INSTITUTE OF ECONOMIC STUDIES

WORKING PAPER SERIES

If you're so smart, why aren't you rich?
Wage inequality with heterogeneous workers

Alison L. Booth and Gylfi Zoega

Address: Alison L. Booth
                ISER, University of Essex
                Wivenhoe Park, Colchester
                England CO4 3SQ

Address: Gylfi Zoega
                Department of Economics, Birkbeck College
                7 – 15 Gresse Street,
                London W1P 2LL

If you're so smart, why aren't you rich?
Wage inequality with heterogeneous workers

Alison L. Booth

ISER, University of Essex
Wivenhoe Park, Colchester
England CO4 3SQ

Gyl... Zoega

Department of Economics
Birkbeck College
7-15 Gresse Street
London W1P 2LL

October 16, 2002

Abstract

This paper provides microfoundations for wage compression by mod-
elling wage-setting in a world of heterogeneous workers and ...rms. Workers
are di¤erentiated by observable innate ability. A high-ability worker con-
fers on a ...rm an externality, since her ability raises the average level of
talent within that ...rm and increases the range of tasks that can be per-
formed. This gives some ...rms monopsony power in the market for labour
trained to do the more advanced tasks. Firms will assign their better
workers to the more advanced tasks performed within their ranks, and
wages are compressed within ...rms, so that low-ability workers are paid
more, relative to their talent, than high-ability workers. The model also
o¤ers an explanation for why wage inequality has recently increased in
some countries: exogenous changes that increase labour market competi-
tion can disporportionately bene...t higher ability workers and widen the
wages distribution.

Keywords: Heterogeneous workers, hierarchical assignment models, monop-
sony, wage compression.

JEL Classi...cation: J24, J31, J42.

1 Introduction

Firms undertake a number of tasks of varying levels of complexity. Some forms of production involve the ordering of tasks in sequential production – from the canvas to a ...nished painting, or from raw steel to an automobile – as modelled by Kremer (1992) and Sobel (1992). Other forms of production involve the simultaneous undertaking of a variety of tasks, some examples being services provided by airlines or by consultancies, police work, and symphony orchestras. An example closer to home is provided by the activities of university depart-ments – in which the output from teaching does not need to feed sequentially into research activities – and where the activities might be done in any order over a particular unit of time.

Our analysis is premised on the plausible assumption that there is an exoge-nously given distribution of ability in the labour market, that some ...rms hire better workers than others, and that the average level of ability within ...rms determines the level of sophistication of tasks performed by each ...rm. An air-line that wants to ‡y the most advanced planes; a defence force that wants to engage in anti-ballistic missile defence; or a university department that wants to produce the most cutting-edge research, will all need workers of high ability. A high-ability worker confers on a ...rm an externality, since her ability raises the average level of talent within that ...rm and increases the range of tasks that can be performed. The high-ability worker acting individually cannot exploit this because it is not her own human capital but the average within the ...rm that is important for determining task complexity. "Good" ...rms employ workers of high average ability and are for that reason able to perform more complex tasks than the "bad" ones. Pilots who want to ‡y the Concorde jet, or workers who want to reach for the sky in a more ...gurative sense, may thus face only a hand-ful of potential employers; and this gives these employers monopsony power.[1] Our model shows how wage compression arises quite naturally in labour mar-kets due to di¤erences in the level of monopsony power across tasks. Such wage compression has been shown to be quite important for ...rm-sponsored training (see inter alia Acemoglu and Pischke, 1999; and Booth and Zoega, 2001).

The remainder of this paper is set out as follows. In Section 2, we set out the model motivation and assumptions about the initial endowment of agents in our economy. In Section 3, we investigate the allocation of heterogeneous workers across tasks. In Section 4 we derive equilibrium wages, and show how these di¤er across tasks and employers. Section 5 outlines some applications of our model and also o¤ers an explanation as to why within-group wage inequality has increased in many countries. Section 6 concludes.

---

[1] In a companion paper (Booth and Zoega, 2001), we derive the implications of this set-up for ...rms' training decisions. In particular, we show how ...rm heterogeneity with regard to the average level of ability among employees generates an incentive for ...rms to fund training that is inherently general, but in practice only useful at a ...nite number of ...rms.

2 The Set-up of the Model

Our model tells a very simple story, and one that can be thought of as a metaphor for wage-setting and worker assignment for knowledge workers. This is because the story is applicable to graduates in physics, economics, or to any other areas of knowledge where ability - both innate and acquired through education - matters. The model is not however intended to describe how wages are set for super-stars.

In our model, ...rms form when workers - di¤erentiated by their innate but observable ability - show up randomly at the workplace. The ...rm observes the average level of innate ability of its workforce, and this determines the complexity of tasks that can be undertaken. The ...rm then allocates workers of varying levels of ability across tasks, with the better workers doing the more complex tasks. There will be fewer ...rms able to do the more complex tasks than able to do the easy tasks, because of the limits to the numbers of higher ability workers. Firms then set wages for each task, taking into account workers' outside opportunities. Clearly if there are few ...rms performing the more complex tasks, ...rms doing these will have some monopsony power.

