

NBER WORKING PAPER SERIES

WAGES AROUND THE WORLD:  
PAY ACROSS OCCUPATIONS AND COUNTRIES

Richard B. Freeman  
Remco H. Oostendorp

Working Paper 8058  
<http://www.nber.org/papers/w8058>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
December 2000

The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

© 2000 by Richard B. Freeman and Remco H. Oostendorp. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Wages Around the World: Pay Across Occupations and Countries  
Richard B. Freeman and Remco H. Oostendorp  
NBER Working Paper No. 8058  
December 2000

### **ABSTRACT**

This study transforms the “October Inquiry” Survey of wages conducted by the International Labour Organization into a consistent data file on pay in 161 occupations in over 150 countries from 1983 to 1998 to examine the pattern of pay across occupations and countries. The new file tells us that:

1. Skill differentials vary inversely with gross domestic product per capita. During the 1980s-1990s, they fell modestly in advanced countries; fell more sharply in upper middle income countries while rising markedly in countries moving from communism to free markets and in lower middle income countries.

2. Wages in the same occupation vary greatly across countries measured by common currency exchange rates and measured by purchasing power parity. Cross-country differences in pay for comparable work increased, despite increased world trade.

3. The principal forces that affect the occupational wage structure around the world are the level of gross domestic product per capita and unionisation/wage-setting institutions.

Richard B. Freeman  
Harvard University  
1050 Mass. Ave. 3rd fl  
Cambridge, MA 02138  
NBER and LSE  
freeman@nber.org

Remco Oostendorp  
ESI, Vrije Universiteit de Boelelaan 1105  
Kamer 4a-23  
1081 HV Amsterdam, Netherlands  
roostendorp@econ.vu.nl

## Wages Around the World: pay across occupations and countries

Consider the wages of workers in two occupations in two countries: say a labourer and software engineer in India and a labourer and software engineer in Germany. There are three useful contrasts one can make with these wages.

1. The percentage (or absolute) differential between the wages of more skilled and less skilled workers within each country. All else the same, this **skill differential** reflects the return to human capital and the degree of inequality in the wage structure in the countries.

2. The ratio of the wages of workers in the same occupation between the countries, measured in a common currency through exchange rates. All else the same, this **cost differential** reflects the different cost of producing goods with these workers.

3. The difference in wages for workers in the same occupation, measured in purchasing power parity units. All else the same, these **living standard differentials** measure differences in the economic well being of similarly skilled workers.

With data for many occupations and countries, one can examine generalizations of these relations: the variance of occupational earnings within countries; the dispersion of cost differentials among countries; and inequality in living standards for similarly skilled workers around the world. With data over time, one can further examine changes in these measures and the tendencies, if any, toward equalization of pay across occupations and countries.

Labour economists have traditionally focused on the pattern of skill differentials.<sup>1</sup> Their chief generalization is that differentials are lower in advanced countries and decline over time with economic development, because of more plentiful supplies of skilled workers.<sup>2</sup> In Tinbergen's "race" between education-induced increases in the supply of skills and technology-induced increased in demand for skills, the shift in supply has dominated, reducing differentials.<sup>3</sup>

Trade economists focus on differentials in the cost of comparable labour across countries. The major issue is the extent to which trade produces factor price equalization. In its strongest form, the Samuelson-Stolper theorem predicts the same pay for workers with the same skills around the world. Still, trade theory allows the wages of similarly skilled workers to differ if countries operate in different cones of diversity or use different technologies.

Development economists focus on how economic growth affects living standards and thus on wages deflated by purchasing power parity price indices. They worry whether growth raises the living standards of the poor in the

early phases of development; whether an effectively infinite supply of labour from the traditional sector keeps real wages down; or if urban-bias in wage-setting creates excessive rural-urban migration and unemployment.

In short, wage differentials around the world are central to various economic questions and theories about the operation of the labour market. Despite this, there is no generally accepted international data on occupational wages. The most widely used cross-country data on wages are the International Labour Organization's annual Yearbook of Labour Statistics manufacturing wages which does not differentiate occupations or skills. Studies of occupational wage structures treat a single country or a small number of countries or the wages of workers in specific occupations across countries. Psacharapoulous's data on earnings by education covers many countries but relies on many different sources and is of dubious quality or comparability.<sup>4</sup>

This study examines wages in 161 occupations around the world using the new **Occupational Wages around the World** (OWW) file. This is a data file that we developed by standardizing wages from the most far-ranging survey of wages, ILO's October Inquiry Survey, into a consistent series of pay across occupations, countries, and time. Using the OWW to examine skill differentials, cost differences, and living standard differentials from 1983 to 1998, we find:

1. Skill differentials vary inversely with gross domestic product per capita. During the 1980s-1990s, they fell modestly in advanced countries; fell more sharply in upper middle income countries; and rose markedly in countries moving from communism to free markets and in lower middle income countries.

2. Wages in the same occupation vary greatly across countries measured by common currency exchange rates and measured by purchasing power parity. Cross-country differences in pay for comparable work increased, despite increased world trade.

3. The principal forces that affect the occupational wage structure around the world are the level of gross domestic product per capita and unionisation/wage-setting institutions.

## **Standardizing the ILO October Survey**

Since 1924 the ILO has conducted an "October Inquiry" to obtain data on pay by occupation across the world. The ILO sends a questionnaire to national governments asking for wages in detailed occupations within particular industries. Table 1 shows the coverage of the October Inquiry by occupation and country over time. In 1924 the survey gathered data on male earners in 18 occupations in 15 countries. In ensuing years the ILO expanded the number of countries and occupations. Country coverage increases fairly steadily so that the 1983-98 Inquiry data files on which we focus had wage statistics for 156 countries in at least one year and wages on up to 76 countries in

any given year. The number of occupations increased from 30 occupations in 1929, to 41 in 1951, to 48 in 1953, and then to 161 in 1983.<sup>5</sup>

To assure comparability of occupational definitions across the countries, the ILO specifies in detail the work involved in each occupation. To get a flavour of the specificity consider the following description of a clicker cutter in the footwear industry:

*Clicker cutter (machine). Operates press machine which cuts out upper parts of footwear; lays material on the table of machine; selects cutting dies; arranges dies on material to cut it economically and avoid weaknesses; cuts out shoe part by lowering press onto dies; removes cut-out parts from material.*

Or this (abbreviated) description of an accountant in a bank:

*Accountant. Plans and administers accounting services and examines, analyses, interprets and evaluates accounting records for the purpose of giving advice on accountancy problems or preparing statements and installing or advising on systems of recording costs or other financial and budgetary data: ... keeps record of all taxes, fees, etc. to be paid by the bank ...conducts financial investigations on suspected fraud ... prepares and certifies financial statements for presentation to the board of directors, executives, shareholders ...*

If each country obtained wages from a nationally representative survey based on ILO definitions, the October Inquiry would be the ideal source for comparing the pay of comparable labour across countries. Clicker cutters, accountants, economists, whomever from Pakistan or Romania or Germany or the US would have essentially interchangeable skills, so that one would truly be comparing equivalent labour.

But the October Inquiry data fall short of ideal. To the contrary, the Inquiry data has so many problems that the survey is one of the least widely used sources of cross-country data in the world. The main problem is that countries respond to the ILO in inconsistent ways. Recorded wages are not directly comparable across countries, in the same country over time, or even among occupations in a country at a point of time, in part because countries report data from differing national sources rather than conducting special surveys to answer the ILO request. Some countries report average earnings in an occupation from an employer survey, others report legislated minimum wages, yet others report collectively bargained wages, while others give hourly wage rates. In 1995 the US reported median weekly earnings from a household survey; Germany reported minimum collectively bargained hourly, daily, weekly, and monthly wage rates, depending on the occupation; India reported minimum daily and monthly wage rates; Nigeria reported average monthly wage rates and average monthly earnings. Some countries give wages for men in some occupations. Others report wages for men and women. Others report wages for women in some occupations. And so on.

Another problem is that countries do not report consistently from year to year. In the 1983-1998 period 156 countries reported wages in at least one year, but just 11 countries reported wages 15 or 16 times (effectively every year given that we have only preliminary data for 1998), 31 reported 10-14 times, 45 reported 5-9 times, and 47 reported 1-4 times and 22 reported just once. Looking across the years, in 1983, 56 countries reported wages, in 1985, 70 reported wages, in 1990, 73 reported, in 1992, 61 reported, in 1995, 76 reported, in 1997, 48 reported.

A third problem is that the Inquiry fails to obtain information on all supplementary labour costs. The ILO asks that countries include cost of living adjustments and other guaranteed and regularly paid allowances in their reports on wages and asks that countries include paid vacation and holidays as well in earnings. But the Inquiry excludes employer contributions to social security, pension funds, severance pay, irregular bonuses, including such important payments as the annual or bi-annual bonuses paid in Japan and some other Asian countries, and contributions to health insurance. To the extent that social charges are proportionate to wages, this will not affect estimates of relative wage structures in countries, but it will underestimate differences in labour costs or living standards across countries.

