then 10 years later in 1999.

Clearly two main influences will have been at work; firstly workers within the organisation will accumulate specific human capital which will increase their productivity and result in earnings growth. Secondly workers are in a market for their services, and may receive (or indeed seek out) alternative offers from other firms in the market, if these offers dominate existing remuneration then they will move. As a consequence of this they will obviously not remain in the cohort and will not be represented in earnings distributions subsequent to the first. We will examine the factors which influence the transition between firms in section V.

The question we pose ourselves is whether we could draw the distribution of earnings that would have existed if exits from the firm hadn't been systematic (or put another way; had been random) because in this scenario the movement of the distribution will reflect the way the productivity and, we are assuming, correspondingly the reward of the workers in the initial cohort would have evolved.

We seek to do this by constructing the distribution that would have prevailed at the subsequent time period if the (distribution) of characteristics had remained as it was initially, since if exit is random the distribution of characteristics will, in expectation, remain the same. The constructed distribution will also reflect other wage setting devices such as back-loading payments to give workers the incentive not to "shirk" early in their careers, (Lazear's Delayed Payment hypothesis) as well as specific capital accumulation, but we don't attempt to disentangle these two things in this paper.

III Data

The data comes from personnel records of all employees of a large financial sector firm based in the UK covering the period January 1989 to November 2001 allowing for a potential total of 154 monthly observations for each employee in the firm. In this paper information on real average log hourly wages including bonuses and other employee characteristics for a subset of full-time employees who started employment with the firm in 1989 is used in the empirical analysis that follows.

Table 1	Composition of employees of cohort 1989 (%) by tenure												
	Tenure in years												
	0	1	2	3	4	5	6	7	8	9	10	11	12
Number of employees	4,475	3,998	3,192	2,862	2,608	2,399	2,186	1,954	1,711	1,520	1,389	1,295	1,189
Male employees	39.08	39.57	39.51	39.34	38.77	37.97	37.65	38.23	39.74	40.39	41.68	43.71	45.08
Female employees	60.92	60.43	60.49	60.66	61.23	62.03	62.35	61.77	60.26	59.61	58.32	56.29	54.92
Leavers (all)	10.66	20.16	10.34	8.87	8.01	8.88	10.61	12.44	11.16	8.62	6.77	8.19	-
Leavers (men)*	35.01	39.85	40.60	46.01	44.78	41.89	31.40	27.63	33.98	26.42	16.07	27.97	-
Leavers (women)*	64.99	60.15	59.40	53.99	55.22	58.11	68.60	72.37	66.02	73.58	83.93	72.03	-
15-19 years old**	47.53	47.62	49.00	49.90	50.84	52.48	53.11	53.28	52.66	53.42	53.42	52.20	52.06
20-24 years old	22.26	22.14	20.93	20.44	20.36	20.26	19.85	19.40	19.05	18.29	17.21	17.76	17.07
25-34 years old	20.92	21.09	20.83	20.51	20.09	18.42	17.93	17.91	18.64	18.49	19.37	19.85	20.19
35-44 years old	7.17	7.03	7.36	7.41	7.17	7.21	7.46	7.88	8.18	8.42	8.57	8.88	9.59
45-54 years old	1.85	1.90	1.66	1.61	1.46	1.63	1.65	1.54	1.46	1.38	1.44	1.31	1.09
55-59 years old	0.22	0.20	0.22	0.14	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
60-65 years old	0.04	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
White	83.69	85.12	87.53	89.34	90.68	90.66	90.81	90.89	91.29	91.12	91.22	91.58	91.59
Asian/Asian British	0.02	0.03	0.06	0.10	1.73	2.75	2.84	2.76	2.86	2.96	3.02	3.24	3.70
Black/Black British	1.43	1.50	1.66	1.68	1.92	1.79	1.92	2.00	2.10	2.37	2.16	2.16	2.35
Chinese/Ethnic	0.00	0.00	0.00	0.07	0.65	0.71	0.69	0.72	0.70	0.66	0.86	0.93	0.93
Unstated ethnic origin	14.86	13.36	10.75	8.81	5.02	4.09	3.75	3.63	3.04	2.89	2.74	2.08	1.43
Degree	11.82	12.83	14.13	14.19	14.26	13.67	13.45	13.41	13.68	14.28	14.33	15.14	14.89
Further Education	1.79	1.95	2.38	2.66	2.95	3.21	3.57	3.68	3.86	3.75	3.96	4.17	4.04
A-level	14.12	14.23	14.07	14.54	14.72	15.01	15.42	15.40	16.07	16.12	16.56	16.29	16.74
O-level	32.96	34.27	37.63	39.27	40.99	43.60	45.43	45.34	44.07	43.55	42.76	42.24	41.72
Other education	8.56	9.50	11.50	12.68	13.42	14.34	15.23	15.86	15.96	16.05	15.91	15.44	15.73
Unknown education	30.75	27.21	20.30	16.67	13.65	10.17	6.91	6.29	6.37	6.25	6.48	6.72	6.90
Staff on entry	83.80	83.02	83.43	84.49	85.28	86.91	87.83	87.51	87.55	88.03	88.12	87.57	87.55
Managerial on entry	16.20	16.98	16.57	15.51	14.72	13.09	12.17	12.49	12.45	11.97	11.88	12.43	12.45