We show that workers at the top end of the ability distribution doing the more complex tasks will earn more than workers at the lower end doing less complex tasks, because they get some of the returns to their ability. However these talented workers will also lose proportionately more of the gains from the employment relationship than will their less gifted colleagues performing simpler tasks. In this sense the wages distribution for heterogenous workers is compressed. However we argue that developments in some countries over the last few decades - which we document later in the paper - may have increased the degree of within-group labour market competition at the top end of the ability range, eroding some of these monosponistic pro...ts, and widening the wages distribution. Our model is thus consistent with the stylised facts summarised in, for example, Gottschalk and Smeeding (1997), which document increases in within-group wage inequality, with the top-end increasing more than the bottom end is declining. However, in the longer run we would expect to observe further wage compression.

How do we obtain these results? We suppose that workers are di¤erentiated by their innate but observable ability measured in terms of their endowment of e¢ciency units of labour h_i. We assume (although this is not important for our analysis) that the allocation of innate ability across workers is normally distributed as

$$h_i \sim N\left(\overline{h}, \frac{3}{4}^2\right) \tag{1}$$

where $\overline{h}$ is the mean ability level in the (...nite) population of available workers in the labour market. The quality of a ...rm is determined by the innate ability of the workforce that applies for its vacancies. Some employers will acquire a better pool of workers than others. In particular, ...rms will di¤er in the average level of innate ability of their workforces. In this way worker heterogeneity

translates into ...rm heterogeneity.[2] And ...rm heterogeneity in our model – in terms of the average innate ability or human capital of its workers – a¤ects the range of tasks that can be performed within the ...rm. The higher the average level of innate ability of a ...rms's workforce, the greater is the range of tasks that can be performed at that ...rm.[3] Higher ability workers are allocated to the more complex tasks.

Examples of such production organisation are not hard to ...nd. For instance, consider the concentration of musicians playing in the Royal Philharmonic Orchestra, the group of world-class physicists who developed the atomic bomb at Los Alamos, and the continuous ‡ow of new products by Microsoft and other concentrations of skilled programmers. University departments, which help create and foster synergy in research, represent yet another example. Departments with higher concentrations of more talented individuals are likely to produce more groundbreaking output.

This production process has an interesting implication, given that workers are potentially mobile across ...rms doing the same task. By de...nition in our model there is only one ...rm capable of performing the most advanced task, two ...rms capable of performing the next-advanced task, and so on, while all ...rms can do the simplest tasks. It is here that an employer's potential monopsony power lies. As the number of ...rms doing a given task increases, employers' monopsony position is diminished. Consider a simple example: as we move towards better and better university departments – which do increasingly sophisticated research and teach more demanding and receptive students – the level of monopsony is increased. An outstanding university may then be able to perform most – if not all – of the tasks while the poor inner-city college can only provide teaching in the basics. We shall explain in a later section the implications of this for wage setting.

Note that the more able workers do not have any o¤setting monopoly power. The reason for this is that, on their own, any one of them cannot enable a ...rm to do a new task. Only a collection of able workers can create the conditions necessary for a new complicated task to be performed. A high-ability worker confers on a ...rm an externality since her presence raises the average level of talent within that ...rm, and in so doing increases the range of tasks that can be performed. However the high-ability worker cannot individually exploit this, because it is not her own human capital that determines task complexity but the average within the ...rm. Think of a bright violinist o¤ering her talents to the Royal Philharmonic or a computer whiz approaching Microsoft for a job. Only if all high-ability employees at a ...rm co-ordinated their activities – for

---

[2] We consider in our model both horizontal and vertical job di¤erentiation. The vertical di¤erentiation arises because workers are heterogeneous with respect to innate ability generating a hierarchy of tasks, as per the hierarchical job assignment literature surveyed in Sattinger (1993). The horizontal di¤erentiation arises because, like Bhaskar and To (1999), we have jobs di¤ering in terms of their non-wage characteristics, over which workers have varying preferences. This generates an imperfectly elastic labour supply to the ...rm.

[3] We assume that a single worker alone cannot perform a new task at a ...rm. We do this in order to capture the fact that collaborations in research can generate results that are not within the individual reach of those involved.

example through a trade union – could this externality be internalised.[4] Thus, in the absence of unions, once the collection of workers is in place, the employer has monopsony power in its relations to any new employee. The employer can give a worker the opportunity to perform a given task – for example only in a good symphony orchestra can a violinist realise the dream of playing the most-challenging of symphonies – but a single worker cannot enable a ...rm to have a new task performed within its ranks.

3   Further assumptions

History matters in our model. When a ...rm starts up, the complexity of its production process is determined solely by the average ability of the workforce that shows up at its doors. The ...rm may however choose not to hire all the workers who show up at its gates, as we demonstrate in Section 4.