A fourth problem is that even with the ILO's detailed specification of skills, the work performed in an occupation will vary across countries. Even in a country, skills differ within the narrow ILO categories. The skill of cooks working in restaurants and hotels (an ILO occupation) in Great Britain range widely, depending on the size of an establishment, cuisine, and number of stars in the guidebooks. Such differences are likely to be even greater across countries. To the extent that differences in skills within occupations are associated with education, workers are likely to be more skilled in advanced countries than in less advanced countries.

Finally, the changing numbers of occupations over time makes comparisons between the post-1983 period and earlier years difficult. In addition, some countries do not in fact provide national data but data from particular regions, such as major cities or urban areas.

Table 2 gives a detailed description of information in the Inquiry files from the 1983 to 1998 period on which we focus. The complicated structure of the data makes it clear why few researchers use these data to analyse wages around the world.

Panel A gives information on the size of the sample. It shows the maximum conceivable number of observations that the Inquiry would contain if each country reported a single wage statistic for each occupation yearly: over 400,000 pieces of data.<sup>6</sup> The actual number of observations is smaller, largely because most countries do not report statistics in many years. On average, countries report wages for 6.3 years out of 16 possible years. Even if we ignore 1998, which has preliminary data for only 14 countries, more than half of country year observations are

empty. In addition, countries do not report data for every occupation in the years when they do report. The bottom line is that there are 74,056 country x year x occupation cells with wage data in the 1983-1998 file.

But there is a further complication. Many countries report more than one wage for a single occupation. Some give hourly wage rates **and** average earnings. Others give wages for men **and** wages for women. Others give wages for one gender and for both genders. Nearly half of the observations (46.3%) contain multiple wage figures. While this will help us calibrate the data to a single form, it makes the raw data difficult to use in cross-country comparisons, particularly since different countries report pay differently. Including multiple wages, we have 124,016 pieces of data.

Panel B shows the frequency distribution of countries by the number of occupations they report; and the frequency distribution of occupations by the number of countries that report statistics on them. The distribution of countries by number of occupations shows that in most countries there are sufficient occupations with wage data to get a good measure of the overall wage structure. It also shows, however, that different countries report on different numbers of occupations, which creates problems in comparing wage structures across countries. The distribution of occupations by country shows that many occupations have wage data for large numbers of countries, which will allow us to contrast labour costs and living standards for workers in the same occupation around the world.

Panel C shows the diverse way in which countries report wages. Most countries report wage rates, presumably from employer surveys or collective bargaining contracts or legislated pay schedules. However, many report earnings, which may come from household surveys. Most give statistics in the form of averages<sup>7</sup> but 23 percent report minimum wages, some from collective bargaining contracts. Some countries report maximum wages. Others give prevailing wages. The US reports median weekly earnings for most occupations (from individual reports on the Current Population Survey). The time period to which the pay refers also varies. The most common period is the month, followed by the hour, but some countries report weekly pay, others give daily rates for some occupations, and so on. There is also variation by gender. Forty-three percent of the observations relate to male workers, 36 percent to all workers, and 21 percent to female workers.

Combining all of these different variants, the vast majority of the Inquiry statistics are non-comparable. Just 5.8 percent relate simultaneously to the most common pay concept (wages), use the most common averaging (mean), cover the most common time span (monthly), and relate to the most common gender (male).<sup>8</sup> How is one to make a valid contrast of pay across country lines when the US reports median weekly earnings for both genders and China reports average monthly wages for females and males separately and Germany gives collectively bargained minimum wages for different time periods. Is it possible to put the Inquiry data into a form that researchers can readily use?

The principal contribution of this paper is to demonstrate that the answer to this question is yes. It is possible to transform the Inquiry data into a usable form by calibrating the diverse statistics into a normalized wage or earnings rate. To see how we standardized the data, consider each observation  $W$  to be the sum of a standard rate  $W^*$ , which we take to be the most common form of data in the Inquiry -- monthly average wages for male workers<sup>9</sup> -- and a data type adjustment term,  $W^a$ , where  $a$  reflects the deviation of the observed wage from the most common type, and an error term,  $v$ :

$$(1) W(i,j,o,t) = W^*(j,o,t) + W^a(i,j,o,t) + v(i,j,o,t)$$

where  $i$  measures the data type;  $j$  refers to the country;  $o$  is the occupation;  $t$  is the time period

The calibration problem is to estimate  $W^*$  for observations where data is reported in non-standard form -- that is, to find adjustment coefficients that measure how non-standard forms of data diverge from  $W^*$  for different countries, occupations, and time periods. Let  $X(i,j,o,t)$  be a row vector of dummies for data type, which takes the value one if the observation is of the particular data type and  $B(i,j,o,t)$  be a column vector of deviations of a particular type of data from the normal. Then we write (1) as:

$$(2) W(i,j,o,t) = W^*(j,o,t) + X(i,j,o,t) B(i,j,o,t) + v(i,j,o,t)$$

The natural way to estimate the  $B$  coefficients is to regress the  $W(i,j,o,t)$  on  $W^*(j,o,t)$  and the dummy variables for observations with both standard and non-standard rates using the observations in which there is data for both the standard form and the non-standard form:

$$(3) W(i,j,o,t)' = W^*(j,o,t)' + X(i,j,o,t)' B(i,j,o,t)' + v(i,j,o,t)'$$

where the observations  $(i,j,o,t)'$  are those for which we have both standard and non-standard wages and where  $v$  is a residual term with  $E(v) = 0$ .

Given the estimated  $B$ s, we can predict  $W^*$  for observations that did not have both types from the following equation:

$$(4) PW^*(j,o,t) = W(i,j,o,t) - X^a(i,j,o,t) PB(i,j,o,t)$$

where  $P$  before a term reflects the predicted or estimated value.

The extent of variation in the Inquiry data complicates this standardization exercise. The big problem is that there are many different types of data. There are 2 types of earnings, 3 types of data by gender, 3 forms of data by time span that cannot be standardized by dimensional analysis;<sup>10</sup> and five distinct forms of averaging. This gives us  $90 (= 2 \times 3 \times 3 \times 5)$  different potential combinations of data types. In addition, there are possible differences in  $B$  coefficients by country, occupation, or time. There is insufficient overlap of observations for all the different possibilities to estimate a full set of  $B$  coefficients.



To cut through this morass, we must simplify the X vector in various ways — for instance, by assuming that different data types affect wages separately rather than interactively (reducing the dimension of X from 90 to 12); or by assuming that B is time invariant or independent of occupation or country. These simplifications put some of the deviation of  $W(i,j,o,t)$  from  $W^*(j,o,t)$  into the error term, introducing heteroskedasticity.

Unfortunately the variation in the Inquiry data is too ‘thin’ to limit the estimation of B to the observations  $(i,j,o,t)$  for which we have both standard and non-standard rates. In principle this would give us 19,010 observations, except that many of the data types occur rarely in conjunction with a standard data type. This happens because in most circumstances these data types are reported in lieu of the normalized data type. Only the wage/earnings and male/female wage differentials can be estimated with any precision in this manner.

For this reason we choose to estimate the normalized wage as well, running the following regression on the entire sample:

$$(5) \quad W(i,j,o,t) = D(j,o,t) A(j,o,t) + X(i,j,o,t) B(i,j,o,t) + \theta(j,o,t) + v(i,j,o,t)$$

where  $D(j,o,t)$  is a row vector of country, occupation and time dummies (including possible interaction terms),  $A(j,o,t)$  a column vector of coefficients, and  $\theta(j,o,t)$  the random component of the normalized wage not captured by the country, occupation and time dummies. The predicted B can be used to predict  $W^*(j,o,t)$  using equation (4).

The most troublesome aspect of our standardization process is that there is no natural single way to simplify the vector of B coefficients. We run the risk that one method of simplification, for instance assuming that the B coefficients in less developed countries are the same as those in advanced countries could yield sufficiently different estimated wages from another simplifying assumption, say that advanced and developing countries have different B coefficients. To take an example, if gender differentials fall by level of development and we adjusted female wages to the male monthly standard using an adjustment parameter estimated for both advanced and developing countries, we would understate male wages in particular occupations in the developing countries and overstate male wages in the advanced countries. The only way to solve this problem is to try several standardization procedures and examine the difference among them in final outcomes.

Even for any given standardization procedure, a multiplicity of predictions or calibrations for the normalized wage will arise if there is more than one wage reported for any given country, occupation and year. If the variance-covariance structure of the error term  $v(i,j,o,t)$  is known, optimal weights can be calculated to derive a weighted prediction with minimal sampling variance. In practice these optimal weights are difficult to derive and different weighting schemes can be used to examine their impact (see Appendix A). Alternatively, equation (5) can be

estimated as a random effects regression model with heteroskedasticity, and a unique prediction of the normalized wage can be derived as

$$(6) PW^*(j,o,t) = D(j,o,t) PA(j,o,t) + P\theta(i,j,o,t)$$

Prima facie, this direct procedure is more efficient, but potentially less robust because it assumes more knowledge about the variance-covariance structure of the regression model.

In Appendix A we show that each of the above calibrations, with alternative simplifications for the coefficient B, different weighting schemes, or different estimation procedures gave similar results, with correlations ranging from 0.9745 to 0.998 for the predicted wages, and from 0.9650 to 0.9987 for the dispersion measures used in this paper.

