Changes In The Distribution Of Male And Female Wages Accounting For Employment Composition

Richard Blundell, Amanda Gosling, Hidehiko Ichimura and Costas Meghir[1]

Institute for Fiscal Studies

February 2002

[1]Richard Blundell, Hidehiko Ichimura, and Costas Meghir are at the Department of Economics, University College London and at the Institute for Fiscal Studies, Amanda Gosling is at the Department of Economics, University of Essex and also at the Institute for Fiscal Studies.

Abstract

This paper presents estimates of the changing distribution of wages that are robust to possible selection effects. We find convincing evidence of an increase in overall inequality, changes in the “return” to education and increases in inequality within age and education groups. On the other hand we find that the increase in the relative wages of women may have been driven by selection.

1 Introduction and motivation

There has been a vast literature on the way the distribution of observed wages has evolved over the past twenty to thirty years. Some economies, such as the US and the UK have seen large and unprecedented increases in inequality among the wages of workers. This illustrated in Figure 1 where we show the way that the interquartile range of male and female log hourly wages has evolved for those who work. These increases in inequality have been associated with increased returns to education, cohort effects and increases in the returns to unobserved skill.[1] A variety of interpretations have been given as to why these events have occurred; these include skill biased technical change, globalization induced increased competition for low skill workers and changes in the supply of graduates. Gosling, Machin and Meghir (2000) show that the increases can be attributed to permanent differences across cohorts and in changes in the returns to education.

Another aspect of the debate about changes in the distribution of wages concerns the relationship of male and female wages. Generally, both in the US and the UK the consensus is that female wages among workers have increased and converged with male wages.

However, in parallel with these momentous changes in the distribution of observed wages, the labour market participation rates for males and females have changed in dramatic ways. Female participation has tended to increase, particularly among women married to employed men. Male participation on the other hand has decreased at all ages to a very large degree. (see Figure 2). The decline in male participation is not confined to older men, reflecting the increase in early retirement. As illustrated in Figure 3 the decline over time occurs at

[1] See Juhn, Murphy and Pierce (1993) for the US and Gosling, Machin and Meghir (2000) for the UK.

all ages, although it is even more pronounced for men close to 55. Female employment on the other hand has risen (see same figures) and this has been particularly true for women younger than 35. Looking at figure 4 we also see that the change in employment has been heavily skill biased. Although male employment has declined for all education groups the highest decline has been for the least skilled group. Moreover for women the unskilled group has shown a slight decline, while most of the increase can be accounted for by the increase in employment of women with more than the lowest level of education.

To the extent that changes in participation are related to wages, part of the changes in the distribution may have been induced by changes in the composition of those employed in terms of their unobserved characteristics. Moreover, since the changes in participation are quite complex it is not at all clear how the composition of those employed has changed in terms of unobservables.

This point has a bearing on the interpretation we give to changes in the distribution of wages. For example in order to interpret the change in return to education as being driven to an increase in the relative demand for higher educated workers (due say to skill biased technical change) we need to establish that the change in the observed return is not an artifact of changes in the composition of employment. In addition it is not obvious how composition is changing. For some (perhaps for young men), the dominant factor in the participation decision will be the unobserved (and observed) productivity. For others, one of the crucial factors will be accumulated wealth (say for those close to retirement). The combination of these two factors implies that we are not necessarily able to state whether it is predominantly the less or the more productive who have been leaving the labour force.

This issue can be potentially resolved with qualitative assumptions, such as an exclusion restriction, combined with parametric assumptions, such one restricting the unobservables to be jointly normally distributed. As far as the impact of composition effects on average wage growth is concerned Blundell, Reed and Stoker (1999) have used information on housing costs to use policy changes as an exclusion restriction for sorting out composition effects from genuine growth. However, the exclusion restrictions they use are better justified for the mean than for the entire distribution of wages. Moreover, the exclusion restriction they use itself is insufficient for the non-parametric identification of the entire distribution of wages.

In this paper we develop new estimation methods of the bounds of the distribution that takes into account the composition effect. We use these methods to bound every quantile of the wage distribution for which bounds can be computed.

We first evaluate the worst case bounds examined by Manski (1994). We then discuss and develop methods for obtaining tighter bounds exploiting various forms of assumptions. Next we describe an economic model of wage determination with which our results are best understood.

2 A broad description of our approach and the worst case bounds

In this section we provide a simple description of our approach. This has applicability to all selection problems, although the focus of the current paper is on understanding the evolution of the wage distribution. As can be expected the interpretation of our results depends on the underlying model of wage determination and we discuss this later in the paper.

The approach that is often followed in the literature is to use a parametric selection model (usually for the mean). This depends on distributional assumptions and/or exclusion restrictions. In practice these are often hard to justify from economic theory. The aim of this paper is different. It is to evaluate what can be said without any (or only minimal) restrictions.

Let y and x denote the dependent variable and the conditioning variable, respectively. In our case the dependent variable should be taken to be the log wage and x should be understood to include gender, age, education and other characteristics. When y is observed, the indicator variable I equals 1 and when y is not observed, I equals 0. In our case I indicates whether the person is working or not. The probability of $I = 1$ given x is written as $P(x)$ We write the conditional distribution of y given x by $F(y|x)$, the conditional distribution of y given x and $I = 1$ by $F(y|x, I = 1)$, and the conditional distribution of y given x and $I = 0$ by $F(y|x, I = 0)$. Then we can write the distribution of wages as

$$F(y|x) = F(y|x, I = 1) P(x) + F(y|x, I = 0) [1 - P(x)] \tag{1}$$

Given that the data identify $F(y|x, I = 1)$ and $P(x)$ (and $1 - Px$) the problem can be respecified as one in which only $F(y|x, I = 0)$ is unknown.

Our starting point for the analysis is the work by Manski (1994) who notes that once the inequality:

$$0 \leq F(y|x, I = 0) \leq 1.$$

is substituted into equation 1, the bounds to the cumulative distribution function can be derived as in equation 2 below[2]

$$F\left(y|x,I=1\right)P\left(x\right)\leq F(y|x)\leq F\left(y|x,I=1\right)P(x)+\left[1-P\left(x\right)\right]. \tag{2}$$

As cumulative distribution functions are monotonic the bounds to the conditional quantiles can then be easily estimated (see below).

The basic idea underlying these worst case bounds is quite simple: it must be the case that assuming that all non-workers are less productive than workers will give us the lowest possible value to the estimated quantiles of wages, conversely assuming that all non workers are more productive than workers will give us the highest possible value. The “truth” must either lie at one of these two extremes or somewhere in the middle.

Assuming that nonworkers are less productive than workers can be justified by the one factor human capital model. Alternatively, suppose that all non-workers are more productive than the non-workers. This will be the case, for instance, in a model where reservation wages depend on accumulated wealth. Hence among older individuals the most productive will have accumulated more wealth and will be more likely to retire early. As this discussion illustrates, which bound is considered more realistic will depend on the characteristics of the subpopulation in question. In general however it is hard to exclude one or the other possibility.