** This is calculated over all leavers, ** The variable here is age-tenure*

The internal hierarchy of this firm is discussed in detail in Treble, van Gameren, Bridges, and Barmby (2001). The firm has a well defined structure of 12 levels that can be divided into four broad categories comprised of training levels, clerical levels, middle managers and senior managers. Discussions with the HR officers responsible for maintaining and updating the personnel records of all employees show that the firm makes a broad distinction between those employees classified as staff and those who are classified as managers. “In house”, the firm is used to grouping the 12 levels into 7 broad categories comprised of an induction level (S01), junior staff levels (S02 and S03), senior staff levels (S04 and S05), junior management levels (M93 and M94), the middle management level (M95), the senior management level (M96), and the executive management levels (M97-M99).

Table 1 summarizes the composition and some characteristics of employees in cohort 89 by tenure. All in all 4475 employees are recruited by the firm in 1989 of which 39.08% are men and 60.92% are women. Throughout the period more women tend to leave the firm than men resulting in a more even gender balance towards the end of the period. Comparing the percentage of men in the cohort in 1989 of 39.08% to that in 2001 of 45.08% signifies a large change in the gender balance in physical terms. By 2001 the initial cohort size has shrunk to 1189 employees who remain to be working for the firm giving an exit rate of 73.4% over time. Of all leavers 36.60% are men and 63.40% are women.

The age dummies are constructed as age-tenure (or alternatively age at entry). This will allow the reader to see changes in the age composition of the cohort without having to take account of the fact that with each additional year of tenure the worker ages by one year. The age composition of the cohort on entry to the firm clearly highlights the firms demand for workers of school leaving age who have completed either their GCSE’s or A-levels. Nearly 50% of all recruits fall within this age bracket of 15-19 year olds, followed by 22.26% of recruits between the ages of 20-24 who could have potentially obtained a degree and 20.97% of recruits between the ages of 25-34. Only very few new entrants are hired who are above 34 years of age. The mean age of all new entrants is 22.9 years and as table 1 reveals, the composition of the cohort in terms of age, does not change very much. This is confirmed by an average age of 34.7 in 2001.

Although some workers have refrained from revealing their ethnic origin, it is clear from table 1 that the vast majority of workers (83.69% on entry) in the cohort primarily originate from a white background and that only 3.67% of entrants come from other ethnic backgrounds and 8.22% did not disclose their ethnic origin to the firm in 1989.

Information on qualification is only available for part of the sample, reflected in the high proportion of employees who display "Other" qualification (8.56%) and those employees who did not state their qualification on entry to the firm (30.75%). Just to give the reader a flavour, the category "other" includes such various qualifications as the Duke of Edinburgh Award, Beginners Italian, Beginners Excel, and others which are not necessarily schooling qualifications. Discussions with our liaison officers in the firm revealed that the missing information on schooling qualification is a consequence of a combination of bad questionnaire design and human error in entering the qualification data into the personnel databank of the firm[2]. The information on qualifications in table 1 is organised such that the qualification variables reflect highest qualification on entry.