Using the above standardization or calibration procedure, what can the OWW tell us about the determination of wages and operation of labour markets around the world? By itself, the new file allows us to examine occupational wage structures within countries over time and the cost of comparable labour and living standards of similarly skilled workers across countries. In conjunction with other data sets it allows us to examine the determinants of wages in comparable occupations around the world and the effect of those wages on other economic phenomena.

## Skill differentials

The OWW is a huge country-occupation-time matrix of wages with many missing elements. Elements are missing because countries report on different occupations in different time periods. New Zealand, for instance, reports wage data continuously from 1984 to 1991, but not earlier or afterwards. It gives wages for 84 occupations in 1984, 131 occupations in 1985, on 143 in 1986, and 135 in 1991. Hungary gives wages for 26 occupations in 1987 and wages for 130 or so occupations in 1995, 1996, and 1997, but gives no figures in other years. And so on. To obtain manageable and comprehensible statistics about skill differentials, we must collapse the OWW matrix into consistent summary measures of the spread of pay among occupations. We compress the data in three ways.

First, for every country we calculated the standard deviation of the log of pay and the ratio of the wages in the occupation in the 90<sup>th</sup> percentile of the occupation wage distribution to the wages in the occupation in the 10<sup>th</sup> percentile of the distribution for **all of the occupations** reporting in a given period.<sup>11</sup> This comparison uses the maximum amount of data but it compares different numbers of occupations across countries, so that the measured spread of wages may be affected by the number of occupations that enter the summary statistic. The most natural way to deal with this problem is to compute measures of skill differences for exactly the same occupations for all countries. But countries give wages for different occupations, so that this common denominator strategy would greatly reduce the sample size.

Instead, we have chosen a different way of compressing the data. This is to treat observations as samples from the distribution of occupational wages for each country, rather than as estimates of wages for a specific occupation.<sup>12</sup> We then estimate the decile distribution of wages by occupation and calculate measures of dispersion from this distribution. Specifically, we ordered occupations by their wage in each country time period; divided the ordering into deciles, and took the mean wage in each decile as that decile's wage in the country.<sup>13</sup> This gives us ten wages for each country, from which we to calculate measures of dispersion. For instance, if the top 10% paying occupations in the US consisted of eight occupations, we would use the mean wage among those eight to represent the top decile in the US, whereas if the top 10% in India had wages only for three occupations, we would use those three wages to estimate the top decile in India. It turns out that the measures of dispersion of pay based on all this decile analysis and measures based on all of the data are highly correlated, which implies that the number of occupations problem is not a serious one in our data. Accordingly, most of the results we report are from analysis of the maximum number of occupational data points, with countries that report on 25 or more occupations, rather than from analysis of the derived decile distribution.

Third, to compare changes in wage structures over time, we limit our analysis to **the same occupations in each country**. If Argentina reports 100 occupations in one year and 56 in another, and just 50 of the occupations overlap, we compare the wages of those 50 occupations over time. If Mexico reports consistently on 30 occupations over time, we use wages in those 30 occupations. In this case we opt for comparability over time rather than comparability in numbers of occupations or in the quantile distribution of occupations.

Table 3 presents our measures of the dispersion of wages by occupation for the mid-period in our data, 1988-1992, organized by the level of development of countries. To obtain the maximum number of comparisons over this period, we report the figures for the year in the interval which gave the most data. Columns 1 and 2 give the statistics based on all of the occupations reporting in the peak year, while columns 3 and 4 give the statistics based on our estimated decile earnings in the occupation-earnings distribution. Both calculations show that skill differences or wage inequality is smaller in the more advanced countries, and is particularly small in the communist countries. Since these facts are well-known from more limited country comparisons, this can be viewed as broadly validating the OWW file.

Figure 1 shows the inverse relation between skill differentials and level of economic development in a somewhat different way. It graphs the standard deviation of ln wages by occupation in a country against the level of GDP per capita in the country for the year for which the highest number of occupations were reported. While there is considerable variation in occupational wage inequality for countries with the same level of GDP per capita -- for instance, the US and UK have high levels of inequality for advanced countries -- the scatter diagram shows clearly the inverse relation between income per head and the spread of wages among occupations.

We pursue the link between GDP per capita and the occupational wage structure in table 4 through regression analysis of a pooled cross-section time series file for all of the years and countries in the OWW. Columns 1-3 record the results of regressions that link the occupational wage structures to the level of GDP per capita in a country. Each regression includes dummy variables for the year of the observation so that the estimates reflect solely the cross section relation between the variables. Since we do not have observations for every year for countries, the file is unbalanced. We report robust standard errors to correct for clustering.

Regression 1 confirms the tendency for skill differentials to fall with rising income per capita and for countries with communist regimes to have particularly low levels of skill differentials. The estimated coefficient on income per capita is sizable, implying that much of the difference in the wage structure between more/less advanced countries is attributable to differing levels of income per head. The inclusion of the squared term in regression 2 provides, moreover, some support for the existence of a Kuznets type curve linking inequality to development.

Inequality rises as GDP per capita rises (positive coefficient on GDP per capita), then falls with GDP per capita (negative coefficient on GDP per capita<sup>2</sup>). The turning point occurs at a relatively low level of GDP per capita. The third regression records coefficients based on a median regression (which finds the coefficients that minimize absolute differences between predicted and actual values rather than squared differences). The case for a Kuznets inverse U shaped curve pattern is stronger than in regression 2, with the estimated effects of both GDP per capita and GDP per capita squared on inequality higher, and the level of GDP where increases reduce inequality rises.

## **Trends in skill differentials**

Figure 2 turns to the trend in skill differentials over time. It gives summary statistics on the spread of wages within countries for **fixed numbers of occupations** over time.<sup>14</sup> It covers three periods, 1983-88; 1989-92; and 1993-98. In each of these periods we picked the year that gave us the maximum number of comparable occupations over time. The figure shows little change in the dispersion of wages among advanced countries as a group, though the underlying data shows increased dispersion in the US, Great Britain, and Italy, among others, consistent with the picture in other data sets (Freeman and Katz). Taken individually, the dispersion of occupational wages rises in 8 of the advanced countries for which we have data and falls in 8 countries. The figure also shows a sharp rise in inequality in the transitional countries (Russia, Romania, the Czech Republic and Slovakia)<sup>15</sup>, again reassuring given the evidence of increased inequality in those countries. In the lowest income countries for which we have comparable data, the decline in dispersion shown in the figure is a bit misleading. This is because the most striking pattern in the individual country data is one of divergence in skill differentials. Inequality falls sharply in two economic disaster countries, the Sudan and Myanmar but rises sharply in Egypt and Sri Lanka. Overall, 7 of 12 of the less developed countries have falls in skill differentials while 5 have rises. The coefficient of variation in the standard deviation of ln wages in the least developed countries was 1.10 in 1993-98 compared to 0.88 in 1983-87.

Columns 4-7 in Table 4 pursue this analysis of trends in skill differentials for all of the countries in our sample. Each of these regressions include country dummies, which subsume cross section patterns, and thus focus on changes within a country over time. Regression 4 replaces the year dummies in the earlier regressions with a linear trend to identify more easily the direction of change over time. Looking across all countries, the coefficient on trend in regression 4 is small and lacking statistical significance, implying that there was little change in occupational wage structures in the 1980s-1990s, on a global scale. Regression 5 gives separate trend coefficients for countries in different income classes; it shows that skill differentials fell in advanced countries and more sharply in upper middle

income countries while rising in lower middle income countries. In low income countries, there is a modest but insignificant downward trend in inequality.

That there is no world trend in occupational wage inequality does not mean that increases in GDP per capita do not reduce skill differentials. Regression 6 includes both year and country dummies, so that it identifies the effect of changes in GDP over time on inequality within a country. The coefficient on GDP per capita is negative, though smaller than in the cross section regression of column 1. Since this calculation includes controls for both country and time, it implies that within the same country, greater growth of GDP was associated with reduced inequality of wages. Regression 7 decomposes the GDP per capita term into separate GDP per capita effects for countries with differing levels of income. It shows that, conditional on trend, GDP per capita has its largest impact in reducing skill differentials in upper middle income countries, while having a positive impact in lower middle income countries.

### **Cost of labour differences across countries**

To analyse differences in the cost of skills across countries, we have deflated the wages for each country-year-occupation by exchange rates for the US dollar. The resulting dollar measure shows great variation across countries in the cost of labour. A carpenter in construction in India, for example, earns \$48 per month, while one in Sweden earns \$2,481 per month and one in Argentina earns \$206 per month on average in the 1988-1992 period. Even among the advanced countries, there are considerable differences in labour costs: a kindergarten teacher in Italy earns \$ 1,786 per month ; a kindergarten teacher in the US earns \$1,501 per month; while a teacher in Japan earns \$ 1,268; and one in Germany earns \$ 2,484 per month, on average over the 1988-1992 period.