[2]Proposition 3, p.152.

2.1 Estimation of The Worst Case Bounds

As cumulative distribution functions are monotonic the bounds to the conditional quantiles can be derived from equation 2 above in the following way:

$$F^{-1}(\frac{\alpha-1+Px}{P(x)}|I=1,x) \leq F^{-1}(\alpha|x) \leq F^{-1}(\frac{\alpha}{P(x)}|I=1,x) \tag{3}$$

where $F^{-1}(\alpha|x)$ as the inverse of the cumulative distribution function defines the y at the α quantile Thus the left hand side of equation 3 defines the lower bound to the conditional quantile and the right hand side the upper bound. The gap between the upper and lower bounds will be negatively related to $P(x)$, approaching zero as $P(x)$ approaches 1. The bounds will also be tighter for those observations where $\alpha/P(x)$ and $[\alpha-1+P(x)]/P(x)$ lie in the area where the right derivative of $F(y|x,I=1)$ is large. Lower and higher percentiles are thus unlikely to have tight bounds when y is both unbounded below and above. In addition, these bounds cannot be defined in all cases. When $P(x)<\alpha$, $\frac{\alpha}{P(x)}>1$ and we cannot measure the upper bound. When $P(x)<1-\alpha$, $\frac{\alpha-1+P(x)}{P(x)}<0$ and we cannot then measure the lower bound.

Note that equation 3 implies

$$q_{\alpha}^{u}(x) = F^{-1}(\frac{\alpha}{P(x)}|x,I=1)$$
$$q_{\alpha}^{l}(x) = F^{-1}(\frac{\alpha-1+P(x)}{P(x)}|x,I=1)$$

where $q_{\alpha}^{u}(x)$($q_{\alpha}^{\ell}(x)$) refers to the upper(lower) bound to the α quantile. These bounds can then be estimating by finding the relevant quantile (a transformation of α) from the sample of those with observed y. Manski thus proposes estimation of the bounds using nonparametric kernel estimation of $F(y|x, I=1)$ and $P(x)$. From this the quantiles can easily be measured using the equalities above. Our approach instead involves a simple adaptation of the usual quantile regression procedure (see Koenker and Basett (1978)).

Let $\Psi_\alpha(s)$ be the check function introduced by Koenker and Basset (1978):

$$\Psi_\alpha(s) = \begin{cases} \alpha \cdot s \text{ if } s \geq 0, \\ (\alpha - 1) \cdot s \text{ if } s < 0, \end{cases}$$

Let $\hat{P}(x)$ be an estimator of $P(x)$, and $K(\cdot)$ be a d-dimensional kernel function to accommodate d regressors x.[3] The solution $\hat{\theta}_0$ that corresponds to the minimizer $\left(\hat{\theta}_0, \hat{\theta}_1\right)$ of

$$\left(\hat{\theta}_0, \hat{\theta}_1\right) = \arg\min_{\theta_0,\theta_1} \sum_{i=1}^{n} \Psi_{\alpha/\hat{P}(x_i)}\left(y_i - \left[\theta_0 + \theta_1\left(x_i - x_0\right)\right]\right) K\left(\frac{x_i - x_0}{h_n}\right) I\left(\hat{P}(x_i) > \alpha\right)$$

is then a local linear estimator of the upper bound function. A nonparametric estimator of the lower bound function can be constructed simply by substituting $\frac{\alpha-1+P(x)}{P(x)}$ for $\frac{\alpha}{P(x)}$. Preliminary analysis shows that this estimator and Manski's estimator have the same asymptotic distribution when θ_1 is set to the zero vector.

When the number of regressors k is high, the inefficiency of nonparametric methods limits applicability of the above approach. Thus we also consider a semiparametric method that reduces the dimensionality problem. Thus we also consider a semiparametric method that

[3] We assume d-continuous regressors for ease of exposition.

reduces the dimensionality problem. The baseline model we consider is

$$y_i = \gamma_0 + x_i'\beta_0 + \varepsilon_i$$

$$D_i = \begin{cases} 1 \text{ with probability } P(x_i) \\ 0 \text{ with probability } 1 - P(x_i) \end{cases}$$

with the added assumption that the conditional distribution of ε_i given D_i and x_i depends only on $P(x_i)$. In this case

$$F(y - \gamma_0 - x_i'\beta_0 | x_i, I = 1) P(x_i) = F(y - \gamma_0 - x_i'\beta_0 | P(x_i), I = 1) P(x_i).$$

Hence when we condition on a fixed value of $P(x_i)$, we can estimate the upper bound of the quantiles of y_i based on the sample of workers ($I_i = 1$) from

$$\left(\hat{\theta}_0, \hat{\theta}_1, \hat{b}\right)$$
$$= \arg \min_{\theta_0, \theta_1, b} \sum_{i=1}^{n} \Psi_{\alpha/\hat{P}(x_i)} \left(y_i - \theta_0 - \theta_1 \left[\hat{P}(x_i) - \hat{P}(x_0)\right] - x_i'b\right) K\left(\frac{\hat{P}(x_i) - \hat{P}(x_0)}{h_n}\right) I\left(\hat{P}(x_i) > \alpha\right).$$

Note that $\hat{\theta}_0$ is an estimator of

$$\gamma_0 + F^{-1}(\alpha/P(x_0) | P(x_0), I = 1)$$

which is the upper bound of the quantile of the unconditional distribution. In turn $\hat{b}$ is an estimator of β_0. As one can define an estimator for each value of x_0, a natural way to combine is to define the estimator as

$$\left(\hat{\theta}_{0j}, \hat{\theta}_{1j}, \hat{b}\right)_{j=1}^{n} = \arg\min_{\theta_0,\theta_1,b} \sum_{j=1}^{n}\sum_{i=1}^{n} \left[\Psi_{\alpha/\hat{P}(x_i)}\left(y_i - \theta_{0j} - \theta_{1j}\left[\hat{P}\left(x_i\right) - \hat{P}\left(x_j\right)\right] - x_i'b\right)\right.$$
$$\left. \times K\left(\left[\hat{P}\left(x_i\right) - \hat{P}\left(x_j\right)\right]/h_n\right) I\left(\hat{P}\left(x_i\right) > \alpha\right)\right].$$

When the objective function is least square based, the estimator is analogous to the Robinson's (1989) estimator. The lower bounds can be estimated analogously.

As we shall see, the worst case bounds are already informative in some cases but not necessarily in many others. We next consider how we can exploit auxiliary assumptions to produce tighter bounds.