Firms in the industry are ranked in terms of the average ability of their workforce. Given that there is a distribution of ability that translates into a distribution of task complexity across our R ...rms, it is notationally convenient to rank ...rms along the real number line in terms of their task complexity $k \in [1; R]$. Thus we denote by R both the number of ...rms and the total number of tasks. Firms are ranked from the best (within-...rm-average) ability workforce ($k = R$) to the worst ($k = 1$). It follows that ...rms di¤er in their within-...rm average level of ability that we denote by H:

$$H_R > ... H_r > H_{r_i 1} ... H_1; \qquad H_i \gg N^{\left[\bar{h}; R_4^2 = N\right]} \qquad (2)$$

The distribution of average ability is normal with mean $\bar{h}$ so that an average ...rm has average ability level $\bar{h}$ while other ...rms have either a higher or a lower average ability level among employees.

We assume that the higher is the average level of innate human capital in the ...rm, the more advanced are the tasks can be performed within its ranks. However, we do not model the within-...rm interactions between workers. Instead we simply assume that the average level of innate ability di¤ers across ...rms depending on workers' endowments, and that more complex tasks can only be performed where the average level of innate ability is greater. What we have in mind is the importance of teamwork, work ethics, and the synergy that often exists within a group of talented individuals. We think of workers as sharing ideas, complementing each other and setting high standards that others have to satisfy.

The ...rm ranked R has a workforce characterised by the highest level of average innate ability and it performs the most complex task in addition to all simpler ones. Firm $R_i 1$ performs all the same tasks apart from the most

---

[4]Thus our approach is not intended to model the behaviour of individual super-stars, such as Pavarotti or Vengerov. Instead we wish to model team players who make up university departments, or who enable Microsoft to develop the most sophisticated software, or who facilitate the mapping of the human genome.

advanced one, which is the one that distinguishes the best and the next-best …rm. Thus, letting M_k denote the number of …rms performing task k, it is clear that M_k is decreasing in the complexity of the task performed. More precisely, we can write

$$M_k = R \mathbin{¡} k + 1 \tag{3}$$

where R is the total number of …rms in the industry and k is the number of tasks performed at a particular …rm, $k \in [1; R]$. Equation (3) implies that there is one …rm doing the most advanced task $k = R$, two …rms doing the second most advanced, and R …rms doing the simplest task.

We assume that the value of output produced by performing a given task is increasing in task complexity. We measure the value of output coming from each of the tasks in terms of units of output emanating from the simplest task. The variable y_k measures the value of output per eΦciency unit of labour devoted to task k measured in units of output coming from the …rst task, which implies $y_1 = 1$ and $y_k \mathbin{\_} y_1$ for all $k > 1$. We assume that the value of production – coming from each task – is a linear function of labour measured in eΦciency units so that the output of worker i doing task k is equal to $y_k h_i$ irrespective of how many other workers are working on the task. One unit of human capital would only produce $y_1 = 1$ in the "worst" …rm (where the average ability level among a worker's colleagues is H_1) – performing the most mundane tasks and receiving wage w_1– but it would produce output valued at y_R in the best …rm doing the most sophisticated task at wages w_R where average ability is H_R. Think of a violinist playing a part in a great symphony – producing $y_R h_v$ where the subscript v refers to the violinist – or, alternatively, going it alone and playing on street-corners without the help of other players and workers doing tasks such as maintaining and operating the concert hall. It is clear that on the street corner the violinist would produce only $y_1 h_v$, since there he would not have the same opportunities for challenging assignments needing the support of the whole symphony orchestra and supporting sta¤.

Each …rm undertaking task k uses a technology that is constant-returns-to-scale with respect to the human capital it employes, $y_k \mathbin{P} h_i$. Total pro…ts of a …rm performing k tasks – de…ned in terms of units of output coming from the simplest task – are the sum of value-added less task-induction training costs over all tasks. Value-added $y \mathbin{¡} w$ is simply the di¤erence between the value of output and wages per eΦciency unit of labour.

Pro…ts of the j-th …rm undertaking k tasks can be written as

$$P_{jr} = \sum_{r=1}^{k(H_j)} (¼_{jr} \mathbin{¡} c_{jr} n_{jr}) \tag{4}$$

where $¼_r$ represents the …rm's pro…ts from workers allocated to task r, n_r is the number of workers allocated to task r, and c_r is the cost of training a worker to

do task r. We shall describe in Section 4 how n_r is determined. Note that the number of tasks performed is a function of the average ability level within the firm; $k(H_j)$; $k^0 > 0$; $k^{00} = 0$:

4 Wages, Monopsony and Task Assignment

Firms' task heterogeneity is determined by the average innate ability of all workers who initially show up. In this section we consider how - once workers are allocated across tasks within a firm - firms set wages for workers performing each task. Obviously firms setting wages will take into account worker's outside opportunities, and these differ depending on what tasks workers are performing.

How are outside opportunities incorporated into our model? Once a firm's task complexity has been established and firms have allocated workers to tasks, workers observe that there are M_k firms in total performing the task k to which they have been allocated, $k \in [1; R]$ and $0 \cdot M_k \cdot R$. Consequently there are M_k firms at which they can perform the task for which their ability best suits them. We assume that they do not move between levels in the hierarchy.

We shall proceed by backward induction, and first consider the optimal wage rate for a given allocation of workers across firms for each task. Then we shall show that – given expected wage rates – the better workers will be allocated to the more complex tasks.

