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  Gyl...Zoega

ISER, University of Essex  Department of Economics
Wivenhoe Park, Colchester  Birkbeck College
England CO4 3SQ  7-15 Gresse Street
London W1P 2LL

October 16, 2002

Abstract
This paper provides microfoundations for wage compression by mod-
eling wage-setting in a world of heterogeneous workers and ...rms. Workers
are differentiated 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
offers an explanation for why wage inequality has recently increased in
some countries: exogenous changes that increase labour market competi-
tion can disproportionately benefit higher ability workers and widen the
wages distribution.

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

JEL Classification: J24, J31, J42.

* We are grateful to John Drifill for helpful comments on an earlier draft.
Financial support from the European Commission TSER Award “Labour De-
mand, Education and the Dynamics of Social Exclusion” is gratefully acknowl-
edged.
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 finished 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 departments - 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 exogenously 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. A airline that wants to fly 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 fly the Concorde jet, or workers who want to reach for the sky in a more figurative sense, may thus face only a handful of potential employers; and this gives these employers monopsony power.\footnote{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 finite number of ...rms.}

Our model shows how wage compression arises quite naturally in labour markets due to differences 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 differ across tasks and employers. Section 5 outlines some applications of our model and also offers an explanation as to why within-group wage inequality has increased in many countries. Section 6 concludes.
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, firms form when workers - differentiated by their innate but observable ability - show up randomly at the workplace. The firm observes the average level of innate ability of its workforce, and this determines the complexity of tasks that can be undertaken. The firm then allocates workers of varying levels of ability across tasks, with the better workers doing the more complex tasks. There will be fewer firms 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 firms performing the more complex tasks, firms 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 monopsonistic profits, 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 differentiated by their innate but observable ability measured in terms of their endowment of efficiency 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

$$h_i \sim N_i, \frac{1}{2}$$

(1)

where $\bar{H}$ is the mean ability level in the finite population of available workers in the labour market. The quality of a firm 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, firms will differ in the average level of innate ability of their workforces. In this way worker heterogeneity
translates into \textit{rm} heterogeneity.\footnote{We consider in our model both horizontal and vertical job differentiation. The vertical differentiation 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 differentiation arises because, like Bhaskar and To (1999), we have jobs differing in terms of their non-wage characteristics, over which workers have varying preferences. This generates an imperfectly elastic labour supply to the \textit{rm}.} And \textit{rm} heterogeneity in our model – in terms of the average innate ability or human capital of its workers – affects the range of tasks that can be performed within the \textit{rm}. The higher the average level of innate ability of a \textit{rm}'s workforce, the greater is the range of tasks that can be performed at that \textit{rm}.\footnote{We assume that a single worker alone cannot perform a new task at a \textit{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.} Higher ability workers are allocated to the more complex tasks.

Examples of such production organisation are not hard to find. 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 flow 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 \textit{rms} doing the same task. By definition in our model there is only one \textit{rm} capable of performing the most advanced task, two \textit{rms} capable of performing the next-advanced task, and so on, while all \textit{rms} can do the simplest tasks. It is here that an employer's potential monopsony power lies. As the number of \textit{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 offsetting monopoly power. The reason for this is that, on their own, any one of them cannot enable a \textit{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 \textit{rm} an externality since her presence raises the average level of talent within that \textit{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 \textit{rm}. Think of a bright violinist offering her talents to the Royal Philharmonic or a computer whiz approaching Microsoft for a job. Only if all high-ability employees at a \textit{rm} co-ordinated their activities - for
example through a trade union – could this externality be internalised. 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 2 [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 differ in their within-rm average level of ability that we denote by H:

\[ H_R > H_{R-1} > \ldots > H_1; \quad H_i \sim N(\overline{H}, \sigma^2/2) \quad (2) \]

The distribution of average ability is normal with mean \( \overline{H} \) so that an average rm has average ability level \( \overline{H} \) 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 differs 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 1 performs all the same tasks apart from the most

---

4Thus 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

$$M_k = R \cdot k + 1$$

where $R$ is the total number of rms in the industry and $k$ is the number of tasks performed at a particular rm, $k \geq 1; R$. Equation (3) implies that there is one rm doing the most advanced task $k = R$, two ...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 efficiency unit of labour devoted to task $k$ measured in units of output coming from the rst task, which implies $y_1 = 1$ and $y_k \cdot y_1$ for all $k > 1$. We assume that the value of production – coming from each task – is a linear function of labour measured in efficiency 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 $y_1 h_1$ 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 staff.