2.2 Estimating bounds to within group inequality

Within our framework one measure we will be estimating are the differences between any two quantiles, such as

$$D(x) = y^{q_2}|_x - y^{q_1}|_x \tag{4}$$

In doing this we can exploit restrictions on the distribution functions that imply that this is monotonic. In particular we need to ensure that the bounds for $D(x)$ are consistent with a monotonically increasing distribution of wages for non-workers (which we do not observe). To see how this affects the bounds to $D(x)$first note that

$$F(y|x, I=0) = \frac{F(y|x) - P(x)F(y|x, I=1)}{P(x)}$$

Assuming differentiability for simplicity to ensure that $\frac{\partial F(y|x,I=0)}{\partial y} > 0$ the above expression implies that

$$\frac{\partial F(y|x)}{\partial y} > P(x)\frac{\partial F(y|x, I=1)}{\partial y}$$

Now consider the upper bound of $D(x)$, say $D^u(x)$.This will be max $D(x)$ such that a. $D(x) \geq 0$, $y^{q_2}|_x$ and $y^{q_1}|_x$ lie within their respective bounds and such that the pair of $y^{q_2}|_x$ and $y^{q_1}|_x$ lie on a curve with the same gradient as $P(x)\frac{\partial F(y|x,I=1)}{\partial y}$ or steeper. Similarly for the lower bound.

3 Restrictions for tightening the bounds

3.1 The distribution of potential wages for non-participants - the median restriction

One of the principal motivations of this paper is to understand the extent to which we can characterise the changes in economic opportunities without imposing strong theoretical or functional form restrictions. However some restrictions may be considered acceptable and turn out to help tighten the bounds considerably. The leading case we have in mind is imposing the restriction that the q-quantile of $F(y|x, I=0)$ is lower than the q-quantile of $F(y|x, I=1)$. For example, we may be willing to assume that the median (potential) wage of the non-working population is less than the median wage of the working population, both conditional on any observable values x.

More generally, we consider restrictions that change the bounding function specified in equation 2 to the form

$$F(y|x, I=1)\,P(x) + G(y|x)\,[1-P(x)]\,.$$

where $G(y|x)$ is an (incomplete) conditional CDF. For example, if we impose that the median wage of the non-working population is less than the median wage of the working population (denoted $q^1_{0.5}(x)$), both conditional on some observables x, then the lower bound function changes to the function of this form with

$$G(y|x) = \begin{cases} 0 \text{ if } y < q^1_{0.5}(x)\,, \\ 0.5 \text{ if } y \geq q^1_{0.5}(x)\,. \end{cases}$$

Hence there is no change to the lower function restriction where $y < q^1_{0.5}(x)$ and at $y = q^1_{0.5}(x)$ the function value jumps to 0.5. Before, the function value was $0.5P(x)$. After this the bound improves to be $F(y|x, I=1)\,P(x) + 0.5\,[1-P(x)]$. Thus in this example, the new restriction is informative about any quantiles above $0.5P(x)$.

One can see that so long as we impose the restriction at a point as above, one can apply the same estimation method with little modification. Note that a very strong restriction would be that this is valid for all quantiles. In this case the workers wage distribution dominates that of the non-workers (i.e. stochastic dominance). We explore the use of this (untestable) restriction.

These two examples are special in that $G(y|x)$ does not contain information other than $F(y|x, I=1)$. More generally one can generate the data from $G(y|x)$ combined with those generated under $F(y|x, I=1)$ to obtain the lower and upper bounds to the quantiles.

3.1.1 The plausibility of the median restriction

The median restriction can be rethought of as one about how the probability of work varies across potential wages. We can respecify the observed distribution functions as follows

$$F(y|x, I = 1) = \frac{\Pr(I = 1|x, Y < y)}{P(x)} F(y|x)$$

Thus $F(y|x, I = 1) \leq F(y|x)$ implies

$$\Pr(I = 1|x, Y < y) \leq P(x)$$

If workers adopt a reservation wage strategy and move into work once their wage moves above a certain threshold, then holding everything else constant, increases in wages will result in increases in the probability of work. As the algebra above shows, this will be consistent with the median restriction. The problem (and rejection of our restriction) will occur if unobserved determinants of the probability of work are strongly negatively correlated with wages.

There are various reasons why this could be the case. First, if wages are correlated over time, higher waged people are more likely to have accumulated assets which may increase their reservation wages. Second if high waged women are matched with high waged men, then the out of work income of women could be increasing in their potential earnings. Lastly as higher waged women tend to delay rather than avoid childbirth, it could be the higher waged women in older age groups who have pre-school children. We suspect that the median restriction will be less plausible for women and for those over 50.

The graph below (figure 5) uses longitudinal data from the British Household Panel Survey 1991-9 (BHPS) to show, that such effects may not be that important. Here we estimated cross sectional wage equations for each year and allocated workers a "residual" (i.e. actual wage minus predicted wage). We then split the sample into those who were unemployed either last year or will be unemployed next year and into those who were not. Figure 5 shows that the distribution of wages of those in work lies below the distribution of those who have been out of work, even controlling for factors such as age and education which are important determinants of unemployment. Hence the median restriction we impose may not be that unreasonable. In fact this graph, taken at face value would provide support for the stochastic dominance assumption.

3.2 Independence restrictions

We next consider exploiting independence restrictions. Suppose we partition the vector of observables into the sub-vectors x_1 and x_2 and suppose that wages can be written as

$$y_i = m(x_i^1, x_i^2) + \varepsilon_i$$

where $F(\varepsilon|x_1, x_2) = F(\varepsilon|x_1)$ In this case none of the quantiles of ε depend on x_2. Hence we can write the qth quantile of y as

$$y^q\left(x_i^1, x_i^2\right) = m(x_i^1, x_i^2) + g(q, x_{i1})$$

In this context the impact of changing x_2 (say education) is easily defined. Moreover the independence restriction can be used to obtain a tight bound for such a return. First note that the impact of x_2 on wages, defined by $y^q\left(x_i^1, x_i^2 = A\right) - y^q\left(x_{1i}, x_{2i} = B\right) = \Delta_{x_2} m(x_{1i}, x_{2i})$ does not depend on q.[4] Then under these assumptions the tightest bound on the return $\Delta_{x_2} m(x_{1i}, x_{2i})$ can be obtained by searching across quantiles. Thus we have that the tightest bound takes the form

$$\max_q \left\{ y^{qLower}\left(x_{1i}, x_{2i} = A\right) - y^{qHigher}\left(x_{1i}, x_{2i} = B\right) \right\}$$

$$\leq \Delta_{x_2} m(x_{1i}, x_{2i}) \leq$$

$$\min_q \left\{ y^{qHigher}\left(x_{1i}, x_{2i} = A\right) - y^{qLower}\left(x_{1i}, x_{2i} = B\right) \right\}$$

where "Higher" and "Lower" refer to the upper and lower bound functions on the quantiles.

As we discuss below, the interpretation of some of our results depends on the separability of the unobservables from the observables, i.e. on index sufficiency type assumptions. Hence testing such assumptions becomes of central importance. Moreover in the first part of the paper we do not actually use these assumptions, so that for some purposes our results are much more general than would be implied by the imposition of such restrictions. Bounds however, can become much tighter by imposing such restrictions on the distribution of wages. In this paper we consider using a restriction that implies that education affects the location of the distribution only (conditional on time and age). We show how such a restriction can tighten the bounds and importantly how this restriction can be tested within our non-

[4] The analysis could also be carried out in terms of a continuous variable

parametric framework.