We assume, following Bhaskar and To (1999) and Booth and Zoega (1999), that firms differ in the nonpecuniary attributes of the jobs they offer. We also suppose, following Salop (1970), that there are R firms equally spaced around a circle of circumference C such that the distance between two adjacent firms in job characteristics space is C=R:

We assume that workers are characterised by heterogeneous preferences x for the job characteristics x_j offered by each firm. The more distant are the j-th firm's job characteristics x_j ($j = 1; ....; M_k$) from the worker's preferred characteristics x, the larger is the worker's disutility cost (denoted by $jx_i \ x_j j$) associated with employment at firm j. We assume $cov(h_i; x_j) = 0$, since there is no reason to suppose that individual ability and preferences for job characteristics should be correlated.[5]

Suppose firm j is the representative firm of the M_k firms performing task k. We follow Salop (1979), Bhaskar and To (1999) and Hamilton, Thisse and Zenou (2000) in assuming that the variable x is distributed continuously and uniformly on a circle of length C. Suppose that firms' job characteristics are equally spaced around the circumference, such that $C=M_k$ is the distance between two adjacent firms in job-characteristics space.[6] The density is given by a constant D and consequently a thick market is associated with a high value of

---

[5] Notice that all workers of ability level h_i doing the kth task are identical from the firm's perspective, since their productivity is observable as $h_i y_k$.

[6] Bhaskar and To (1999) and Booth and Zoega (1999) also assume that workers have heterogenous preferences for job characteristics. Bhaskar and To (page 200) cite various empirical studies supporting this assumption.

D while a thin market is associated with a low value. Hence D is capturing the number of workers of each preference type while C measures the heterogeneity of workers doing task k. We assume that C and D are the same across all tasks.

The jth firm performing the k-th task is going to have an effective pool of labour whose outer boundary in job-characteristic space is the midpoint between this firm's job characteristics x_j and the two adjacent firms x_{j+1} and x_{j-1}. Some workers in these two sub-segments (those closest to the jth firm and further from the midpoints) are going to be more satisfied than workers whose preferences are closer to the midpoints between pairs of firms. Wages are set to satisfy these marginal workers at the midpoints.

Firms simultaneously choose their wage levels. The net wage a worker receives per efficiency unit is thus $w_j - j x - x_j j$, where we omit the subscript k for clarity. To find the labour pool for firm j, we first establish the upper and lower bounds for that firm in job-characteristic space, represented by $\bar{x}$ and $\underline{x}$ respectively. Workers will choose to work at the firm giving them the highest wage per efficiency unit net of their disutility associated with working at that firm. First we consider the lower bound for the jth firm. Given the wages w_{j-1} and w_j set by the two adjacent firms, the marginal worker at x will be indifferent between working at firm j and firm $j-1$. Thus we find $\underline{x}$ by solving the following;

$$w_j - (x_j - \underline{x}) = w_{j-1} - (\underline{x} - x_{j-1}) \tag{5}$$

which yields

$$\underline{x} = \frac{w_{j-1} - w_j + (x_j + x_{j-1})}{2} \tag{6}$$

Similarly, the upper bound for the jth firm $\bar{x}$ solves:

$$w_{j+1} - (x_{j+1} - \bar{x}) = w_j - (\bar{x} - x_j) \tag{7}$$

which yields

$$\bar{x} = \frac{w_j - w_{j+1} + (x_{j+1} + x_j)}{2} \tag{8}$$

Firm j doing the kth task sets wages for that task by maximising profits π_{jk}, written as:

$$\max_{w_{jk}} \pi_{jk} = \int_{\underline{x}_k}^{\bar{x}_k} D_k (y_k - w_{jk}) H_{jk} dx_k = D_k (y_k - w_{jk}) H_{jk} (\bar{x}_k - \underline{x}_k) \tag{9}$$

8

where wages per e¢ciency unit of labour for the jth ...rm are denoted by w_k so the wages of a worker with human capital H_{jk} working on task k at the jth ...rm are $H_{jk}w_{jk}$.[7] This brings us to our ...rst proposition:

Proposition 1 The smaller is the number of ...rms M_k performing a given complexity of task, the larger is the gap between wages per e¢ciency unit and productivity. Hence the gap is increasing in task complexity. This implies that a ...rm's level of monopsony power is rising in task complexity.

Proof:
The ...rst-order condition of (9) yields the following equation:[8]

$$D_k H_{jk} (y_k \mathbf{i} \ w_{jk}) = D_k H_{jk} (\overline{x}_k \mathbf{i} \ \underline{x}_k) \qquad (10)$$

The term on the left-side of equation (10) shows the marginal bene...t of raising wages due to a larger workforce. The term on the right-hand side has the marginal cost of raising wages, which consists of higher wage payments to all workers. Using equations (6) and (8), and after some manipulations, we get the following equation:

$$(y_k \mathbf{i} \ w_{jk}) = \frac{[(x_{j+1} \mathbf{i} \ x_{j \mathbf{i} 1}) \mathbf{i} \ (w_{j+1} + w_{j \mathbf{i} 1}) + 2w_j]}{2} \qquad (11)$$