Figure 3 displays the variation in pay across countries for five occupations in the 1988-92 period: a high-wage occupation (general physician); a low wage occupation (logger); a white collar occupation (insurance agent); a blue collar occupation (clicker cutter); and a high tech occupation (computer programmer). For ease of presentation, we have standardized the wages relative to the highest wage in the category. That is, the wage in the highest wage country in a given occupation is scaled as 1.0 and the wages of workers in that occupation in other countries are fractions between 0 and 1. The frequency distribution shows the number of countries with given wages relative to the highest paying country in each occupation. As a summary of the variation in each occupation, we record the ratio of the wage in the country at the median of the distribution relative to the wage in the country with the highest wage. This is not a common measure of dispersion, but it is an easily interpreted statistic:<sup>16</sup> the smaller the median wage relative to the maximum wage, the further from the maximum wage are occupational wages in other countries. The median wage/maximum wage for all five occupations is relatively low -- it ranges from .050 to .159 -- implying huge

differences in the cost of nominally comparable labour across the world. We also report the standard deviation of ln pay for each occupation across the reporting countries. The standard deviation is also large, compared, say, to the standard deviation of ln wages among occupations within a country.

Figure 4A extends this analysis to all 161 occupations. For this figure, we calculated the ratio of the wage in the median wage country in each occupation divided by the wage in the highest wage country and the standard deviation of ln wages across countries in each occupation. We then took the average of these statistics for all the occupations for which there are at least 25 countries reporting (145 occupations). The figure confirms the evidence in figure 3 that there are huge differentials in the cost of nominally comparable labour across countries: the average median/max wage is just 0.10 while the standard deviation of the ln wage is 1.21.

The bottom parts of figure 4A present comparable statistics for countries with different income levels and for occupations disaggregated by skill level. The statistics for the wage in the median paying country/maximum wage shows that advanced countries have occupational wages reasonably close to the highest wage (which of course is in an advanced country, though usually a different one). The average median/maximum wage for less developed countries pay show a much wider spread from the maximum wage. In addition, these data show that higher skilled occupations have modestly less spread in pay than lower skilled occupations.

The standard deviation of ln wages in the figure give the standard deviation in the ln wage around the average **within each country grouping**, rather than around the average for all countries. These standard deviations thus show the variation in wages within income classes. Within groupings, the advanced countries have less cross-country variation in occupational pay than do the less advanced countries and more skilled occupations have less cross country variation than the less skilled occupations.

To what extent, if at all, did the economic developments of the 1980s and 1990s, such as increased globalisation and the spread of technology, reduce differences in wages across occupations?

One way to answer this question is to regress measures of the country dispersion of pay in occupations on trend variables. Table 5 presents estimates of these trend effects, with individual occupation dummy variables included in each regression to eliminate cross-occupation variation in dispersion. The regressions show that variation in wages across countries **rose** over this period using either our median country wage/maximum country wage measure of the spread of pay within an occupation (column 1) or the standard deviation of ln wages in the occupation (column 3). To see if this pattern differs according to the skill or wage level in an occupation, we ranked occupations by their rate of pay in a single country, Germany<sup>17</sup>, giving the lowest paying occupation a 1 and giving the highest paying occupation a 161. Then we interacted the rank with the time trend. The results, given in columns 2 and 4 of

the table, show that for both the median/max pay and standard deviation of ln pay, wages became more dispersed across countries in the higher skilled occupations.

The pattern for rising country differentials of occupational wages in a period of increased globalisation suggests that factors beyond trade pressures greatly affected the dispersion of pay across countries. Determining what those factors were — the debt crisis of the 1980s? the deterioration of African economies? smaller absolute increases in schooling in less developed countries than in more developed countries? brain drains? -- lies beyond the scope of this essay.

### **Comparing living standards**

If exchange rates reflected differences in the cost of goods across countries, wages deflated by exchange rates would suffice to measure living standards for similar workers. But exchange rates invariably overstate the cost of non-tradeable goods and vary wildly on foreign currency markets for reasons having little to do with real economic conditions. Accordingly, to compare standards of living for comparably skilled workers across countries, we deflate observed wages by the purchasing power parity prices in the Penn World Tables.<sup>18</sup> Purchasing power parity prices diverge greatly from exchange rate prices. Virtually all less developed countries have higher purchasing power for their currency than exchange rates show. Summers and Heston show that the ratio of PPP to the exchange rate rises with GDP from below 1 to levels above one, and that the divergence is particularly sharp for non-tradables (Summers and Heston, table 1). The magnitude of the increase in PPP/Exchange rates is sizeable, ranging from 0.38 in low income countries to 1.17 in high income countries in 1988.

The lower cost of living in low income countries implies that the variation in wages across countries will be lower when we deflate wages by purchasing power parity prices than when we deflate by exchange rates. Panel B of figure 4 shows that this is the case. Here we calculate our two measures of the dispersion of wages across countries using wages deflated by PPPs rather than by exchange rates and report the average across occupations of these statistics. With the PPP adjustment, the variation in wages among occupations is reduced considerably compared to what it was with the exchange rate adjustment. The median to maximum wage statistic rises from 0.10 to 0.27 while the standard deviation of ln wage drops by 0.37 ln points, from 1.21 to 0.84. Still, even with wages deflated by PPPs there remains huge cross country variation in the payment to workers in the same occupation.

### **Determinants of wage structures**

What factors beyond GDP per capita determine the spread of occupational wage structures and wages? We have analysed the impact of three potential determinants of occupational wages:



the educational attainment of workers in the economy, which most analyses view as the primary cause for the lower level of differentials in advanced countries;

the extent of institutional determination of wages and/or prices in a country, which most analyses view as reducing pay differentials;

the link to the global economy, which is expected to reduce skill differences in less developed countries and raise them in more developed countries due to the pressures of factor price equalization on countries with differing levels of skill among workers.

Our measures of these factors are crude, though no more so than in studies of cross-country growth rates, since we have simply imported some variables from the cross country growth literature. Still, we view the results as tentative, subject to revision given better measures of the independent variables or more structured models or better specified empirical tests.

We measure **educational attainment** by years of schooling in 1985 from the Barro-Lee data set. For some observations, the attainment figures are based on UNESCO data on average years of schooling; for other observations, they are derived from historical enrollment data using a perpetual inventory method of transforming enrollment flows into educational stock. While the data surely distinguish between countries with very high and very low levels of schooling, they are subject to considerable error (Behrman and Rosensweig), which makes any analysis of changes a problem (Krueger and Lindahl). Ideally, we would want educational levels for occupations within a country for each of our three periods, but the current analysis is limited to cross-country differences using a single years' education for each country.

We have two types of measures of the role of **institutions in wage-setting**. First, we have obtained estimates of union density for a larger number of countries in 1990-95 from Visser (1999); these numbers are also subject to considerable measurement error. But perhaps the biggest problem with them is the union density does not necessarily translate into collective bargaining and wage determination. Some countries, such as the Netherlands or Austria have modest union density but near universal collective bargaining. Other countries, such as the Ukraine or China have high union density, but little or no collective bargaining. Accordingly, we have modified the density data in two ways. For OECD countries, we have replaced unionism with OECD estimates of collective bargaining coverage from OECD (1997). For communist countries like China and for ex-communist countries, where the union figures do not reflect free trade unions, we have coded unionism as zero.<sup>19</sup> Our second type of data are subjective evaluations of the extent of institutional intervention in an economy taken from the Heritage Foundation's Index of Economic Freedom. We have taken two indicators: the rank of country in its economic freedom scale, and the score

given the country in the extent of wage/price interventions in the economy. As the economic freedom scale places market economies, including those that intervene significantly in wage-setting such as the Nordic countries, high in its ranking, it is less appropriate for analysing occupational wage dispersion than the more narrowly defined wage/price intervention scale. We have taken these variables from the 1998 Index of Economic Freedom, because this gives us the maximum number of country coverage.

For **trade** we chose two types of measures. We use the ratio of exports plus imports to GDP from the World Bank's World Development Indicators to indicate the extent to which a country is connected to the global economy. While some trade economists are dubious of quantity measures of the openness to trade, on the grounds that trade can have sizable effects on wages through prices, even with no trade flows, others argue that quantitative measures of trade may still affect wage and other outcomes, not least because it may reflect barriers such as tariff or non-tariff policies.<sup>20</sup> As our second indicator of the potential influence of trade on the structure of wages, we use the Heritage Foundation measure of trade policies from its economic freedom report. To the extent that this subjective measure ignores actual flows of goods and services to focus on presumably exogenous policy factors, it has particular merit for our analysis. Because trade is expected to have a differential effect on skill differentials in advanced and less advanced countries, raising dispersion in high income countries with relatively high levels of labour skill and lowering dispersion in low income countries with relatively low levels of labour skill, we have introduced both of these trade measures interactively with gross domestic product per capita.