3.3 Exploiting information from a determinant of participation

This next section examines how economic restrictions about the relationships between wages and another set of variables (z) can be used to obtain tighter bounds to the estimated quantiles[5]. One example of this would be an exclusion restriction. We show, however, that other weaker assumptions may also be employed. We also show how these restrictions can be rejected empirically.

3.3.1 An exclusion restriction

Manski (1994) shows that if y is independent of z conditional on x i.e.

$$F(y|x,z) = F(y|x)\ \forall y,x,z$$

then

$$\max_z q^l(x,z) \leq q(x) \leq q^u(x,z)$$

where $q(x)$ denotes the q^{th} quantile of wages conditional on x, and $q^{l(u)}$ denotes the lower (upper) bound to the q^{th} quantile. Implementation is thus simply a matter of finding the bounds to the quantiles of wages conditional on x and z and then searching across z within x to find the lowest upper bound and the highest lower bound.

It might be thought that these minima and maxima will be found at the value of z

[5]The analysis here draws heavily on the work of Manski (94) and Manski and Pepper (2000)

where the proportion of those with $I = 1$ is highest. This is not necessarily the case. To illustrate this note that we can rewrite the lower bound to the estimated distribution function $(F(y|I = 1, x, z) \Pr(I = 1|x, z))$ when z is independent of y as

$$F(y|x) \Pr(I = 1|Y < y, x, z)$$

Thus the difference across two values of z will become

$$F(y|x) \left[\Pr(I = 1|Y < y, x, Z = z_2) - \Pr(I = 1|Y < y, x, Z = z_1)\right]$$

This will be positive when $\Pr(I = 1|Y < y, x, Z = z_2) > \Pr(I = 1|Y < y, x, Z = z_1)$. $Pr(I = 1|x, Z = z_2) > Pr(I = 1|x, Z = z_1)$ is neither a necessary or a sufficient condition for this. One could think of an example (albeit an unlikely one) where $\Pr(I = 1|x)$ does not vary across z, but the $I = 1$ sample are positively selected for some values of z, randomly selected for others and for the last group negatively selected. One important difference between this model and many other selection models is that the latter usually assume the process (but obviously not the degree) of selection to remain the same across z. COSTAS DO WE WANT MORE HERE?

There is nothing in the estimation procedure suggested that forces the min of the upper bounds to lie above the max of the lower bounds. Thus, in cases where y is not in fact independent of z it is perfectly possible for the bounds to cross and the upper bound for some values of z to lie below the lower bound for others. In these cases we can reject the exclusion restriction. Thus, just as in the independence restriction discussed above, the data together with the worst case bounds provide us with a weak test of our identification restrictions.

3.3.2 A monotonicity restriction

Strong exclusion restrictions of the type discussed above may not be available. We might, however, be prepared to assume the direction of the relationship between y and z. (This idea is very similar to the monotone instrument variable (MIV) for the mean proposed by Manski and Pepper (2000) and the following analyses borrows from their exposition.) The restriction then becomes:

$$F(y|x, z \geq z_1) \leq F(y|x, z \leq z_1) \quad \forall y, x$$

Thus we must then get

$$F(y|x, z_1) \geq \max_{z>z_1} F(y|x, z, I = 1)P(x, z)$$

and

$$F(y|x, z_1) \leq \min_{z<z_1} F(y|x, z, I = 1)P(x, z) + 1 - P(x, z)$$

Then so long as it is the case that

$$F(y|x, z_1, I = 1)P(x, z_1) \leq \max_{z>z_1} F(y|x, z, I = 1)P(x, z)$$

or

$$F(y|x, z_1, I = 1)P(x, z1) + 1 - P(x, z_1) \geq \min_{z<z_1} F(y|x, z, I = 1)P(x, z) + 1 - P(x, z)$$

we will get tighter bounds for the conditional distribution where $z = z_1$.

Implementation The bounds to the distribution of wages conditional only on x are then estimated in the following way. First find for each value of z the upper ($UB^y_{z_i|x}$) and lower ($LB^y_{z_i|x}$) bounds that are consistent with the assumption that $F(y|x,z)$ is never increasing in z and then average over the sample to get bounds to $F(y|x)$.

$$\sum_{i=1}^{N}\left[LB^y_{z_i|x}\Pr(Z=z_i|x)\right] \leq F(y|x) \leq \sum_{i=1}^{N}\left[UB^y_{z_i|x}\Pr(Z=z_i|x)\right]$$

Once upper and lower bounds for the distribution function conditional on only x are obtained, it is then easy to use it to estimate various quantiles.

To see an easy way of estimating $UB^y_{z_i|x}$and $LB^y_{z_i|x}$, look at figure 6. In this graph the estimated (unrestricted) lower bound is plotted as the dashed line. From this we can find the lower bound that is consistent with our restriction that $F(y|x,z)$ should never be increasing in z. As can be seen the restricted function has 3 ranges. From Z=$Z_{\min}$ to Z=Z_1, it is flat taking the value of LB(Z_1). From Z=Z_1 to Z=Z_2 the unrestricted bound is consistent with our restriction and so the restricted function and the unrestricted one is identical. From Z=Z_2 to Z=Z_3 the restricted function is again flat taking the value of LB(Z_1).

When will this result in tighter bounds? Using these procedure, we will get a tighter bounds when there exists at least one pair of $z = z_1$, $z = z_2$ ($z_1 < z_2$) for which for the lower bound

$$F(y|I=1,x,Z=z_1)\Pr(I=1|x,Z=z_1) \leq F(y|I=1,x,Z=z_2)\Pr(I=1|x,Z=z_2)$$

Again if we use the respecified bounds as above we get

$$\begin{aligned} & LB^y_{z_2|x} - LB^y_{z_1|x} \\ &= \left[\Pr(I=1|Y<y,x,Z=z_2) - \Pr(I=1|Y<y,x,Z=z_1)\right] F(y|x,Z=z_{1,2}) + \\ & \left[F(y|x,Z=z_2) - F(y|x,Z=z_1)\right] \Pr(I=1|Y<y,x,Z=z_{1,2}) \end{aligned}$$

Rearranging and setting to be positive

$$\begin{aligned} & \frac{\Pr(I=1|Y<y,x,Z=z_2) - \Pr(I=1|Y<y,x,Z=z_1)}{\Pr(I=1|Y<y,x,Z=z_{1,2})} \\ > \quad & \frac{F(y|x,Z=z_1) - F(y|x,Z=z_2)}{F(y|x,Z=z_{1,2})} \end{aligned} \tag{5}$$

Equation 5 suggests that we will get a higher estimate of the lower bound to the wage distribution when the relative difference in the probability that someone with low wages is observed in work across z is large enough to compensate for the fact that there are fewer of them. This finding is very similar to that of Manski and Pepper (2000) namely the closer we get to the exclusion restriction case, the more bite we find in employing an MIV.