Now we know that ...rms are located symmetrically around the circle of job-characteristic space, such that the distance between ...rms is given by $C=M_k$. Consequently $x_{j+1} \mathbf{i} \ x_{j \mathbf{i} 1} = 2C=M_k$. In symmetric equilibrium with $w_j = w_{j+1} = w_{j \mathbf{i} 1}$ we can now simplify the equation above to obtain:

$$w_{jk}^{\tt a} = y_k \mathbf{i} \ \frac{C}{M_k} \qquad (12)$$

Notice in (12) that as $M_k \ ! \ 1$, $w_{jk}^{\tt a} \ ! \ y_k$. This is the wage scenario for a perfectly competitive labour market. However, as $M_k \ ! \ 1$, $w_{jk}^{\tt a} \ ! \ y_k \mathbf{i} \ C$.[9] ¥
The intuition behind this result is as follows: When there is a lot of heterogeneity for a given number of ...rms – C is large – each ...rm is faced with many workers who prefer its job characteristics to those of other ...rms. This makes the cost of raising wages high because by o¤ering higher wages – aimed

---

[7]Note that we have made the assumption that ...rms do not observe the ability levels of marginal workers and as a result use the average ability level H_{jk} in the pro...t function. This simpli...es the exposition without any loss of generality.

[8]The second order condition holds, $\frac{d}{dw_{jk}} \ \frac{d\frac{1}{4}_{jk}}{dw_{jk}} \ < 0$:

[9]In order to ensure $w_j^{\tt a} > 0$, we need to impose the restriction that $y > C$.

at attracting new recruits whose preferences do not match the jobs being o¤ered as well – ...rms will have to pay higher wages to many workers who would have stayed with them anyway. However, if there are very many ...rms – M_k is large – then not many workers would settle for any given ...rm in the absence of wage incentives. This reduces the marginal cost of o¤ering higher wages. It follows that as the number of ...rms M_k goes up, the marginal cost of raising wages falls, which makes the wage level rise towards marginal productivity as the number of ...rms approaches in...nity.

Notice that labour supply n_k to ...rm j is given by

$$n_k = (\overline{x}_k \,\textrm{¡}\, \underline{x}_k)\, D_k = D_k \left[w_j \,\textrm{¡}\, \frac{1}{2}\,(w_{j+1} + w_{j\,\textrm{¡}\,1}) + \frac{C}{M_k} \right] \tag{13}$$

and is increasing in the ...rm's own wage, in contrast to the perfectly competitive case where the labour supply is in...nitely elastic. Moreover, the labour supply is a positive function of the density D and the heterogeneity of workers' preferences C and a negative function of the number of ...rms competing for the workers M_k. The more workers there are of each preference type D, and the more heterogeneous are workers, the greater the number of workers who want to join a particular ...rm. However, taking the density and the heterogeneity as given, a greater number of ...rms implies that fewer workers will apply to any single ...rm.

Having established the optimal wage rate and labour supply we now turn to the optimal allocation of workers of di¤erent abilities across tasks, for given optimally chosen wages for each task. The jth ...rm will only hire a worker as long as it pro...ts from doing so and there may be workers of such low abilities that it is not in the interest of any ...rm to hire them. Equation (4) implies that the following condition will hold for every worker trained to do task k:

$$(y_k \,\textrm{¡}\, w_{jk})\, h_i \,\textrm{,}\, c_k \tag{14}$$

The term on the left-hand side is the expected pro...ts from employing worker i. Only if this condition holds does the ...rm allocate and train a worker for task k. Since the number of workers applying for jobs at the representative ...rm is given by supply – and the ...rm is willing to hire any worker as long as inequality (14) holds for at least one of the tasks – we are primarily interested in the allocation of workers across tasks. We will show that the best workers should be allocated to the most advanced tasks within the ...rm.

Assume that there is a worker with ability level h_j doing task r – the most sophisticated within ...rm r – and another doing the least advanced task – task 1 – with a higher ability level h_i; $h_i > h_j$. If the ...rm could, ex-post, relocate them at zero cost – that is recant its earlier decision and assign workers di¤erently to the tasks – so that the latter would be trained to do task r and the former to do task 1, the expected bene...t to the ...rm would be the following since total training costs would be unchanged:

$$[(y_r - w_r^\pi) - (y_1 - w_1^\pi)](h_i - h_j) > 0 \qquad (15)$$

which is positive if and only if

$$(y_r - w_r^\pi) - (y_1 - w_1^\pi) > 0 \qquad (16)$$

that is if the …rm bene…ts more from the employment of a worker doing the sophisticated task. We will show this to be the case because of a greater degree of monopsony power for task r. It will then become clear that the best workers would be allocated to the most advanced task and the worst to the least advanced one.