Table 6 presents regression coefficients that link the standard deviation of ln occupational wages to these indicators of education, wage-setting institutions, and trade in a pooled cross-section analysis over all the years for which we have country data. The number of observations in the various regressions differs somewhat due to the availability of measures for particular countries. Since the explanatory variables are from single years, we include year dummies in each regression and report robust standard errors that take account of the clustering of observations. Column 1 shows that countries with greater years of schooling have a smaller spread in skill differentials than other countries but that the country's rank by the Heritage Foundation economic freedom indicator has effectively no relation on the dispersion of pay. The negative estimated effect of education on differentials should not, however, be interpreted as an elasticity of relative wages to exogenous changes in schooling, since schooling will itself respond to the returns to skill. It is a reduced form estimate that understates the potential effect of exogenous changes in schooling on the spread of pay across occupations. The negligible relation between the broad economic freedom indicator and occupational differentials suggests that countries with differing reliance on market forces have considerable option to choose different levels of inequality -- the Nordic countries, being the most notable case in

point. Addition of log GDP per capita to the regression greatly reduces the effect of schooling on the occupational wage structure. This reflects the strong correlation between levels of schooling and levels of GDP per capita in cross section country contrasts. It may also indicate that rises in GDP tend to reduce dispersion of pay through a broader set of mechanisms than increased schooling. In columns 3 and 4 we have dropped the economic freedom ranking and entered our measures of institutional wage-setting: unionism in market economies and the Heritage Foundation measure of wage/price regulation. The column 3 regression shows that unionism has a substantial reductive effect on the dispersion of occupational pay. This reflects the greater unionisation of lower skilled workers in many countries and the tendency for collective bargaining of various forms to improve the wages of the less skilled. The wage-price regulation index in column 4 takes on only 5 values, with lower scores reflecting greater market freedom in wage and price setting and higher scores reflecting greater regulation. It also obtains a negative coefficient, indicating that regulation reduces inequality.

The most problematic regressions in table 6 relate to the impact of trade on the occupational wage structure. Our calculations using the Heritage Foundation's score for free trade showed no relation between the score and occupational inequality. But we obtained some fragile relations between the ratio of exports and imports to GDP and occupational wage dispersion. As column 5 shows, when this measure of trade  $\times \ln$  of GDP is included in the regression without the union and institutional wage-setting indicator, it obtains the anticipated positive effect, indicating that greater trade raises dispersion of pay more in high income countries. But column 6 shows that inclusion of unionism by itself reduces the measured trade effect substantially. The relation between the extent of trade and the occupational wage structure is sufficiently weak to be highly sensitive to inclusion of other variables and to the way we treat Singapore and Hong Kong, which are extreme outliers in terms of their reliance on trade. If trade has sizable effects on occupational wage structures, it will take a much more detailed analysis than ours to nail them down with any precision.

In short, the results in table 6 suggest that within the limits of extant measures, institutional factors relating to wage-setting and the level of GDP per capita have the most pronounced impact on occupational wage structures across countries. While additional or better measures of the various potential determinants and extension of the analysis from cross section to time series variation could alter this broad conclusion, it is sufficiently consonant with studies from more limited country samples (Freeman and Katz) that we expect the results to be fairly robust to further investigation.

## Conclusion

Lack of an internationally accepted body of data on wages by skill has hampered analysis of occupational wage structures and of the wages of comparably skilled workers across the world. In this study we introduce the Occupational Wages around the World (OWW) data file, which gives consistent pay in 161 occupations in over 150 countries from 1983 to 1988 and propose this file as the best source of data by skill around the world. We developed the OWW by standardizing the diverse data in the rarely used International Labor Organization's "October Inquiry" survey into rates of pay defined in the same units for workers by occupation in all countries and time periods. Variant standardizations yielded similar results, so that the OWW file is robust to plausible alternative ways of standardizing the Inquiry data. The OWW data file that we use can be extended back in time to at least the 1950s, though with less occupational detail.

Our analysis of the OWW shows that, consistent with extant wisdom, economic development was associated with smaller skill differentials, due in part to the difference in educational attainment between more and less developed countries but in larger part to institutional factors such as collective bargaining; but that contrary with extant wisdom and fears about the effects of globalisation on wages, the disparity of pay for similar work across countries widened in both exchange rate and purchasing power parity terms in the period studied.

The OWW offers a rich data set on the price side of the labour market. Combined with additional data from the quantity side of the labour market, such as the structure of employment or unemployment,<sup>21</sup> and with more refined measures of economic institutions and policies, we hope that it will illuminate important questions about the position of labour in the global economy in the future.

**TABLE 1 : The Development of the ILO October Inquiry Survey**

	<b>Number of Occupations</b>	<b>Number of Countries Reporting In Given Year</b>	<b>Number of Countries Reporting Cumulative Reporting</b>
1924:	18 occupations	15	15
1925-28:	18 occupations	16-19	20
1929:	30 occupations	17	20
1930-1950	30 occupations	16 - 37	69
1951	41 occupations	19	69
1952	41 occupations	48	77
1953	48 occupations	47	82
1954-1982	48 occupations	54-103	181
1983	161 occupations	56	182
1984-1997	161 occupations	48-76	194
1998p	161 occupations	14	194

Source: International Labour Organization, October Inquiry Survey; 1983-1998 from computer data files; earlier years from published reports:

1924-1950: International Labour Review (Vol.10-54)

1935-1957: ILO Yearbook of Labour Statistics (Issue X-XVIII)

1951-1963: International Labour Review - Special Issue of the Statistical Supplement (Vol.66,70,72,74,76,78,80,82,84,86,88,90)

1964-1982: Bulletin of Labour Statistics (1965-1983)

**TABLE 2:** Observations in the 1983-1998 October Inquiry Computer Files**A. SAMPLE SIZE**

<b>Maximum Conceivable Observations</b>	<b>401,856</b>
observations missing because country did not report in given year	242,466
observations missing because occupation missing in year country reported	85,344
<b>Actual year/country/occupation observation</b>	<b>74,056</b>
observations with multiple figures	34,262
multiple figures	49,960
<b>Total, including all multiple observations</b>	<b>124,016</b>

**B. COUNTRIES AND OCCUPATIONS WITH AT LEAST ONE REPORTED WAGE STATISTIC****Countries With Reported Wage Statistic for different numbers of occupations**

#s of occupation	# of countries (total: 156)
<30	11
30-59	21
60-79	24
80-99	18
100-119	34
120-139	15
140+	33

**Occupations with one Reported Wage Statistic for different numbers of countries**

#s of countries reporting on occupation	# of occupations (total: 161)
<59	17
60-79	28
80-99	42
100-119	53
120+	21

**C. ACTUAL OBSERVATIONS**

<b>Pay concept</b>	
Wages (141 countries)	81,869
Earnings (93 countries)	42,147
<b>Averaging concept</b>	
Mean	83766
Minimum	28541
Maximum	4515
Prevailing	4596
Median	1803
Other	148
Missing	647

(Table 2: continued)

<b>Period concept</b>	
Monthly	75882
Hourly <sup>1</sup>	22775
Weekly	15184
Daily	6359
Annual	2444
Fortnight	1197
Other	133
Missing	42
<b>Gender</b>	
Males	53605
Both	44707
Females	25704

Source: Tabulated from ILO October Inquiry Files, 1983-1998

---

<sup>1</sup> The hourly figures include small number for whom the data tell us that they relate to hours paid for and another small number for whom the wages relate to hours worked.

TABLE 3: Measures of the Occupational Wage Structure Within Countries, 1988-1992

	# Occup	Sd Ln W		p90/p10	
		All	Decile	All	Decile
<b>High Income Economies</b>					
AT – Austria	144	0.33	0.33	2.30	2.45
AU – Australia	149	0.24	0.24	1.71	1.90
BE – Belgium	42	0.17	0.17	1.49	1.52
DE – Germany	159	0.37	0.35	2.56	2.59
DK – Denmark	57	0.19	0.19	1.57	1.67
FI – Finland	124	0.24	0.24	1.80	1.90
GB – United Kingdom	50	0.28	0.26	2.04	2.01
HK – Hong Kong	52	0.41	0.38	2.34	2.74
IT – Italy	144	0.27	0.25	1.76	1.92
JP – Japan	41	0.35	0.30	1.96	2.25
NL – Netherlands	70	0.15	0.13	1.43	1.41
NO – Norway	31	0.18	0.17	1.68	1.56
NZ – New Zealand	136	0.35	0.31	2.25	2.30
SE – Sweden	130	0.20	0.17	1.55	1.57
SG – Singapore	121	0.54	0.50	3.72	3.77
US – United States	86	0.34	0.34	2.31	2.45
<b>Upper Middle Income Economies</b>					
	# Occup	Sd Ln W		p90/p10	
		All	Decile	All	Decile
AR – Argentina	133	0.75	0.71	7.46	6.95
GA – Gabon	60	0.62	0.63	4.96	5.39
KR – Korea, Republic of	129	0.39	0.40	2.45	2.92
MU – Mauritius	89	0.50	0.49	4.19	3.74
MX – Mexico	46	0.18	0.11	1.44	1.35
PR – Puerto Rico	48	0.34	0.31	2.37	2.24
TT – Trinidad and Tobago	118	0.56	0.52	3.29	3.95
UY – Uruguay	45	0.50	0.50	2.70	3.79
VE – Venezuela	142	0.44	0.41	2.91	2.92