For the upper bound we have

$$UB^y_{z_2|x} - UB^y_{z_1|x} =$$

$$[\Pr(I=1|Y<y,x,Z=z_2) - \Pr(I=1|Y<y,x,Z=z_1)]\, F(y|x,Z=z_{1,2}) +$$
$$[F(y|x,Z=z_2) - F(y|x,Z=z_1)]\, \Pr(I=1|Y<y,x,Z=z_{1,2})$$
$$-\Pr(I=1|x,Z=z_2) + \Pr(I=1|x,z=z_1)$$

Rearranging and setting to be positive

$$\begin{aligned} & \frac{\Pr(I=1|Y<y,x,Z=z_2) - \Pr(I=1|Y<y,x,Z=z_1)}{\Pr(I=1|Y<y,x,Z=z_{1,2})} - \\ & \left(\frac{\Pr(I=1|x,Z=z_2) - \Pr(I=1|x,z=z_1)}{\Pr(I=1|Y<y,x,Z=z_{1,2})F(y|x,Z=z_{1,2})}\right) \\ > \quad & \frac{F(y|x,Z=z_1) - F(y|x,Z=z_2)}{F(y|x,Z=z_{1,2})} \end{aligned} \tag{6}$$

The conditions are thus slightly different. Here for the same change in $\Pr(I=1|Y<y,x,z)$ and $F(y|x,z)$ across z we are less likely, than with the lower bound, to be able to tighten the bounds when the change in the probability of work across z ($\frac{\delta \Pr(I=1|x,z)}{\delta z}$) is also positive. However, when the relationship between z and the probability of participation is non monotonic the first term in the above equation could well be smaller than the second ($F(y|x,z)<1$) and so the *sum* could be positive. This possibility, that the relationship between participation and z could be non monotonic, is very real for some of our potential "instruments" (husbands income, leisure etc.). However, the key thing is still that these should be a better determinant of participation than wages.

4 A model for wages that justifies our approach

Our results are best interpreted within a specific type of human capital model. We assume a competitive labour market. The production sector employs a variety of human capital types (s=1,...,S) and the production function for a firm j takes the form.

$$Y_{jt} = F_t(H_{jt}^1, H_{jt}^2, ..., H_{jt}^S, M_{jt})$$

where H_{jt}^s is the quantity of human capital s employed by the firm and M_{jt} is a vector of other inputs. Demand for human capital of each type by each firm is determined by the condition

$$e^{r_t^s} = \frac{\partial F_t(H_{jt}^1, H_{jt}^2, ..., H_{jt}^S, M_{jt})}{\partial H_{jt}^s}$$

where $e^{r_t^s}$ is the price of a unit of human capital of type s which equalises aggregate supply and aggregate demand for that type.

We assume that each individual can supply one type of human capital only. This is the result of earlier investment decisions that are irreversible. Given the type of Human capital that each person possesses he/she has a potential wage if not working and an actual wage if she does. There are no search frictions.

Human capital of each type is produced by a production function of the form

$$H_{it}^s = e^{H^s(ed_i, X_{it})} e^{u_{it}}$$

where s denotes the type of human capital that the individual has associated himself with.

We assume this is observable and in the empirical section we associate this with education levels, although this could represent occupation or profession. The important assumption is that the unobservable u_{it} is separable from education and is assumed independent of education. Implicitly we assume that the choice of education has been made based on unobserved costs that do not have a direct impact on wages.

In this world there is a different price for each type of human capital $e^{r_t^s}$. In this case observed log wages can be written as

$$\ln w_{it} = \sum_{s=1}^{S} 1(s_i = s)\left[r_t^s + H^s(ed_i, X_{it})\right] + u_{it} \tag{7}$$

This formulation allows for changes in the returns to the type of human capital s since each sector has its own time effect. More generally, the returns to any characteristic that defines the type of human capital will change over time. Our approach is consistent with this wage determination model since the unobservables are additively separable in (7). Note that the human capital/ability composition of those unemployed will depend on the vector of prices r_t^s: Even if one were willing to accept that the least productive in a sector are out of work (assuming identical reservation wages), this does not hold across sectors necessarily.

Within this framework one might add a participation model, which helps in imposing restrictions on the relationship between the distribution of wages for those working and the unconditional "potential" distribution of wages. In interpreting our results we suppose that workers make a utility comparison between working and not working taking into account the utility costs of working and the benefits when unemployed. An example of a simple labour force participation rule is

$$U^1(w_{it}, X_{it}, ed_i) - U^0(B_{it}, X_{it}, ed_i) > \varepsilon_{it},$$

where ε_{it} represents unobserved preferences for leisure and where U^1 and U^0 are indirect utilities when in work and out of work respectively. This is the Heckman (1974) model of participation and wages. Who is selected into work depends on the way that u_{it} and ε_{it} are correlated as well as on observables not entering the wage equation. This will imply that we cannot in general know who selects out of work without imposing restrictions on how unobservables preferences and wages relate.

4.1 Non-separable model

A more general model than the one above arises when we assume that the unobservables are not separable from the observables. An example is where the returns to education are heterogeneous. The wage equation becomes

$$lnw_{it} = \sum_{s=1}^{S} 1(s_i = s)\left[r_t^s + H^s(ed_i, X_{it}, v_{it})\right] + u_{it}$$

where v_{it} represents a vector of unobservables. In this case the returns to characteristics X_{it} or education are no longer deterministic functions. Here again the notion of a potential wage for an individual out of work is still valid. However returns to education (or impacts of other characteristics) are heterogeneous. The notion of a single return is no longer well defined. Moreover it will in general be impossible to bound average returns, unless one restricts the support of such returns. In this case the consistent strategy is to bound the distribution of returns to education which we address in our companion paper (Blundell, Gosling, Ichimura and Meghir, 2001)

Finally, a further generalisation can be obtained if we allow education to be correlated with the unobservables. This leads to a further selection problem, over and above the one induced by selection into the labour market and can be dealt with using bounds as well. However, how informative such bounds are likely to be remains a topic for investigation.

5 Empirical Results

5.1 Data source and variable definitions

The data we used for the analysis was pooled the Family Expenditure Survey (FES) cross sections from 1978 to the first quarter 2000. We included in our sample all men and women between the ages of 23 and 59 who were not in full time education[6] This gave us a sample of 187,467 individuals in total. We defined individuals to be in "work" (i.e. $I = 1$) if they were either employed, whether full time or part time or were self employed over the last week[7]. Hourly wages were defined as usual weekly earnings divided by usual weekly hours (inclusive of overtime) and were deflated by the consumer all items quarterly retail price index . Deflated wages lower than 50p an hour and the wages of the self employed were treated as missing.