Let us now take a look at the …rm rank R – which performs all tasks of complexity R and below. The most exceptional workers will then be devoted to task R. It follows from equation (4) that the threshold ability level is;

$$h_j \geq \frac{c_R}{y_R - w_R^\pi} \qquad (17)$$

Because of the high degree of monopsony for task R, the …rm bene…ts from allocating all of its workers to the task but this comes at a cost; the cost of training a worker to do task R is higher than the cost of training him to do any other task. Workers who do not meet the grade for task R may then be allocated to task $R - 1$ as long as the following condition holds:

$$\frac{c_{R-1}}{y_{R-1} - w_{R-1}^\pi} \leq h_i < \frac{c_R}{y_R - w_R^\pi} \qquad (18)$$

The inequality to the right says that worker i is not good enough for the most advanced task while the inequality on the left says that he is good enough for task $R - 1$. Note that while the monopsony pro…t form task R is greater than that for task $R - 1$, the cost of training the worker is lower. Importantly, a worker who is judged not to be su¢ciently able for task R will only be trained to do task $R - 1$ if lower training costs o¤set the e¤ect of lower monopsony power. If not, he and all others who are not su¢ciently able to do task R will be turned away and the …rm will only do the most advanced task R. Finally, there will be some workers who are not even su¢ciently good for task 1:

$$h_i < \frac{c_1}{y_1 - w_1^\pi} \qquad (19)$$

These workers will not be trained for any task within this …rm – in spite of the low costs of training a worker to do the most basic task – and will be turned away. This yields our second proposition:

Proposition 2 The highest ability workers should be allocated to the most advanced task and the worst to the least advanced one.

Again going back to our earlier examples, a good airline with employees of such high average ability that it can operate the most advanced aircraft, should allocate its best pilots to do the most sophisticated task, which is ‡ying the most recent types of aircraft. Similarly, at a top-ranking university, the most talented members of sta¤ should be directed to the most challenging research tasks. While this comes as no surprise, the reasons given by our model – while intuitive – are new in the literature: By training the best pilots to ‡y the Concorde, an airline gains monopsony power in its relations with high-ability individuals. This enables it to pro...t more than it would otherwise do since these pilots will perform better than other pilots and the airline gets a share of their output every inch – or every air-mile! – of the way.

In the sequential production-task models of Kremer (1992) and Sober (1992), mistakes made doing the early tasks can be easily corrected – without destroying the output of any subsequent worker. However, mistakes made performing the ...nishing touches can ruin the work performed doing all earlier tasks. Hence they show that a ...rm should put the more skilled and reliable workers in charge of the ...nal tasks. While our model is also directed at task complexity, it is in a di¤erent context, since our production process is not sequential. Almost any employee of an airline can cause a Concorde to crash – be it due to incorrect fuelling, poor loading and maintenance, incorrect ‡ight-path, failure to detect an explosive device, or pilot error – and any member of a symphony orchestra can ruin a good concert. Nonetheless, we still ...nd that the more able workers should be delegated to the more advanced tasks, because the ...rm then reaps maximum pro...ts from the greater monopsony position associated with the more di¢cult tasks, as we show below. It is more pro...table to be in a monopsony position vis-a-vis the more productive workers.

It follows that the more advanced is the relevant task, the higher are pro...ts from performing the task within the ...rm: all ...rms can perform the simplest task, hence the level of competition is highest for this task. But as ...rms become better, they gain elements of monopsony power in the market for labour trained to do tasks that are only feasible within their ranks – labour that can only produce output by enjoying the company of the high-quality workers currently employed.

According to Proposition 2 the better workers should always be assigned to the more advanced task. Since these workers have higher ability – they embody more e¢ciency units of innate human capital – they have higher take-home wages $w_k h_i$ on that count. But Proposition 1 tells us that monopsony pro...ts increase as we move to more advanced tasks. Combining the two results gives a proposition involving wage compression, which we now de...ne.

De...nition 1 Absolute wage compression occurs when the di¤erence between the value of output produced by two workers of di¤erent ability ($h_i > h_j$) doing

12

separate tasks (k and $k-1$) is greater than the difference between their wages, that is $h_i y_k - h_j y_{k-1} > h_i w_k - h_j w_{k-1}$.

It is easy to combine Propositions 1 and 2 to show that the wage distribution is compressed in our model. This yields Proposition 3.

Proposition 3 Wage compression arises quite naturally in a model with heterogeneous workers: As we move from a less advanced to a more advanced task, the increase in output per worker is greater than the increase in take-home wages, i.e. there is wage compression. For a worker h_i doing task k and a lower ability worker h_j doing task $k-1$, it follows that:

$$h_i y_k - h_j y_{k-1} > h_i w_k - h_j w_{k-1} \tag{20}$$

Proof:
This proposition is easily veri…ed using our earlier equations. We know that

$$h_i (y_k - w_k) > h_j (y_{k-1} - w_{k-1}) \tag{21}$$

since according to Proposition 1

$$y_k - w_k > y_{k-1} - w_{k-1} \tag{22}$$

given equations (9)-(12) and $h_i > h_j$. This can then be rewritten as

$$h_i y_k - h_j y_{k-1} > h_i w_k - h_j w_{k-1} \tag{23}$$

which is our de…nition of wage compression.¥

The results in Proposition 3 shows that wage compression arises quite naturally in a world of heterogeneous workers in the absence of any speci…c institutional arrangements. This is in contrast to the previous literature on wage compression, which typically uses the representative-agent model and posits institutional explanations for wage compression.[10]

Acemoglu and Pischke (1999) show that wage compression is a necessary condition for …rm-sponsored training. In our model, we can show that – if a high-ability worker doing an advanced task were to quit or die – the …rm may bene…t from retraining a lower-ability worker doing a less advanced task to take her place. This is due to the wage compression.