It was already highlighted earlier that the firm predominately recruits workers in 1989 aged 15 to 19 which is directly reflected in the high proportion of new entrants with highest attainment at O-levels (32.96%) and A-levels (14.12%). Of those recruits who have stated their schooling on entry 11.82% hold a degree and only a small fraction of 1.79% have completed other further education courses. For 38.30% of leavers information on qualifications is unknown. But we do know that 29.68% have highest attainment at O-levels, 13.09% have A-levels, 10.99% hold a degree, 1.15% have attended further education courses and 6.87% hold other qualifications.

The last two rows in table 1 describe the cohort in terms of whether employees have been assigned to a staff or managerial position on entry to the firm.

[2] The bank only started to collect employees' qualifications data in 1995 when analysts in the bank were specifically asked to incorporate information on employee qualifications amongst others in a particular report to the bank. The questionnaire design was such that employees were asked to list their qualifications with no particular reference to schooling or other qualifications obtained.

Unsurprisingly, 90.36% of recruits entered the firm as staff in 1989 whereas only 9.64% were hired into managerial levels. Of the 90.36% recruited into staff grades, 67.35% were women compared to 32.65% who were men. On the other hand, of those 9.64% new entrants recruited into managerial levels, 80.62% were men and only 19.38% were women. Of all leavers, 82.75% entered as staff and only 17.25% as managers. Not surprisingly 73.58% of those who left and entered the firm into managerial positions are men compared to 26.42% of their female counterparts holding a managerial job on entry. Again, the picture reverses for those who left the firm and started their career at a staff level of which 28.9% are men and 71.1% are women.

IV. Empirical Analysis

Our empirical problem will be to interpret the way in which earnings distributions within the organisation change. For each individual worker we will observe *w*, the natural log of their hourly earnings, some characteristic(s), *z* and the date at which we observe these, *t*.

A basic conceptual device in this analysis is the idea that each observation is a realisation of a random vector (*w*,*z*,*t*) which is drawn from a joint distribution F(*w*,*z*,*t*) of wages *w*, (an) individual attribute *z*, and a date *t*. The reader is referred to DiNardo, Fortin, and Lemieux (1996) for a full discussion of this methodology of constructing counterfactual densities.

As a starting point we consider the *marginal* distribution of wages at a point in time t=1, say

$$f(w \mid t=1) = \int_z f(w, z \mid t=1) dz \qquad (3.1)$$

Since $f(w \mid z, t=1) = \dfrac{f(w, z \mid t=1)}{f(z \mid t=1)}$ and $\dfrac{dF(z)}{dz} = f(z)$ we can write

$$\begin{aligned} f(w \mid t=1) &= \int_z f(w \mid z, t=1) f(z \mid t=1) dz \\ &= \int_z f(w \mid z, t=1) dF(z \mid t=1) \end{aligned} \qquad (3.2)$$

Now it is clear that everything here refers to period 1, however if we wrote the RHS of the above as

$$\int_z f(w \mid z, t=2) dF(z \mid t=1) \qquad (3.3)$$

we would be constructing a *counterfactual* density of the wages which would have prevailed in t = 2 if the distribution of characteristics had remained as they were at t=1. Notationally we can indicate this (following DiNardo) as

$$\int_z f(w \mid z, t=2) dF(z \mid t=1) = f(w \mid t_w = 2, t_z = 1) \qquad (3.4)$$

The question is: how might we obtain this density ? The answer is surprisingly simple and elegant. We only need to rewrite the above as

$$\begin{aligned} f(w \mid t_w = 2, t_z = 1) &= \int_z f(w \mid z, t=2) \frac{dF(z \mid t=1)}{dF(z \mid t=2)} dF(z \mid t=2) \\ &= \int_z f(w \mid z, t=2) \psi(z) dF(z \mid t=2) \end{aligned} \qquad (3.5)$$

It is clear from the above that the *counterfactual* density of interest is simply the density you would plot at t = 2 except that the observations at t = 2 are *reweighted* according to the function

$$\psi(z) = \frac{dF(z \mid t=1)}{dF(z \mid t=2)} \qquad (3.6)$$

Intuitively the adjustment of the density at t=2 essentially amounts to taking account of *non-random attrition*. This idea is clearer if we consider

$$\psi(z) = \frac{dF(z \mid t=1)}{dF(z \mid t=2)} = \frac{\frac{dF(t=1 \mid z)dF(z)}{dF(t=1)}}{\frac{dF(t=2 \mid z)dF(z)}{dF(t=2)}} = \frac{\Pr(t=1 \mid z)}{\Pr(t=2 \mid z)} \frac{\Pr(t=2)}{\Pr(t=1)} \qquad (3.7)$$