[6] In the UK university education is completed for most students by the age of 22

[7] We treated the self-employed as workers because excluding them from the analysis would result in another source of selection bias. We believe that the assumption that self-employment is just another form of work, albeit one with a different tax treatment is an acceptable one. We do however treat the wages of the self employed as missing at random

As mentioned above, the proportion of individuals in work, $P(x)$ will determine which quantiles can be estimated. Upper bound quantiles can only be identified up to $P(x)$ and lower bound quantiles can only be identified down to $1 - P(x)$. For women, during most of our time period female participation was below 60% and so in practice we can really only estimate the bounds around the median.

6 Results

6.1 Overall trends in male wage inequality

We start by considering whether the oft cited conclusion that wage inequality has risen since the late 1970s is robust to compositional or selection effects. Figure 6 thus plots the upper and lower bound to the 75^{th}-25^{th} percentile differential from 1978 to 2000 for the male wage distribution. Since the participation rate for women is often quite low we cannot produce the same figure for them

The central line shows, for comparison, what has happened to wage inequality amongst workers and the dotted lines give 95% confidence intervals for the upper and lower bounds (see below). We can only say for certain that inequality has gone up if the lower bound at the end of the period is higher than the higher bound at the beginning of the period. Figure 1 thus shows strong evidence of an increase in inequality. The lower bound in 1998 is higher than the highest bound in 1978, suggesting that inequality must have risen and a comparison of the limits of the bootstrapped confidence intervals suggest that this change is significant implying that these bounds are quite precisely estimated and that there is less than a 5%

chance that inequality was the same in 1999 as in 1978.[8]

The bounds above were estimated by placing no restrictions on the data. We have shown earlier that imposing restrictions on the relationship of the wage distribution for workers and non-workers can lead to tighter bounds. Thus we now impose the restriction that the median wage of workers is at least as high as the median wage for those out of work at any point in time and for any group defined by x. We then follow this up by deriving bounds assuming stochastic dominance, i.e. that all quantiles of the distribution of wages for non-workers are lower than those of workers. As described earlier panel data suggested these assumptions were reasonable overall.

The first panel of figure 7 is simply the earlier bounds shown in figure 6 . The next panel shows the bounds that result from imposing the median restriction; the third panel shows the bounds obtained by exploiting the assumption of stochastic dominance. The 95% confidence intervals for these bounds are reported in Figure 8. The results are particularly strong when we impose stochastic dominance. In this case we can establish that inequality rose *vis a vis* 1978 in all periods after 1986. Taken together, these results give strong evidence for an increase in overall inequality between 1978 and 1999, and that the observed changes cannot be just an artifact of selection.

[8]see the appendix below for the description of the bootsrapping procedure

6.2 The determination of median wages across education, time, age and gender.

6.2.1 Methodology and estimation procedure

To estimate the bounds for the qth quantile of the wage distribution, conditional on x we need to estimate the probability of work, conditional on x $(P(x))$ and then use this to estimate the upper and lower bounds using the formula below:

$$F^{-1}(\frac{q-1+P(x)}{P(x)}|I=1,x) \leq F^{-1}(q|x) \leq F^{-1}(\frac{q}{P(x)}|I=1,x) \tag{8}$$

Generally, the main difficulty, if one is to remain non-parametric[9], is the dimensionality of x. In our case x includes, gender, age, year and education and the principal concern is the cell size, particularly for higher education groups among older individuals.

We define data by age, year, education and gender. To avoid problems with small cell sizes we use seven age groups (23-27,28-32,53.-57).[10]. We define three education groups based on the age individuals reported leaving full time education (less than 17, 17 or 18, above 18). Finally We use ten year groups (78-79..98-99). The probability of work $P(x)$ was estimated as the (weighted) proportion in work for each cell. We then used these probabilities to estimate the upper and lower bounds for the median for each group.

[9]Note that even if the underlying relationship between the quantiles of wages and x were linear, there is nothing to say that the bounds should have the same functional form, thus we need to be as nonparametric as possible in our estmation procedure

[10]As these age groups are quite broad, we assigned a weight for each age group to reach a maximum for workers in the middle of the age category (25,30...). Effectively, then, this procedure is equivalent to a non-parametric regression with a bandwidth of 3 and triangular non overlapping kernels..

6.2.2 Wage growth over the life cycle and across cohorts

Figures 9 and 10 show bounds for the median wage by cohort, and education for three age groups for men and women respectively. For each cohort, age and education group we show the observed median wage (the middle line in the group of three) its upper bound and the lower bound.

For most cases these graphs imply that there has been wage growth across cohorts in all education and age groups. However there are exceptions among the unskilled. For those with the lowest level of education the observed median wage for 25 year olds has hardly grown and the bounds suggest that wages may have declined (or increased perhaps). For the older individuals in this education group we can hardly say anything. This is because there has been a very rapid decline in the labour force participation of older men. For women we can only confirm wage growth across cohorts for the youngest age group of the highest education level.

These graphs can also be used to investigate wage growth over the lifecycle, at least for some cohorts which we observe for a long enough time period. Comparing the upper bound of median wages of a younger age group for a particular birth cohort to the lower bound of an older group we obtain the lowest possible life-cycle growth of median wages for the underlying distribution of wages. Comparing the observed median of the younger age group to the lower bound of the median of the older age group we obtain the lower bound of life-cycle wage growth which is consistent with the median restriction we have imposed.

Consider first the wage distribution for the highest education group. For both men and women there is clear and unambiguous evidence of wage growth between the ages of 25 and

40.For men this is at least 0.5 log point (approximately 50%) while for women the wage growth seems to be considerably lower. For men we can also confirm that wages continue to grow up to the age of 55 (although much lower. This is an important result because it demonstrates that the positive wage growth for middle-aged men is not an artifact of selection. Moreover this results re-confirms that the decline of wages often observed in cross section data past 40 is just an artifact of cohort effects. Here controlling for both selection and cohort we find unambiguous wage growth. For the middle education group we can confirm that there is substantial wage growth form men up to the age of 40. Beyond that we can again confirm that wages do not decline. However the extent of growth is unclear - it could be as high as 40% and as low as zero. For the lowest education group (again for men) the lower bound to lifecycle wage growth is about 10% and when we impose the median restriction it is about 20/% (or a bit higher in both cases for the older cohorts where this is observed). However when we look at wage growth beyond 40 (for the older cohorts) there is no clear picture emerging. Neither can we say much about life-cycle wage growth for the two lower education groups for women. The precision of all the bounds can be assessed by looking at figure 11 where we report 95% confidence intervals for the bounds to the median for men.

6.2.3 Returns to education

We now consider the returns to education and the extent to which they have changed. We compare wages of the highest ($ed = 1$) to the lowest ($ed = 0$) education group for this purpose. When we examine differentials across groups we have to allow for the fact that workers may be positively selected in one group and negatively selected in another. Thus the bounds to wage differentials (D) are set as :

$$w^q_{low}(ed = 1) - w^q_{up}(ed = 0) \leq D \leq w^q_{up}(ed = 1) - w^q_{low}(ed = 0)$$

where $w^q_{low(up)}(x)$ is the lower(upper) bound of the q^{th} quantile of wages conditional on x.