**Lower Middle Income Economies**

	# Occup	Sd Ln W		p90/p10	
		All	Decile	All	Decile
BO – Bolivia	117	0.65	0.62	4.89	5.49
CI – Cote d’Ivoire	134	0.66	0.67	5.34	6.38
CM – Cameroun	51	0.71	0.74	6.77	6.49
CO – Colombia	41	0.65	0.59	5.12	4.82
DZ – Algeria	135	0.33	0.31	2.57	2.37
JO – Jordan	105	0.43	0.38	3.12	2.81
PE – Peru	34	0.60	0.52	3.46	4.19
PH – Phillipines	36	0.14	0.15	1.33	1.46
SN – Senegal	73	0.44	0.40	2.87	3.00
SV – El Salvador	138	0.54	0.54	3.70	4.38
TH – Thailand	125	0.49	0.49	3.68	3.72
TN – Tunisia	67	0.16	0.10	1.21	1.23
TR - Turkey	45	0.27	0.26	2.07	1.97

**Low Income Economies**

	# Occup	Sd Ln W		p90/p10	
		All	Decile	All	Decile
BD – Bangladesh	133	0.48	0.44	3.10	3.33
BF – Burkina Faso	110	0.43	0.43	2.90	3.11
BI – Burundi	69	0.76	0.77	6.11	7.62
BJ – Benin	72	0.74	0.72	6.72	6.65
CF – Central African Republic	91	0.78	0.78	6.05	7.97
EG – Egypt	81	0.36	0.34	2.18	2.50
HN – Honduras	109	0.56	0.56	4.53	4.72
ID – Indonesia	33	0.29	0.29	1.96	2.10
IN – India	92	0.60	0.52	4.09	3.73
LS – Lesotho	106	0.69	0.71	5.61	7.08
ML – Mali	112	0.60	0.58	4.42	4.85
MM – Myanmar	138	0.29	0.28	2.09	2.14
MZ – Mozambique	117	0.45	0.45	3.19	3.48
NP – Nepal	32	0.41	0.36	2.20	2.13
RW – Rwanda	126	0.70	0.72	6.85	7.01
SD – Sudan	130	0.42	0.35	2.35	2.59
SL – Sierra Leone	102	0.59	0.57	2.93	4.79
TD – Tchad	91	0.76	0.77	6.13	7.25
TG – Togo	39	0.52	0.50	3.40	3.60
ZM – Zambia	132	0.56	0.53	3.87	4.23

**Economies with Small Populations**

	# Occup	Sd Ln W		p90/p10	
		All	Decile	All	Decile
AG – Antigua and Barbuda	64	0.41	0.42	2.53	3.12
AI – Anguilla	56	0.37	0.34	2.59	2.46
AN – Netherlands Antilles	47	0.40	0.36	2.64	2.69
BB – Barbados	97	0.44	0.46	3.11	3.43
BH – Bahrain	84	0.54	0.51	3.67	3.98
BM – Bermuda	46	0.42	0.38	2.11	2.54
BS – Bahamas	39	0.47	0.51	3.25	3.66
BZ – Belize	100	0.56	0.55	3.23	4.23
CY – Cyprus	112	0.47	0.44	3.43	3.25
FK – Falkland Islands (Malvinas)	65	0.39	0.34	2.33	2.49
GF – Guyane, French	63	0.37	0.33	2.45	2.41
GI – Gibraltar	32	0.30	0.28	1.91	2.05
IM – Isle of Man	58	0.33	0.32	2.26	2.40
KM – Comores	74	0.55	0.54	4.01	4.37
LC – St. Lucia	97	0.54	0.56	3.61	4.51
PF – Polynesia, French	87	0.46	0.44	3.40	3.23
PM – Saint-Pierre-et-Miquelon	67	0.30	0.32	2.10	2.40
SC – Seychelles	58	0.39	0.40	3.28	3.06
SR – Suriname	57	0.53	0.49	3.71	3.87
VI – Virgin Islands	70	0.40	0.40	2.79	2.93

**Communist & Ex-Communist Economies**

	# Occup	SD Ln W		p90/p10	
		All	Decile	All	Decile
BG – Bulgaria	112	0.28	0.25	1.88	1.97
CN – China	81	0.24	0.23	1.77	1.85
CU-- Cuba	129	0.28	0.26	2.00	1.90
CZ – Czech Republic	110	0.20	0.20	1.61	1.73
RO – Romania	160	0.24	0.24	1.88	1.89
RU – Russian Federation	41	0.28	0.28	2.05	2.19
SI – Slovenia	57	0.34	0.35	2.93	2.58
SK – Slovakia	110	0.20	0.20	1.61	1.73
YU – Yugoslavia	159	0.36	0.36	2.40	2.69

NOTE: In some tabulations we allocate the small economies or communist economies to their appropriate income class.

**TABLE 4:** Regression Estimates of the Effect of Per Capita Income and Trend on Standard Deviation of Log Occupational Wages (Standard Errors in Parenthesis)

	Cross Section			Time Series			
	OLS	OLS	MEDIAN				
	1	2	3	4	5	6	7
Log GDP per Capita	-.098 (.011)	.135 (.186)	.413 (.261)			-.064 (.031)	
(Log GDP per Capita) <sup>2</sup>		-.014 (.011)	-.031 (.016)				
Communist (1=comm. years)	-.19 (.02)	-.20 (.02)	-.23 (.03)	-.11 (.02)	-.05 (.03)	-.08 (.02)	-.10 (.03)
Trend				-.0010 (.0008)			
Trend, by Income Level							
High Income					-.002 (.001)		
Upper Middle Income					-.006 (.002)		
Lower Middle Income					.007 (.002)		
Low Income					-.002 (.002)		
Log GDP per Cap., by Income Level							
High Income							-.068 (.054)
Upper Middle Income							-.174 (.046)
Lower Middle Income							.072 (.049)
Low Income							-.085 (.062)
Country Dummies				✓	✓	✓	✓
Year Dummies	✓	✓	✓			✓	✓
Constant	✓	✓	✓	✓	✓	✓	✓
Observations	652	652	652	830	830	652	652
R <sup>2</sup>	.424	.431	.291	0.84	0.84	0.88	0.88

Notes: These regressions used dispersions based on raw data for countries that reported on 30 or more occupations. Regressions based on standard deviations of wages based on estimates of decile wages gave nearly identical results. Standard errors are corrected for clustering. Standard errors for median regression are based on (clustered) bootstraps with 1000 replications.

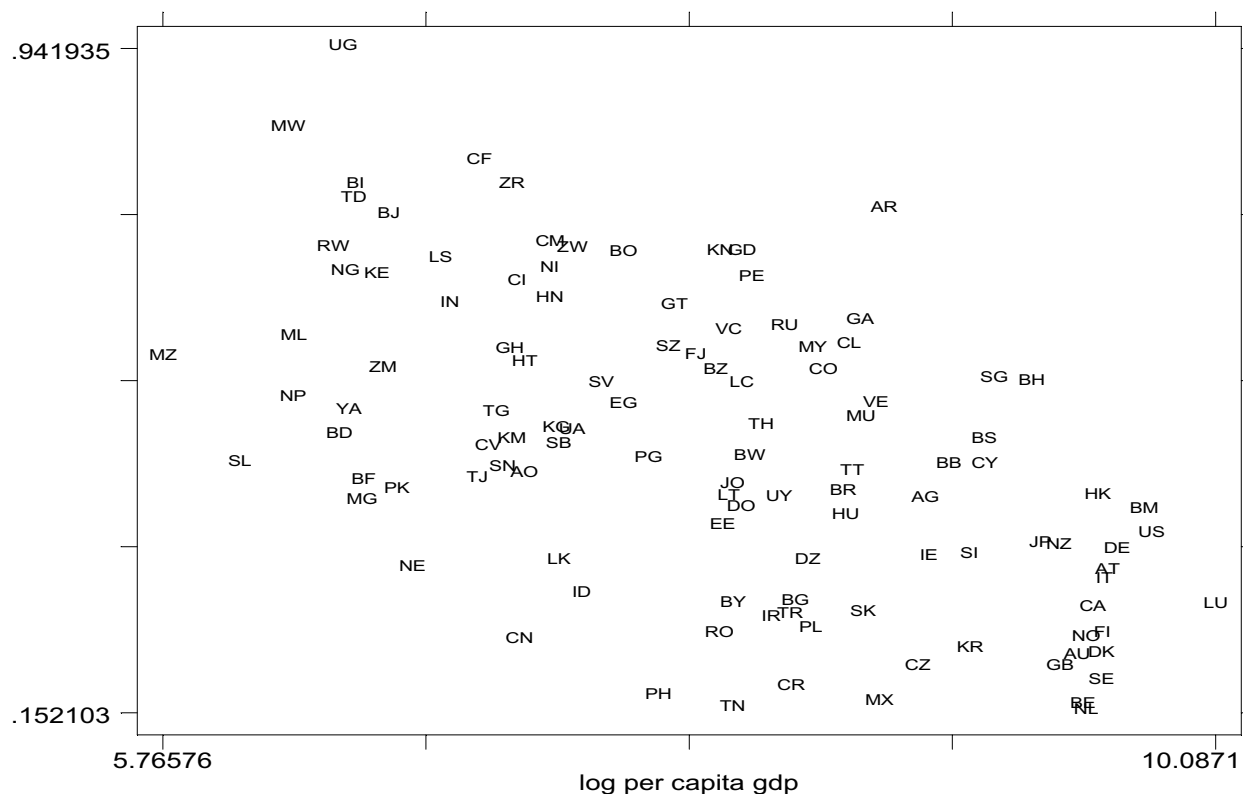
**TABLE 5:** Estimates of Differences in Wages in Occupations Across Countries  
(Using Exchange Rates)