---

[10] Recent contributions to the wage compression literature o¤er explanations such as minimum wages and other wage ‡oors, unions, asymmetric information about a worker's skills between current and prospective employers, and the complementarity between …rm-speci…c and general skills (alternatively between physical capital and general skills). In contrast, our non-representative-agent model makes these mechanisms unnecessary, although not irrelevant.

5 Applications

What might reduce wage compression within our model? Giving higher-ability workers access to more potential employers would widen wage dispersion. This might be achieved if controls preventing international labour mobility were relaxed, or if working for employers in other countries became easier. For example, improvements in transportation, communications or the invention of the internet have given workers of higher ability a wider range of ...rms from which to choose employment. While this would improve worker welfare at the top end of the distribution, it would worsen the (absolute) position of employers and the (relative) position of lower-ability workers.

The employment of agents by higher ability workers could also improve the position of such talented workers through two mechanisms. First, the agent might be able to bargain away some of the ...rm's monopsony pro...ts through the threat of labour withdrawal. This would be possible if workers doing the most advanced tasks were to combine into a labour union that would act as an agent on their behalf. Second, an agent representing a single talented individual may be able to introduce the individual to more potential employers. Thus – to take another example – top ...ction publishing houses rely for their sales on signing up talented authors. The role of the literary agent is not only to verify worker ability but also to introduce the author to a wider range of ...rms than would otherwise be possible for the individual. In so doing the agent reduces the monopsony power of the publishing house and ensures that the author receives a higher share of royalties than might be achieved individually. While the author on her own might threaten to take her oeuvre elsewhere, this threat is less credible when not backed by an agent with the requisite portfolio of ...rms from which to choose. On her own she faces only the local pool of publishers. With the agent, she faces a wider – perhaps even global – pool.

Our theory has the potential to explain changes in the wage distribution over time. Recent evidence shows a widening of wage inequality across some developed countries, especially the US and the UK (see for example Gottschalk and Smeeding, 1997). Popular explanations for this stylised fact include globalisation, technological change, and increased demand for skilled workers. In this paper we have provided an alternative explanation in a model of heterogeneous agents. We show that a ...rm's degree of monopsony power over knowledge workers – that is the ones of higher ability doing the more advanced tasks – is rising in task complexity. In our model, a widening of the wage distribution would occur whenever more talented individuals are given access to more potential employers.[11]

Consider our model as initially applying to the intra-country distribution of wages for knowledge workers prior to the changes listed below. Now suppose that there are changes that facilitate the globalisation of the labour market for knowledge workers. Such changes might include: (i) The removal or reduction

---

[11] Saint-Paul (2001) also looks at explanations for widening wage inequality in a model with a ...xed supply of heterogenous workers. We di¤er from his approach in a number of respects, including in our emphasis on a hierarchical model of job assignment.

of immigration controls for knowledge workers, as has happened for example in the UK and the US. (ii) The emergence of international and national agency ...rms (the growth of the temporary agency ...rm is one example) that specialise in matching workers to jobs at the international level.[12] (iii) The revolution in electronic communication allowing knowledge workers to view global job opportunities costlessly. Coupled with point (i) above (removal of migration controls for knowledge workers) this means that the number of ...rms accessible has increased. (iv) The growth of the English language as a international medium of communication, which e¤ectively increases the range of ...rms where knowledge workers can gain employment.

The net result will be a widening of the within-country wage distribution due to increased labour-market competition at the top of the market.[13] In contrast, the wages of knowledge workers should become more equalised across countries. In sum, global wage inequality should be declining within groups of knowledge workers, although country-speci...c wage inequality has grown.

6   Conclusions

We have shown how wage compression arises quite naturally in market economies with heterogeneous ...rms and workers. In the training literature, such wage compression has been shown to be an important condition for ...rm-sponsored training: employers are willing to pay for the training of their workforce if the training adds more to productivity than it adds to wages. Importantly, we do not need to resort to labour unions or minimum wages to generate the wage compression. This result follows from several intermediate results of our analysis:

- ² The number of tasks that can be performed within a ...rm depends on the average level of ability within the ...rm. It follows that there are good ...rms where the most advanced tasks can be performed as well as bad ...rms where only the more simple ones can be done.

- ² The degree of monopsony power is rising in task complexity.

- ² Firms allocate the best workers to the most sophisticated tasks. While more talented workers get paid more on the basis of their higher ability, they do not receive the full return to their talent owing to the monopsony power enjoyed by their employers.

---

[12] Our model shows that there are pro...ts to the agency ...rms from so doing – from monopsony rents - and the agency ...rm can cream o¤ these and share them with the agency worker.

[13] Of course, while these changes outlined above may explain why/suggest that the monopsony pro...ts of ...rms are diminishing within each group of knowledge workers, the logical outcome is that - for the global labour market - there will still be monopsony at the top end once we are a global village.

While each of these results is plausible – as well as expected – combining the three to generate microeconomic foundations for wage compression may turn out to be a useful result in labour economics.