Figure 12 shows the bounds for the changes in the education differentials for men. The first panel looks at changes in the return to education over the entire period, the next looks at changes over the 1980s and the last looks at changes over the 90s. For men in their 30s, returns to education have unambiguously risen over the last 20 years for workers in their 30s. This conclusion is robust to possible selection effects into work[11].

Looking at the next two panels, it is clear that this increase really occurred in the 1980s. Indeed results for each year suggest that education returns are falling in the first half of the 90s and are now rising again.

As Figure 13 shows, for women the bounds to the changes are far too wide to draw any conclusions at all. When the median restriction is imposed for men (Figure 14), the picture becomes much clearer showing that differentials increased for all but older men over the 1980s. Figure 15 reports the confidence intervals for the bounds. These suggest that these changes are not statistically significant.

Finally note that these differentials at the median only really make sense if the observables and the unobservables are separable. In this case we can further tighten the bounds by using the independence assumptions, which we do in the section below.

[11] Of course we have not dealt with here about the problem of endogenous acquisition of education.

6.2.4 Gender wage differentials

Generally the worst case bounds for women are too wide to draw any firm conclusions. In this section we investigate the male/female wage differentials using the results that impose the median restriction.

For low education workers, the observed wage differentials (see figure 16) have tended to decline or remain constant for all age groups. They have also been quite cyclical. However, even with this restriction the bounds are too wide to be sure of what has happened to the underlying wage distribution.

For more educated workers (figure 17) the bounds are tighter. From this we can conclude that wage differentials have fallen for younger higher educated workers. In the first part of the period they were between 5% and 10%. In 1998 they fell to between -5% and 5% (basically zero). Note that the gaps in the graphs occur when there is insufficient information to bound the median wage for women. It is interesting to note though that the lower bound is usually up against the observed differential, implying that if anything selection would imply a higher than observed differential for unskilled women. In other words the changes in participation if anything bias downward the differential, implying that women are catching up with men. However, this may be an artifact of selection.

6.3 Within Group inequality

A feature of the increase in inequality in Britain (as well as the US) has been the large increase in within group inequality. Looking at the point estimates, as Figures 18 and 19 show, we can confirm that this increase cannot have been driven by selection. Looking first at figure 18, which documents the bounds to the changes in inequality (the 75th-25th differential) for those men leaving school at or before 16, inequality has risen for some if not all age groups. The first panel shows that even when we place no restrictions on the data, inequality amongst younger ages has risen, the next two panels suggest that if we make the assumption that those in work have higher potential wages than those out of work, inequality must have risen for all age groups. Figure 19 shows the changes for the higher education group (those leaving school after 18). Note first that the bounds to the changes are much sharper as employment rates are much higher for this group and the observed changes have been relatively smaller. For this group, there is relatively little qualitative gain in making either the median restriction or assuming stochastic dominance, all panels show inequality rising amongst younger cohorts, little changes amongst those in their 40s and rising for older workers.

The picture of an unambiguous increase in within group inequality becomes less clear when we look at the bootstrapped confidence intervals (shown in figure 20). The point estimates suggest that the observed increase cannot be driven by selection but the bootstrapped confidence intervals show that any changes are not significant at the 5% level. For 30 year olds, they are, however, significant at the 20% level for the lower educated group and at 6% for the higher educated group. We are thus relatively confident that our data do show increases in within group inequality.

6.4 Independence

The final set of results is obtained by imposing the assumption that the unobservables in the wage equation are separable from the observables and that they are independent of education. Implementation of this involves estimating the education wage differentials for each quantile and then setting the lower bound as the maximum of the lower bounds and the upper bounds as the minimum of the upper bounds. Note that this restriction can (and is) rejected by the data as it is possible for the maximum of the lower bounds to be greater than the minimum of the lower bounds. The next graph (Figure 21) shows educational differentials for men imposing independence assumption but with no other restrictions.

For 25 year olds the independence restriction is not rejected and, as expected, in some cases the bounds are very tight, even providing a point estimate. This graph also shows that differentials increased in the 1980s similar to the results we reported earlier; the picture for the 90s however is much more ambiguous with no definite confirmation that the differentials have actually gone up.

For 40 year olds the bounds do cross at one year - this may not be significant [see above]. This result, taken at face value implies that differentials did not increase for 40 year olds. Finally for 55 year olds the picture is less clear: There is a rejection of the restriction in the first year of data. After that the bounds start widening out, reflecting the increase in early retirement; this does not allow us to conclude much on the way the differential changed.

These results have exploited an assumption that education is independent of the unobservables. We could combine this with the median restriction. However, there are indications that the assumption may not be valid as in the results we report the bounds cross in two instances. Moreover it is completely clear that we cannot impose independence in any other dimension (cohort or age). For example, the lowest upper bound for the 25 year olds is clearly below the highest lower bound. Hence this confirms that the distribution of unobservables has been changing over time. Similarly we can reject independence across ages.

In the absence of independence there is a distribution of differentials, an issue which is not addressed here. In our companion paper we address directly the issue of bounding the distribution of differentials

6.5 Exploiting information from a determinant of participation

6.5.1 An exclusion restriction

To illustrate how the exclusion restriction might work, we take wage and employment data on low educated women and assume that partners earnings are independent of wages. Partner's earnings are a good proxy for out of work income and should therefore be correlated highly with reservation wages.

Figure 23 plots the relationship between employment and partner's earnings for 25 year old women who left school at 16 in 1980. The first two points denote the employment proportions amongst those with no partner and those whose partner does not work. The next five connected points denote, for each quintile of the partner's earnings distribution, the mean earnings and the proportion of women in work. As expected there is a negative relationship so that those women living with high earning men are less likely to work.

The next figure (Figure 24) plots the worst case upper and lower bounds for those women married to working men across quintiles of the partners earnings distribution. For the group as a whole log median wages are estimated to lie between about 1.2 and 1.6. The horizontal lines drawn across the graph are the points where the lower bound reaches a maximum and the upper bound reaches a minimum. Here the data do not reject the exclusion restriction and it can be seen that assuming that women's earnings are not related in any way to partner's income allows one to obtain much tighter bounds. Estimated log median earnings now lie between 1.45 and 1.4.

This is quite encouraging, but the Figure 25 plots the bounds with the median restriction (that assumes the median wage of workers is not below the median wage of those out of work) with less positive results. Here the upper bound with an exclusion restriction lies below the lower bound with an exclusion restriction. This suggests that the data reject either the median restriction or our exclusion restriction. The next graph shown in figure 26) plots the worst case bound for 25 year olds in 1990 show the exclusion restriction to be rejected for this group. Without making any other assumption, median wages of those women living with partners earning £150 a week are shown to be less than median wages amongst women married to higher earning men. It is probably safe to conclude, therefore, that the best interpretation of figure 25 is that the exclusion restriction is rejected.