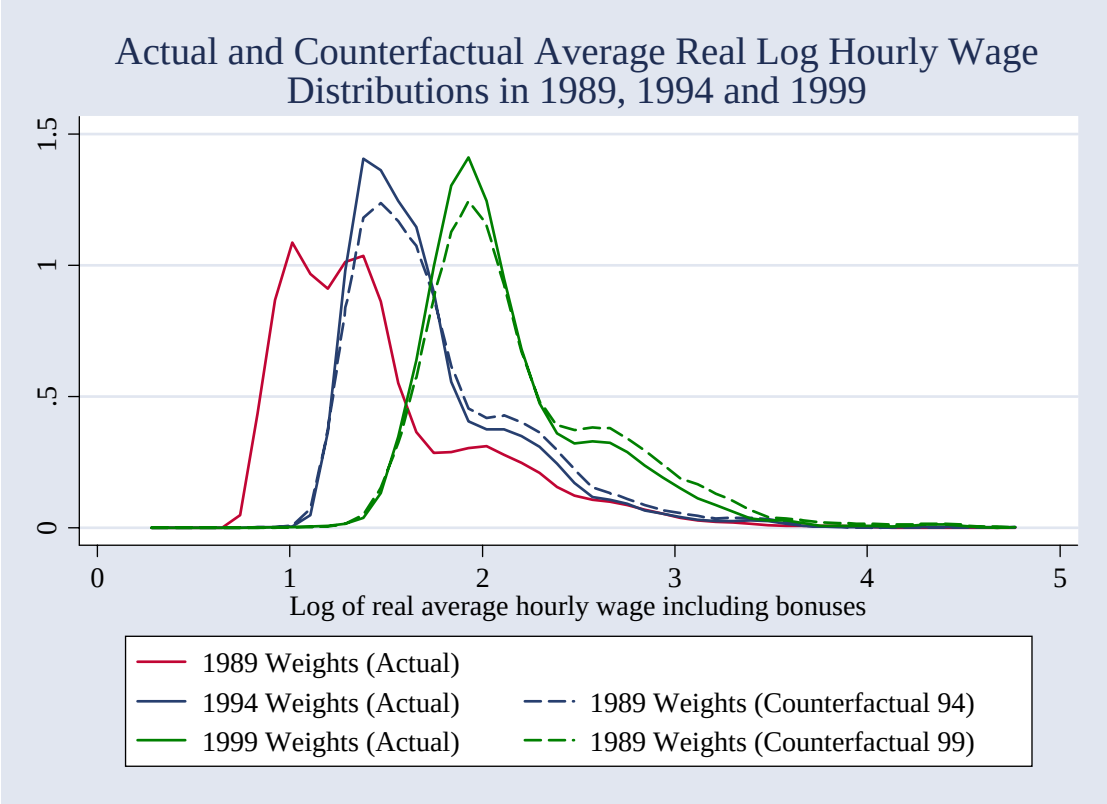
$\Pr(t=k \mid z)\ k=1,2$ is just an *attrition* probability which can be estimated by a conventional logit or probit model. Note these are not exactly probabilities of individual's remaining with the firm t= 1,2,.. years etc, as the sample space would be N, the initial number of individuals in the cohort. Rather it is the probability of observing the date t = 1,2,… etc *given* particular characteristics, and the sample space would be the individual/date points. Note if there was no systematic transition out of the firm, $\Pr(t=k \mid z) = \Pr(t=k) \forall k$ and $\psi(z) = 1$[3].

VI Counterfactual Distributions

Applying the methodology outlined in section IV to our data yields the following results. Consider figure 2 where the three solid densities are the actual densities of log hourly wages which are observed on entry in 1989 and then subsequently in 1994 and 1999 (after 5 and 10 years respectively), we have already seen these in figure 1. The dotted lines are the adjusted densities which would have prevailed if the distribution of characteristics had remained as on entry in 1989, or stated another way if the exits from the firm had been random.