References

[1] Acemoglu, Daron and Jörn-Ste¤en Pischke (1999), 'The Structure of Wages and Investment in General Training,' Journal of Political Economy, 107(3), June, p. 539-572.

[2] Bhaskar, V. and Ted To (1999), 'Minimum Wages for Ronald McDonald Monopsonies: A Theory of Monopsonistic Competition,' The Economic Journal, Vol. 109, No 455, p. 190-203.

[3] Booth, Alison L and Gyl... Zoega (1999), 'Do Quits Cause Under-training?,' Oxford Economic Papers, 51, p. 374-386.

[4] Booth, Alison L and Gyl... Zoega (2001), 'Worker Heterogeneity and General Training,' working paper, Birkbeck College.

[5] Gottschalk, Peter and Timothy M. Smeeding (1997), 'Cross-National Comparisons of Earnings and Income Inequality,' Journal of Economic Literature, Vol. 35, No. 2., p. 633-687.

[6] Hamilton, Jonathan, Jacques-Francois Thisse and Yves Zenou (2000), 'Wage Competition with Heterogeneous Workers and Firms,' Journal of Labour Economics, 18 (3), p. 453-472.

[7] Kremer, Michael (1992), 'The O-Ring Theory of Economic Development,' Quarterly Journal of Economics, volume 108, p. 551-575.

[8] Saint-Paul, Gilles (2001), 'On the Distribution of Income and Worker Assignment under Intra...rm Spillovers, with an Application to Ideas and Networks,' Journal of Political Economy, 109(1), p. 1-37.

[9] Salop, Steven (1979), 'Monopolistic Competition with Outside Goods,' Bell Journal of Economics, vol. 10, p. 141-156.

[10] Sattinger, M. (1993), 'Assignment Models of the Distribution of Earnings,' Journal of Economic Literature, 31, p. 831-880.

[11] Sobel, Joel (1992), 'How to Count to One Thousand,' The Economic Journal, CI, p. 1-8.

INSTITUTE OF ECONOMIC STUDIES WORKING PAPERS 1987-2002
Formerly Iceland Economic Papers Series

Editor Tryggvi Thor Herbertsson

A complete list of IoES Working Papers and Reprints can be accessed on our World Wide Web site at http://www.ioes.hi.is

W00:14 Sveinn Agnarsson: Productivity in Icelandic Fish Processing Industry 1985 – 1995: A Comparison of Methods

W00:15 Sveinn Agnarsson: Development of Efficiency in Icelandic Fish Processing Firms: A DEA Approach

W00:16 Jon Danielsson, Bjorn N. Jorgensen and Casper G. de Vries: Risk Management and Regulation in Incomplete Markets

W00:17 Ragnar Arnason, Gylfi Magnusson and Sveinn Agnarsson: The Norwegian Spring Spawning Herring Fishery: A Stylised Game Model

W00:18 Helgi Tomasson: Estimation of Correlations in Financial Markets when Trading is Infrequent

W00:19 Helgi Tomasson: Computations of Bayesian Estimators in ARMA Models

W00:20 Helgi Tomasson: Monitoring the trading intensity of a stock market under infrequent trading

W01:01 Tryggvi Thor Herbertsson: The Economics of Early Retirement

W01:02 Tryggvi Thor Herbertsson and J. Michael Orszag: The Costs of Early Retirement in the OECD

W01:03 Asta Herdis Hall and Solveig Frida Johannsdóttir: Generational Equality in Iceland

W01:04 Gylfi Zoega and Yu -Fu Chen: Exchange Rate Volatility as Employment Protection

W01:05 Tryggvi Thor Herbertsson and Gylfi Zoega: The Modigliani "Puzzle"

W01:06 Thorvaldur Gylfason: Lessons from the Dutch Disease: Causes, Treatment and Cures

W01:07 Tor Einarsson and Milton H. Marquis: Bank Intermediation over the Business Cycle

W01:08 Tor Einarsson and Milton H. Marquis: Bank Intermediation and Persistent Liquidity Effects in the Presence of a Frictionless Bond Market

W01:09 Tryggvi Thor Herbertsson, Edmund Phelps, and Gylfi Zoega: Demographics and Unemployment

W01:10 Tryggvi Thor Herbertsson: Shrinking Labour Forces and Early Retirement

W01:11 Tor Einarsson: Small Open Economy Model with Domestic Resource Shocks: Monetary Union vs. Floating Exchange Rate

W02:01 Tor Einarsson and Milton H. Marquis: Banks, Bonds, and the Liquidity Effect

W02:02 Fridrik M. Baldursson and Nils -Henrik M von der Fehr: Prices vs. Quantities: The Case of Risk Averse Agents

W02:03 Thorvaldur Gylfason: The Real Exchange Rate Always Floats

W02:04 Tor Einarsson: Small Open Economy Model with Domestic Resource Shocks: Monetary Union vs. Floating Exchange Rate

W02:05 Gudmundur Magnusson and Saso Andonov: Basel Capital Adequacy Ratio and the Icelandic Banking Sector: Quantitative Impact, Structural Changes and Optimality Considerations

W02:06 Alison L. Booth and Gylfi Zoega: If you're so smart, why aren't you rich? Wage inequality with heterogeneous workers