Each rm undertaking task $k$ uses a technology that is constant-returns-to-scale with respect to the human capital it employs, $y_k h_k$. Total profits of a rm performing $k$ tasks – defined 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 $y_i w$ is simply the difference between the value of output and wages per efficiency unit of labour.

Profits of the $j$-th rm undertaking $k$ tasks can be written as

$$p_{jr} = \left( \frac{1}{y_i c_r n_r} \right)$$

where $\frac{1}{y_i}$ 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 \( \text{..rm; } k(H_i); \ k^0 > 0; \ k^0 = 0 \):

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 \( \text{..rm} \) - \( \text{..rms} \) set wages for workers performing each task. Obviously \( \text{..rms} \) 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 \( \text{..rm} \)'s task complexity has been established and \( \text{..rms} \) have allocated workers to tasks, workers observe that there are \( M_k \) \( \text{..rms} \) 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 \( M_k \) \( \text{..rms} \) 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 \( \text{..rst} \) consider the optimal wage rate for a given allocation of workers across \( \text{..rms} \) 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 \( \text{..rms} \) differ in the nonpecuniary attributes of the jobs they offer. We also suppose, following Salop (1970), that there are \( R \) \( \text{..rms} \) equally spaced around a circle of circumference \( C \) such that the distance between two adjacent \( \text{..rms} \) 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 \( \text{..rm} \). The more distant are the \( j \)-th \( \text{..rm} \)'s job characteristics \( x_j \) \( (j = 1; \ldots; M_k) \) from the worker's preferred characteristics \( x \), the larger is the worker's disutility cost (denoted by \( jx_i \cdot x_j \)) associated with employment at \( \text{..rm} \). We assume \( \text{cov}(h; x) = 0 \), since there is no reason to suppose that individual ability and preferences for job characteristics should be correlated.\(^5\)

Suppose \( \text{..rm} \) is the representative \( \text{..rm} \) of the \( M_k \) \( \text{..rms} \) 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 \( \text{..rms} \)'s job characteristics are equally spaced around the circumference, such that \( C = M_k \) is the distance between two adjacent \( \text{..rms} \) 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 \( k \)-th task are identical from the \( \text{..rm} \)'s perspective, since their productivity is observable as \( h_i \cdot y_k \).

\(^6\)Bhaskar and To (1999) and Booth and Zoega (1999) also assume that workers have heterogeneous 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 \(j\)th \(..\)rm 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 \(..\)rm's job characteristics \(x_j\) and the two adjacent \(..\)rms \(x_{j+1}\) and \(x_{j-1}\). Some workers in these two sub-segments (those closest to the \(j\)th \(..\)rm and further from the midpoints) are going to be more satisfied than workers whose preferences are closer to the midpoints between pairs of \(..\)rms. 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 \(w_j = j x_j x_{j,j}\), where we omit the subscript \(k\) for clarity. To find the labour pool for \(..\)rm \(j\), we first establish the upper and lower bounds for that \(..\)rm in job-characteristic space, represented by \(\bar{x}\) and \(\underline{x}\) respectively. Workers will choose to work at the \(..\)rm giving them the highest wage per efficiency unit net of their disutility associated with working at that \(..\)rm. First we consider the lower bound for the \(j\)th \(..\)rm. Given the wages \(w_{j-1}\) and \(w_j\) set by the two adjacent \(..\)rms, the marginal worker at \(x\) will be indifferent between working at \(..\)rm \(j\) and \(..\)rm \(j - 1\). Thus we find \(\bar{x}\) by solving the following:

\[
w_j = j x_j x_{j,1} = w_{j-1} + \left( x_j + x_{j,1} \right)
\]

which yields

\[
\bar{x} = \frac{w_{j-1} + \left( x_j + x_{j,1} \right)}{2}
\]

Similarly, the upper bound for the \(j\)th \(..\)rm \(x\) solves:

\[
w_{j+1} = j x_j x_{j,1} = w_j + \left( x_{j+1} + x_j \right)
\]

which yields

\[
\underline{x} = \frac{w_j + \left( x_{j+1} + x_j \right)}{2}
\]