The fact that this exclusion restriction is rejected should come as no surprise. Many economic models of household level decision making and of partnership formation would predict a correlation between productivity and partners earnings. The above empirical example above is therefore best viewed as demonstrating the power of the bound approach to reject economic restrictions.

Variables that determine participation that are not related to wages are notoriously difficult to find in general, particularly as any determinant of participation will effect future wages through experience effects. in this case it is even harder. First the assumption of conditional independence is a lot stronger than that imposed by many other selection models. These typically only need to assume that the *mean* of wages does not vary across z, conditional on x. Second many possible instruments such as the variations in welfare payment used by Blundell et al. (2000) are only plausible instruments once a whole set of variables proxying household composition are controlled for. Our non-parametric approach limits the amount of possible control variables we can use and so makes instruments like this impractical. We thus next discuss whether it is possible to use a weaker assumption to gain tighter estimates of the upper and lower bounds to wages

6.5.2 A monotonicity restriction

We now illustrate this restriction by looking at the wages of 30 year old low educated women over the late 80s and 90s[12]. The dotted lines in figure 27 denote the worst case bounds. As participation for this group is so low, these bounds are incredibly wide. The solid line in this figure are those obtained from imposing a monotonicity restriction. It can be seem that there is a significant gain for this group in doing so. The bounds are now over 20% narrower. As the discussion in the theory section predicted, the drop in the upper bound (derived from the lower bound to the distribution function) is larger than increase in the lower bound. The next figure shows the gain when both the median restriction and the monotonicity restriction are imposed. As can be seen, the median restriction is much more powerful than the monotonicity restriction but imposing them both together results in much narrower bounds, particularly at the start of the period.

[12] Between 1981 and 1983 participation for this group was under 50% as so the median was censored

Although there is some potential for obtaining narrower bounds using the monotonicity restriction, the bounds for women are in fact still too large to answer questions about what has happened to the gender wage gap or the return to eduction for women. Future work will utilise variables such as variation in simulated benefit income that come much closer to an exclusion restriction

7 Conclusions

In this paper we have asked a very simple question: What can be learned about the evolution of economic opportunity (as reflected by the wage) without making strong behavioural assumptions or relying on any narrowly defined economic model. In particular we focus on the question of whether observed changes could be an artifact of changes in composition induced by the momentous changes in male and female participation over the last 25 years or so. We show in fact that in some cases we can get quite definite conclusions on how things have changed. However, we also show that many conclusions taken for granted (e.g. male/female differentials have declined) are not robust and critically depend on what one assumes about the employment model one has in mind.

We have explored ways to tighten Manski's worst case bounds and implemented the methods using the UK Family Expenditure Survey to explore the changes in wage distribution accounting for the compositional effect explicitly. We showed that while the worst case bound is sometimes useful, for example in examining the wage growth over cohorts, that alone in many other cases is not sufficient to provide definite results. We proposed and examined the effect of imposing the restriction that wages of workers at the median are higher than the potential wage of non-workers at the median and demonstrated that the condition is informative as described above.

In this version of the paper we have not presented results for bounding the distribution of returns to education when errors are not separable. Neither have we presented such bounding methods for the case where the unobservables (whether separable or not) are not independent of education.

References

[1] Blundell R, Gosling A, Ichimura H and Meghir C (2001) "Composition Effects on the changing structure of Male and female wage distributions", UCL mimeo

[2] Gosling A S Machin and C Meghir (1994) "What has happened to the wage of men since the mid 1960s?" *Fiscal Studies* November pp 63-87

[3] Gosling A S Machin and C Meghir (2000) The Changing Distribution of Male Wages in the UK, *Review of Economic Studies* Oct 2000, Vol.67(4), No.233, pp.635- 666

[4] Heckman J (1974) "Shadow prices, market wages and labour supply" *Econometrica* 42 pp 679-694

[5] Heckman J LaLonde R and Smith J (1999) "The economics and econometrics of active labour market programmes" in O Ashenfelter and D Card eds *Handbook of Labour Economics* Vol 3a North Holland

[6] Heckman J and Honoré B (1990) "The empirical content of the Roy model" *Econometrica* 58(5) pp 1121-1149

[7] Heckman J and Sedlacek G (1985) "Heterogeneity, aggregation and market wage functions; an empirical model of self selection into the labour market" *Journal of Political Economy* 98 (6) pp 1077-1125

[8] Juhn C, Murphy K and Pierce B (1993) "wage inequality and the rise in returns to skill" *Journal of Political Economy* 108 pp 410-42

[9] Koenker, Roger and Gilbert Bassett Jr. (1978) "Regression Quantiles", *Econometrica,* 46, 33-50.

[10] Manski C (1994) "The Selection Problem," in *Advances In Econometrics, Sixth World Congress* Vol 1., Editor C. Sims, Cambridge University Press.

[11] Manski C and J Pepper (1999) "Monotone Instrumental Variables: with Application to the Returns to Schooling," working paper

[12] Schmitt J (1994) "The changing structure of male earnings in Britain 1974-1988" in R Freeman and L Katz (eds) *Differences and Changes in Wage Structures* Chicago, University of Chicago Press

A Using the Bootstrap to obtain confidence intervals for the inequality bounds

There are different ways to construct a confidence interval. What we need is an interval I_α such that

$$\Pr\left\{\left[w_{low}^q, w_{up}^q\right] \in I_\alpha\right\} \leq \alpha$$

but there are many such intervals. There are at least three different criteria we could use: Equal distance interval, equal probability interval, and the shortest distance interval. In the standard normal theory with a point estimate, these three concepts coincide but in our case they do not. In particular, the asymptotic variances of the two bounds are different and hence the equal distance approach produces wide interval with much smaller coverage probability than the nominal probability α. Below we construct equal probability intervals.

To construct our interval we repeated our estimation procedure on 200 bootstrapped samples of the data[13] We then defined the 95% confidence interval as that which we obtained after the following steps:

- Define $w_{up}^q[200]$ as the maximum of the upper bound and $w_{low}^q[1]$ as the minimum of the lower bound in a given cell across replications
- Take the range defined by the pair $w_{up}^q[199]$ and $w_{low}^q[2]$ and find the proportion of replications in which both the upper and lower bounds lie within it
- If this proportion is greater than 0.95 then go on to the range defined by the pair

[13] In fact when we bootstrapped the results for overall inequality there were 1000 replications but the estimated confidence intervals did not differ that much when the smaller number of replications was considered.

$w_{up}^{q}[198]$ and $w_{low}^{q}[3]$, repeat until the proportion goes under 0.95.

For differentials which are obtained from values which are not independent (say a within group differential) we have to conduct this analysis on the differential rather than the value of each quantile

This procedure obtains tighter estimates than if we started at the point estimates and then added (subtracted) a fixed number from the upper (lower) bounds. This is because the precision with which the upper bound is estimated is not the same as that with which the lower bound is estimated, particularly for quantiles other than the median.