[3] *The Probits used in constructing the weights ψ(z)* are reported in Appendix 1

Figure 2



The empirical results of this exercise consistently suggest that the observed (unadjusted) distribution appears to be under representing the growth in earnings that would have taken place within the firm if the transitions out of the firm had been random. Put another way there appears to be empirical evidence of a negative bias in the returns to tenure.

To put some figures to the extend of the bias, consider Table 2 which computes the expected hourly wages in pound sterling for the actual and counterfactual hourly wage distributions corresponding to figure 2. It is clear from rows two and three that the actual average wages observed in 1994 and 1999 fall short of average counterfactual wages as was already graphically shown in figure 2. The second row in table 2 also shows average wage growth over time. Between 1989 and 1994 average hourly wage growth is 22.4% comparing the actual distributions in 1989 and 1994 with an implied annual growth of 4.13%. If characteristics of employees had stayed as they were on entry in 1989, average wage growth would have reached 31.61% over the five year period, resulting in an annual growth rate of 5.65%. This

implies a downward bias of 1.52%. Actual average wage growth observed for the cohort over a ten year period is 82.32% (annual rate 6.19%) but would have been 101.72% (annual rate 7.27%) if the distribution of characteristics had remained at their entry level in 1989, giving a discrepancy of 1.08% pa.

Table 2

	Average 89	Actual Mean 94	Counterfactual Mean 94	Actual Mean 99	Counterfactual Mean 99
E(w)	5.312	6.504	6.991	9.686	10.716
Wage growth		22.4%	31.61%	82.34%	101.72%
Implied annual growth rate in w		4.13%	5.65%	6.19%	7.27%

The next section considers theoretical reasons why this might be the case

V Theory

The notion, which we hinted at in section I, that to fully understand observed wages in firms we need to understand the process by which external offers are made, internal counter offers might be forthcoming and which of these are accepted, was discussed in Lazear (1986) and more recently by Stevens (2003)[4].

In essence we need to consider what drives the wage offers of the firm the worker is working for and also alternative firms that he/she might work for. To do this we consider the worker's productivity in firm 0 (the firm he/she presently works for). We follow similar notation to that used by Stevens in setting out this theoretical framework.

$$v_0 = g + k + \varepsilon_0 \qquad (4.1)$$

Here productivity is v, general human capital g, and specific human capital k. There are a number of alternative employers $i = 1,\ldots,n$ in which the workers productivity is described by the equation

[4] See also Burdett and Coles (2003) and Stevens (2004), for related discussions of matching models.

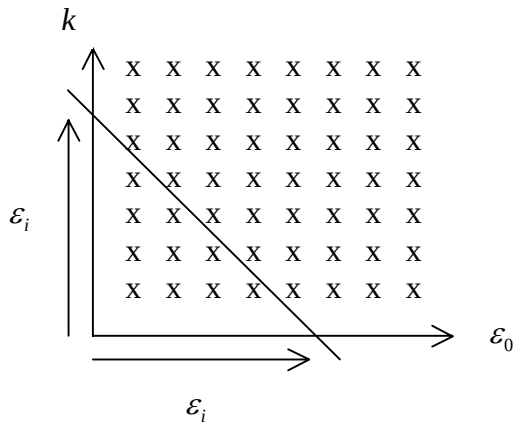
$$v_i = g + \varepsilon_i \tag{4.2}$$

In both of these equations ε represents the quality of the worker's match with the particular employer. The match quality is distributed

$$\varepsilon \sim f(\varepsilon); \quad \varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}] \tag{4.3}$$

Each worker in a particular cohort of workers employed by a given firm will have a probability of receiving an offer from an alternative employer in any given period. Assuming that the existing employer and the prospective employer observe their specific match quality with certainty then the worker if he/she receives an offer will remain with the existing employer if $v_i < v_0 \Rightarrow g_i + \varepsilon_i < g_0 + k_0 + \varepsilon_0$. Assuming that $g_i = g_0$, this implies that $\varepsilon_0 + k > \varepsilon_i$. This is describing a non-random selection from the distribution of match quality. The process describing how workers transit between firms will induce a *negative* correlation between ε_0 and *k*. To see this more clearly consider the following diagram

Figure 3



The points within the axes represent drawings from the joint distribution of *k* and ε_0 prior to any outside offers. If the distributions of *k* and ε_0 are independent there will be no correlation between these two quantities. Consider an outside offer with match quality ε_i being made, if the individual worker's (k, ε_0) lies above and to the

right of the offer line then the worker will remain with the firm, below and to the left he will transit between firm 0 and *i*. Consider many such offers being made, some of these will of course be accepted and some rejected. It is the rejected offers we are particularly interested in here and it appears that the effect will be to induce a negative correlation between *k* and ε_0 in the stayer group. The effect of this negative correlation will be to induce a *negative* bias on a regression estimation of earnings on tenure for a given cohort.