Firm \(j\) doing the \(k\)th task sets wages for that task by maximising profits \(\frac{\partial \pi}{\partial w_k}\), written as:

\[
\max \frac{\partial \pi}{\partial w_k} = D_k (y_k, w_k) H_{jk} d\bar{x}_k = D_k (y_k, w_k) H_{jk} (x_k, x_k)
\]
where wages per efficiency unit of labour for the \( j \)th rm are denoted by \( w_k \) so the wages of a worker with human capital \( H_{jk} \) working on task \( k \) at the \( j \)th rm are \( H_{jk}w_k \).\(^7\) This brings us to our first proposition:

**Proposition 1**

The smaller is the number of rns \( M_k \) performing a given complexity of task, the larger is the gap between wages per efficiency 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:\(^8\)

\[
D_k H_{jk} (y_k w_k) = D_k H_{jk} (x_k - x_k)
\]

(10)

The term on the left-side of equation (10) shows the marginal benefit 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:

\[
(y_k w_k) = \frac{[(x_j + x_j) (w_j + w_j) + 2w_j]}{2}
\]

(11)

Now we know that rns are located symmetrically around the circle of job-characteristic space, such that the distance between rns is given by \( C = M_k \). Consequently \( x_j + x_j = 2C = M_k \). In symmetric equilibrium with \( w_j = w_j = w_j \) we can now simplify the equation above to obtain:

\[
w_{jk} = y_k \frac{C}{M_k}
\]

(12)

Notice in (12) that as \( M_k \rightarrow \infty \), \( w_{jk} \rightarrow y_k \). This is the wage scenario for a perfectly competitive labour market. However, as \( M_k \rightarrow \infty \), \( w_{jk} \rightarrow y_k \).\(^9\)

The intuition behind this result is as follows: When there is a lot of heterogeneity for a given number of rns - C is large - each rm is faced with many workers who prefer its job characteristics to those of other rns. This makes the cost of raising wages high because by offering higher wages - aimed

\(^7\)Note that we have made the assumption that rns do not observe the ability levels of marginal workers and as a result use the average ability level \( H_{jk} \) in the proft function. This simplifies the exposition without any loss of generality.

\(^8\)The second order condition holds, \( \frac{d}{d w_j} \frac{d y_k}{d w_j} < 0 \).

\(^9\)In order to ensure \( w_j > 0 \), we need to impose the restriction that \( y > C \).
at attracting new recruits whose preferences do not match the jobs being offered 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 ...ms - M_k is large - then not many workers would settle for any given ...m in the absence of wage incentives. This reduces the marginal cost of offering higher wages. It follows that as the number of ...ms M_k goes up, the marginal cost of raising wages falls, which makes the wage level rise towards marginal productivity as the number of ...ms approaches infinity.

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

$$n_k = (x_k \cdot x_k) D_k = D_k \cdot w_j \cdot \frac{1}{2} (w_j + w_{j+1}) + \frac{C}{M_k}$$

and is increasing in the ...m's own wage, in contrast to the perfectly competitive case where the labour supply is infinitely 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 ...ms competing for the workers M_k. The more workers there are of each preference type, and the more heterogeneous are workers, the greater the number of workers who want to join a particular ...m. However, taking the density and the heterogeneity as given, a greater number of ...ms implies that fewer workers will apply to any single ...m.

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

$$y_k \cdot w_j \cdot h \cdot c_k$$

The term on the left-hand side is the expected profits from employing worker i. Only if this condition holds does the ...m allocate and train a worker for task k. Since the number of workers applying for jobs at the representative ...m is given by supply - and the ...m 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 ...m.

Assume that there is a worker with ability level h_i doing task r - the most sophisticated within ...m r - and another doing the least advanced task - task 1 - with a higher ability level h_i; h_i > h_1. If the ...m could, ex-post, relocate them at zero cost - that is recant its earlier decision and assign workers differently to the tasks - so that the latter would be trained to do task r and the former to do task 1, the expected benefit to the ...m would be the following since total training costs would be unchanged:
which is positive if and only if

\[(y_r \cdot w_r^c) \cdot (y_1 \cdot w_1^c) > 0\]  

(15)

that is if the rm benefits 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:

\[h_j \leq \frac{G_R}{y_R \cdot w_R^c}\]  

(17)

Because of the high degree of monopsony for task R, the rm benefits 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:

\[\frac{G_{R-1}}{y_{R-1} \cdot w_{R-1}^c} \cdot h_i < \frac{G_R}{y_R \cdot w_R^c}\]  

(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 profit 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 sufficiently able for task R will only be trained to do task R-1 if lower training costs offset the effect of lower monopsony power. If not, he and all others who are not sufficiently 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 sufficiently good for task 1:

\[h_i < \frac{c_1}{y_1 \cdot w_1^c}\]  

(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 flying the most recent types of aircraft. Similarly, at a top-ranking university, the most talented members of staff 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 fly the Concorde, an airline gains monopsony power in its relations with high-ability individuals. This enables it to profit 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 finishing touches can ruin the work performed doing all earlier tasks. Hence they show that a firm should put the more skilled and reliable workers in charge of the final tasks. While our model is also directed at task complexity, it is in a different 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 flight-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 find that the more able workers should be delegated to the more advanced tasks, because the firm then reaps maximum profits from the greater monopsony position associated with the more difficult tasks, as we show below. It is more profitable 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 profits from performing the task within the firm: all firms can perform the simplest task, hence the level of competition is highest for this task. But as firms 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 efficiency 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 profits increase as we move to more advanced tasks. Combining the two results gives a proposition involving wage compression, which we now define.

Definition 1 Absolute wage compression occurs when the difference between the value of output produced by two workers of different ability \( h_i > h_j \) doing
separate tasks \( k \) and \( k_1 \) is greater than the difference between their wages, that is \( h_i y_k - h_i y_{k_1} > h_i w_k - h_i w_{k_1} \).

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:

\[
h_i y_k - h_j y_{k_1} > h_i w_k - h_j w_{k_1}
\]

**Proof:** This proposition is easily verified using our earlier equations. We know that

\[
h_i (y_k - w_k) > h_j (y_{k_1} - w_{k_1})
\]

since according to Proposition 1

\[
y_k - w_k > y_{k_1} - w_{k_1}
\]

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

\[
h_i y_k - h_j y_{k_1} > h_i w_k - h_j w_{k_1}
\]

which is our definition 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 specific 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 firm-sponsored training. In our model, we can show that - if a high-ability worker doing an advanced task were to quit or die - the firm may benefit from retraining a lower-ability worker doing a less advanced task to take her place. This is due to the wage compression.

¹⁰Recent contributions to the wage compression literature offer explanations such as minimum wages and other wage floors, unions, asymmetric information about a worker’s skills between current and prospective employers, and the complementarity between firm-specific and general skills (alternatively between physical capital and general skills). In contrast, our non-representative-agent model makes these mechanisms unnecessary, although not irrelevant.
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 firms 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 firm’s monopsony profits 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 fiction 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 firms 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 firms 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 firm’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

11Saint-Paul (2001) also looks at explanations for widening wage inequality in a model with a fixed supply of heterogeneous workers. We differ 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 effectively 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-specific 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 on 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


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

W99:09 Kenneth F. Wallis: Macroeconometric Modelling

W99:10 Tryggvi Thor Herbertsson, Marta G. Skúladóttir, and Gylfi Zoega: Three Symptoms and a Cure: A Contribution to the Economics of the Dutch Disease


W99:13 Ragnar Arnason: Costs of Fisheries Management: Theoretical and Practical Implications

W99:14 Ragnar Arnason: Economic Instruments to Achieve Ecosystem Objectives in Fisheries Management

W99:15 Ragnar Arnason: Property Rights as a Means of Economic Organization

W00:01 Jerry Coakley, Ana-Maria Fuertes and Gylfi Zoega: Testing the Persistence and Structuralist Theories of Unemployment

W00:02 Thrainn Eggertsson: Norms in Economics – With Special Reference to Economic Development

W00:03 Thorvaldur Gylfason: Growing Apart

W00:04 Jon Danielsson: The Emperor has no Clothes: Limits to Risk Modelling

W00:05 Thorolfur Matthiasson: The Icelandic Debate on the Case for a Fishing Fee: A Non-Technical Introduction

W00:06 Willem H. Buiter: Is Iceland an Optimal Currency Area?