VI. Concluding Remarks

Our findings support the theoretical conjecture of Stevens (2003). We also offer the thought that the approach taken in this paper offers a way of explaining the apparent puzzle mentioned in Medoff and Abraham (1981) where the introduction of performance indicators into an OLS earnings equation appears to increase the estimated coefficient on tenure (this is a puzzle if tenure and performance are positively correlated; as exclusion of the performance ratings would result in tenure acting as a proxy for them and inflated the estimated coefficient; inclusion therefore should *reduce* the estimated coefficient). However in the framework used in this paper there is an *existing* negative bias on tenure, the performance ratings added to the equation will act as a proxy for match quality and reduce some of the negative bias.

References

Becker Gary S (1993) Human Capital 3rd Edition, University of Chicago Press.

Burdett Ken and Coles Melvyn (2003) "Equilibrium Wage-Tenure Contracts" *Econometrica* Vol 71:(5) pp 1377-1404

DiNardo John, Nicole Fortin and Thomas Lemieux (1996) "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach" *Econometrica* Vol 64:(5) pp 1001-44

Lazear Edward P (1986) "Raids and Offer Matching" *Research in Labour Economics*, Vol 8A, pp 141-165

Lazear Edward P (1979) "Why is there Mandatory Retirement" *Journal of Political Economy* Vol 87 pp1261-64.

Medoff James L and Abraham Katharine (1981) "Are Those Paid More Really more Productive? The Case of Experience" *The Journal of Human Resources* Vol 16:(2) pp 186-216.

Stevens Margaret (2004) "Wage-Tenure Contracts in a Frictional Labour Market: Firms' Strategies for Recruitment and Retention" *Review of Economic Studies* Vol 71 pp 535-551.

Stevens Margaret (2003) "Earnings Functions, Specific Human Capital, Job Matching: Tenure Bias is Negative" *Journal of Labor Economics* Vol 21:(4) pp 783-805
Treble, John, Van Gameren, Edwin, Bridges, Sarah, Barmby, Tim (2001). "The internal economics of the firm: further evidence from personnel data". *Labour Economics* Vol 8(5): 531-552.

Appendix 1

Probit estimations for weighting function			
	t=89	t=94	t=99
	Coefficient	Coefficient	Coefficient
	(Std.Error)	(Std.Error)	(Std.Error)
Constant	-0.924	-1.506	-1.798
	(0.026)***	(0.033)***	(0.040)***
Grade on Entry†	0.036	0.003	0.041
	(0.032)	(0.039)	(0.046)
Gender	0.150	-0.003	-0.318
	(0.056)***	(0.073)	(0.107)***
A-levels	-0.466	0.124	0.247
	(0.047)***	(0.049)**	(0.055)***
O-levels	-0.450	0.143	0.189
	(0.037)***	(0.040)***	(0.046)***
Asian/Asian British	-5.210	0.306	0.384
	(0.343)***	(0.136)**	(0.146)***
Black/Black British	0.029	-0.017	-0.083
	(0.140)	(0.170)	(0.206)
Gender*A-levels	0.012	-0.012	0.016
	(0.063)	(0.066)	(0.075)
Gender*O-levels	-0.009	-0.006	0.043
	(0.045)	(0.049)	(0.058)
Gender*Asian/Asian British	3.479	0.076	-0.024
	(0.000)	(0.161)	(0.178)
Gender*Black/Black British	-0.262	0.018	0.226
	(0.163)	(0.192)	(0.229)
Gender*Grade on Entry	-0.113	0.029	0.237
	(0.060)*	(0.077)	(0.111)**
χ^2	762.37	67.46	124.07
Log Likelihood	-12380.52	-8390.89	-5598.56
N	30778	30778	30778

† If staff on entry, grade on entry=1, 0 otherwise
* significant at 10%; ** significant at 5%; *** significant at 1%