W00:07 Alison L. Booth and Gylfi Zoega: Why do Firms Invest in General Training? ‘Good’ Firms and ‘Bad’ Firms as a Source of Monopsony Power

W00:08 Eduard Hochreiter: “Exchange rate regimes and capital mobility: Issues and some lessons from central and eastern European applicant countries”

W00:09 Thorvaldur Gylfason: Fix or Flex? Alternative Exchange Rate Regimes in an Era of Global Capital Mobility

W00:10 Thorvaldur Gylfason: Natural Resources, Education and Economic Development


W00:12 Thorolfur Matthiasson: Changing Rules for Regulation of Icelandic Fisheries

W00:13 E. Tumusiime-Mutebile: Economic Reforms and their Impact in Uganda
<table>
<thead>
<tr>
<th>Session</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>W00:15</td>
<td>Sveinn Agnarsson: Development of Efficiency in Icelandic Fish Processing Firms: A DEA Approach</td>
</tr>
<tr>
<td>W00:16</td>
<td>Jon Danielsson, Bjorn N. Jorgensen and Casper G. de Vries: Risk Management and Regulation in Incomplete Markets</td>
</tr>
<tr>
<td>W00:17</td>
<td>Ragnar Arnason, Gylfi Magnusson and Sveinn Agnarsson: The Norwegian Spring Spawning Herring Fishery: A Stylised Game Model</td>
</tr>
<tr>
<td>W00:18</td>
<td>Helgi Tomasson: Estimation of Correlations in Financial Markets when Trading is Infrequent</td>
</tr>
<tr>
<td>W00:19</td>
<td>Helgi Tomasson: Computations of Bayesian Estimators in ARMA Models</td>
</tr>
<tr>
<td>W00:20</td>
<td>Helgi Tomasson: Monitoring the trading intensity of a stock market under infrequent trading</td>
</tr>
<tr>
<td>W01:01</td>
<td>Tryggvi Thor Herbertsson: The Economics of Early Retirement</td>
</tr>
<tr>
<td>W01:02</td>
<td>Tryggvi Thor Herbertsson and J. Michael Orszag: The Costs of Early Retirement in the OECD</td>
</tr>
<tr>
<td>W01:03</td>
<td>Asta Herdis Hall and Solveig Frida Johannsdóttir: Generational Equality in Iceland</td>
</tr>
<tr>
<td>W01:04</td>
<td>Gylfi Zoega and Yu-Fu Chen: Exchange Rate Volatility as Employment Protection</td>
</tr>
<tr>
<td>W01:05</td>
<td>Tryggvi Thor Herbertsson and Gylfi Zoega: The Modigliani “Puzzle”</td>
</tr>
<tr>
<td>W01:06</td>
<td>Thorvaldur Gylfason: Lessons from the Dutch Disease: Causes, Treatment and Cures</td>
</tr>
<tr>
<td>W01:07</td>
<td>Tor Einarsson and Milton H. Marquis: Bank Intermediation over the Business Cycle</td>
</tr>
<tr>
<td>W01:08</td>
<td>Tor Einarsson and Milton H. Marquis: Bank Intermediation and Persistent Liquidity Effects in the Presence of a Frictionless Bond Market</td>
</tr>
<tr>
<td>W01:09</td>
<td>Tryggvi Thor Herbertsson, Edmund Phelps, and Gylfi Zoega: Demographics and Unemployment</td>
</tr>
<tr>
<td>W01:10</td>
<td>Tryggvi Thor Herbertsson: Shrinking Labour Forces and Early Retirement</td>
</tr>
<tr>
<td>W01:11</td>
<td>Tor Einarsson: Small Open Economy Model with Domestic Resource Shocks: Monetary Union vs. Floating Exchange Rate</td>
</tr>
<tr>
<td>W02:01</td>
<td>Tor Einarsson and Milton H. Marquis: Banks, Bonds, and the Liquidity Effect</td>
</tr>
<tr>
<td>W02:02</td>
<td>Fridrik M. Baldursson and Nils-Henrik M von der Fehr: Prices vs. Quantities: The Case of Risk Averse Agents</td>
</tr>
<tr>
<td>W02:03</td>
<td>Thorvaldur Gylfason: The Real Exchange Rate Always Floats</td>
</tr>
<tr>
<td>W02:04</td>
<td>Tor Einarsson: Small Open Economy Model with Domestic Resource Shocks: Monetary Union vs. Floating Exchange Rate</td>
</tr>
<tr>
<td>W02:05</td>
<td>Gudmundur Magnusson and Saso Andonov: Basel Capital Adequacy Ratio and the Icelandic Banking Sector: Quantitative Impact, Structural Changes and Optimality Considerations</td>
</tr>
<tr>
<td>W02:06</td>
<td>Alison L. Booth and Gylfi Zoega: If you’re so smart, why aren’t you rich? Wage inequality with heterogeneous workers</td>
</tr>
</tbody>
</table>