	Median/Max Wage		SD Ln Wage	
	1	2	3	4
Trend	-.011 (.004)	-.006 (.0005)	.023 (.002)	.011 (.003)
Occupation Dummies	✓	✓	✓	✓
Rank by Skill x trend (1=least skilled)		-.000056 (.000005)		.00016 (.00003)
Constant	✓	✓	✓	✓
R <sup>2</sup>	.69	.72	.62	.64
Number of Observations	1416	1416	1416	1416

**TABLE 6:** Regression Coefficients and Standard Errors for Factors that Affect the Standard Deviation of Ln Wages Across Countries

	1	2	3	4	5	6
<b>Economy Wide</b>						
GDP per capita		-.071 (.036)	-.046 (.034)	-.054 (.032)	-.096 (.030)	-.062 (.039)
Years of Schooling	-.032 (.006)	-.017 (.010)	-.007 (.012)	-.013 (.011)	-.005 (.010)	-.003 (.013)
Economic Freedom Rank	-.000 (.001)	-.001 (.001)				
Communist	-.112 (.054)	-.146 (.050)	-.202 (.025)	-.177 (.030)	-.186 (.025)	-.220 (.027)
<b>Wage-Setting</b>						
Union			-.173 (.052)	-.146 (.043)		-.161 (.052)
Wage-Price Regulations				-.045 (.027)		
<b>Trade</b>						
(X+M)/GDP x GDP per capita (X 1000)					.046 (.025)	.020 (.028)
Year Dummies	✓	✓	✓	✓	✓	✓
<b>R<sup>2</sup></b>	.34	.42	.39	.46	.41	.41
<b>Number of Observations</b>	522	459	401	386	503	389

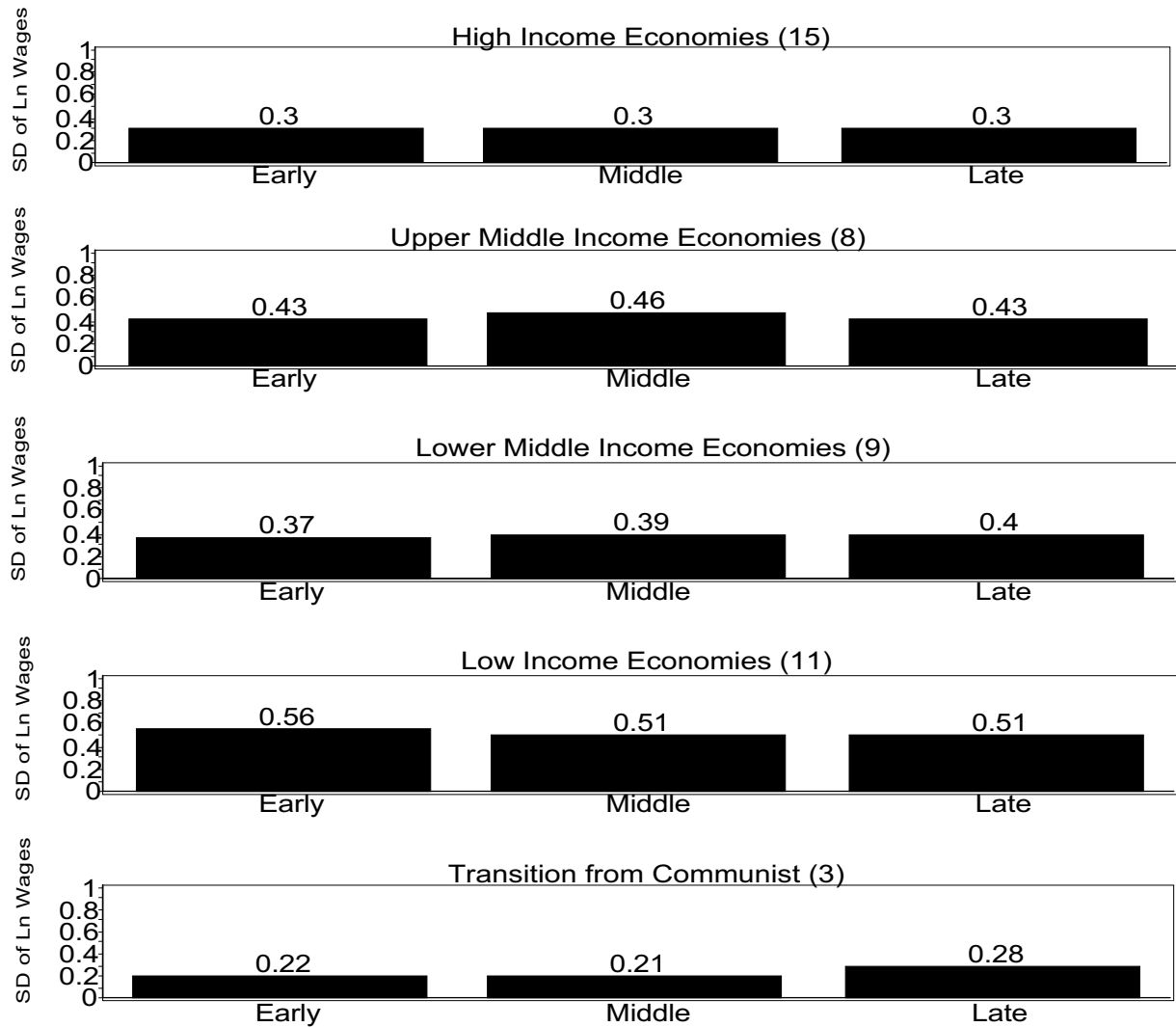
Figure 1: Standard deviation of ln occupational wages vs log GDP per capita, by country



NOTE: Countries included are:

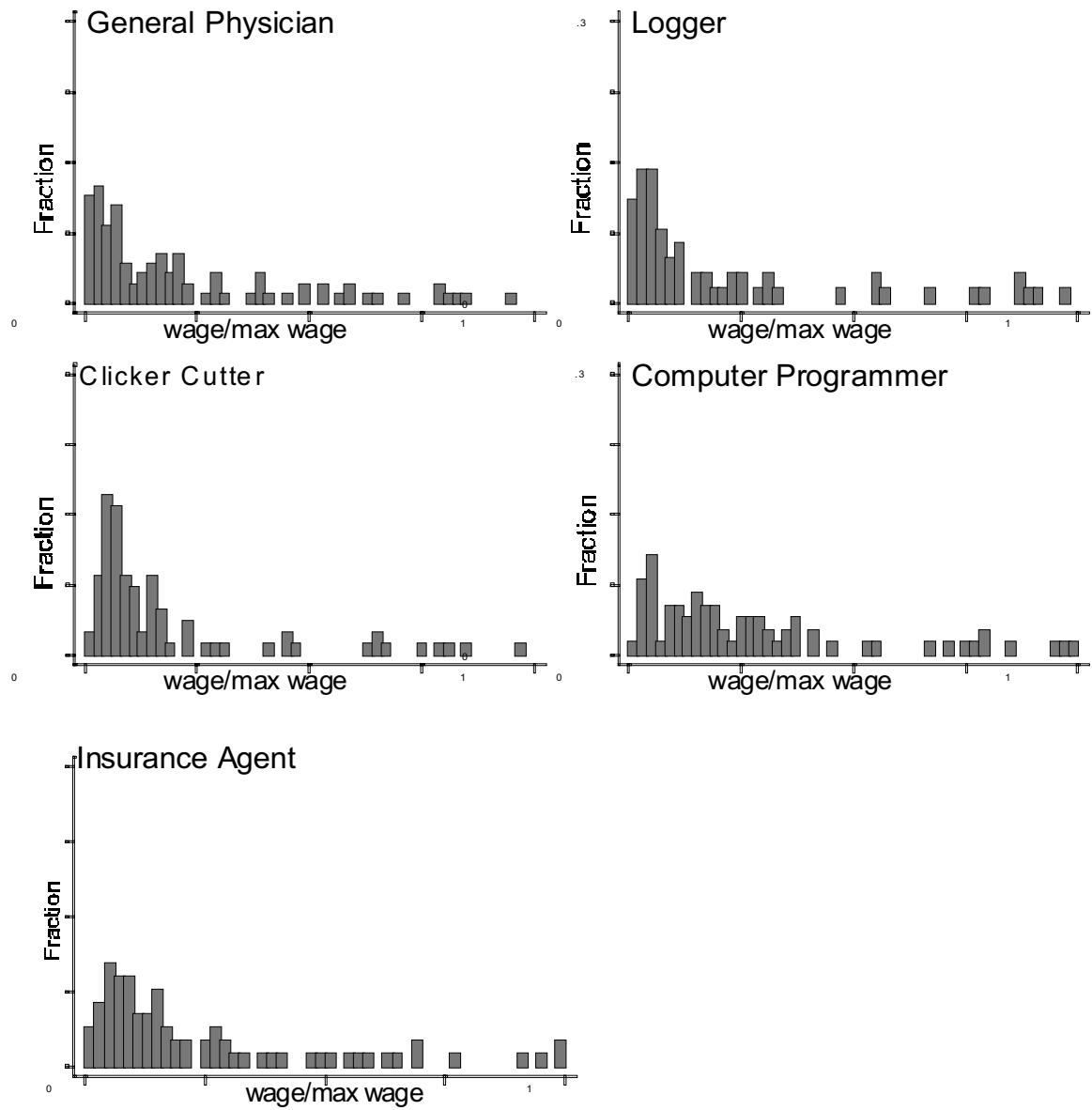
AO-Angola; AG-Antigua and Barbuda; AR-Argentina; AT-Austria; AU-Australia; BB-Barbados; BD-Bangladesh; BE-Belgique; BF-Burkina Faso; BG-Bulgaria; BH-Bahrain; BI-Burundi; BJ-Benin; BM-Bermuda; BO-Bolivia; BR-Brasil; BS-Bahamas; BW-Botswana; BY-Belarus; BZ-Belize; CA-Canada; CF-Central African Republic; CI-Cote d'Ivoire; CL-Chile; CM-Cameroun; CN-China; CO-Colombia; CR-Costa-Rica; CV-Cap Vert; CY-Cyprus; CZ-Czech Republic; DE-Germany; DK-Denmark; DO-Dominican Republic; DZ-Algeria; EE-Estonia; EG-Egypt; FI-Finland; FJ-Fiji; GA-Gabon; GB-United Kingdom; GD-Grenada; GH-Ghana; GT-Guatemala; HK-Hong Kong; HN-Honduras; HT-Haiti; HU-Hungary; ID-Indonesia; IE-Ireland; IN-India; IR-Iran, Islamic Rep.of; IT-Italie; JO-Jordan; JP-Japan; KE-Kenya; KG-Kyrgyzstan; KM-Comores; KN-Saint Kitts and Nevis; KR-Korea, Republic of; LC-St.-Lucia; LK-Sri Lanka; LS-Lesotho; LT-Lithuania; LU-Luxembourg; MG-Madagascar; ML-Mali; MU-Mauritius; MW-Malawi; MX-Mexico; MY-Malaysia; MZ-Mozambique; NE-Niger; NG-Nigeria; NI-Nicaragua; NL-Netherlands; NO-Norway; NP-Nepal; NZ-New Zealand; PE-Peru; PG-Papua New Guinea; PH-Philippines; PK-Pakistan; PL-Poland; RO-Roumanie; RU-Russian Federation; RW-Rwanda; SB-Solomon Islands; SE-Sweden; SG-Singapore; SI-Slovenia; SK-Slovakia; SL-Sierra Leone; SN-Senegal; SV-El-Salvador; SZ-Swaziland; TD-Tchad; TG-Togo; TH-Thailand; TJ-Tajikistan; TN-Tunisie; TR-Turkey; TT-Trinidad and Tobago; UA-Ukraine; UG-Uganda; US-United States; UY-Uruguay; VC-St.Vincent and the Grenadines; VE-Venezuela; YA-Yemen; ZM-Zambia; ZR-Zaire; ZW-Zimbabwe

**Figure 2:** Trends in skill differentials by groups of countries (based on consistent occupations within countries over time)



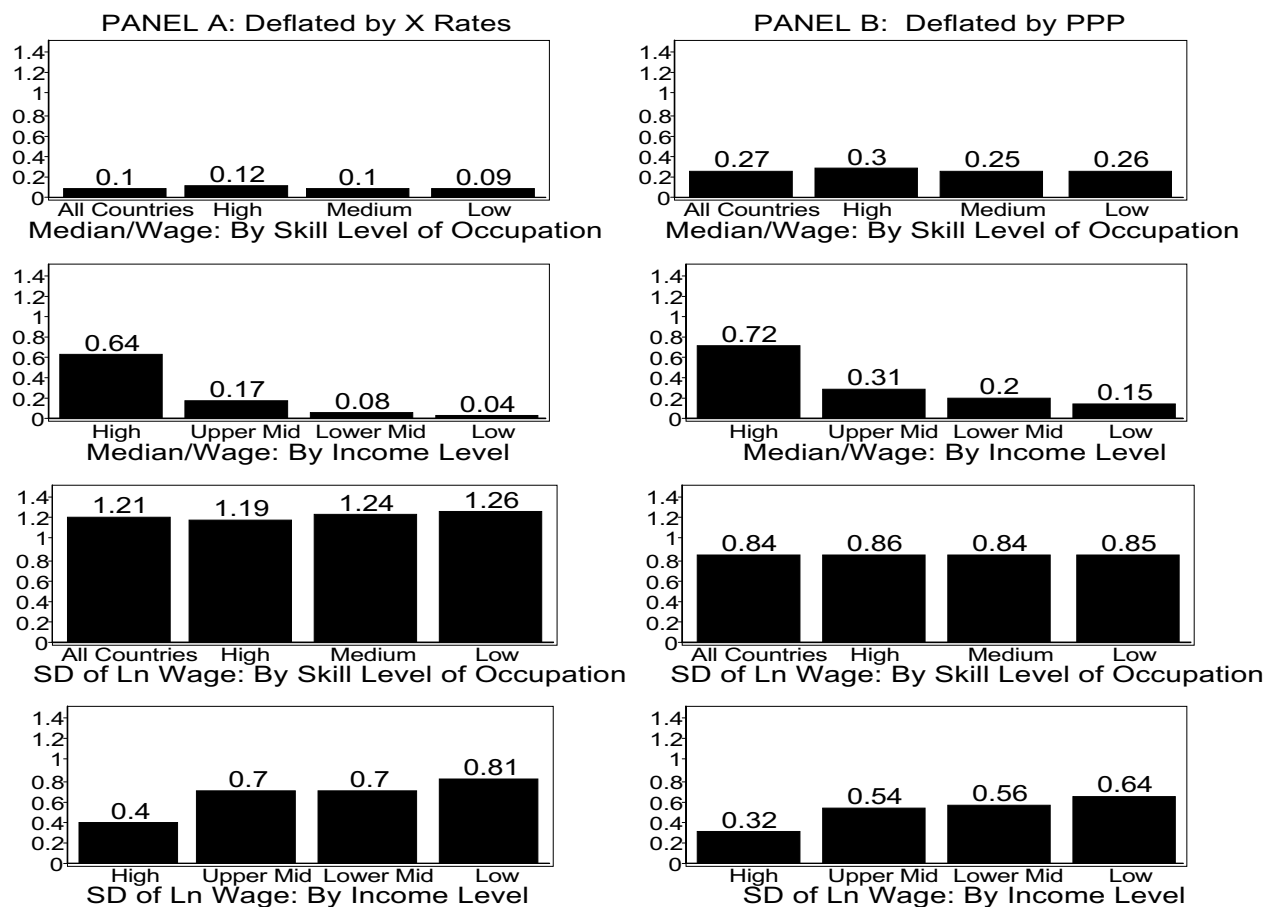
NOTES: This analysis is limited to countries that reported 30 or more occupations. The results are comparable if we include countries with less than 30 occupations as well.

**Figure 3:** Variation in wages around the world in five occupations, 1983-98





**Figure 4:** Cross-country variation in wages for workers in the same occupation



NOTES: Panel A based on 45 countries, with 11 in the high income group; 10 in the upper middle income group; 9 in the lower middle income group; and 15 in the low income group. Occupations classified as high, medium, and low skill based on 1992 German occupational wage distribution, with high skill defined as top 30 occupations in terms of wages, low skill defined as the bottom 30 in terms of wage, and the medium skill defined as the 30 occupations in the middle of the distribution (around the median) in terms of wage.

Panel B based on 38 countries, with 11 in the high income group; 6 in the upper middle income group; 7 in the lower middle income group; and 14 in the low income group. Standard deviation of logs for the country groups are standard deviations within those groups.

NB: Similar results are obtained if we restrict the countries to those for which we have PPPs.

## Appendix. Data Structure and Data Calibration

### I. Data structure

Each wage or earnings observation can be written as the sum of a 'normalized' wage or earnings rate  $y_{n,j,o,t}^*$ , a 'data type' correction factor  $x_{i,j,o,t}\beta$ , and an error term  $v_{i,j,o,t}$ :

$$y_{i,j,o,t} = y_{n,j,o,t}^* + x_{i,j,o,t}\beta + v_{i,j,o,t} \quad (1)$$

where  $y_{i,j,o,t}$  is a log wage or earnings observation of data type  $i$  ( $i=1,\dots,I$ ), country  $j$  ( $j=1,\dots,J$ ), occupation  $o$  ( $o=1,\dots,O$ ), and period  $t$  ( $t=1,\dots,T$ ),  $y_{n,j,o,t}^*$  is the (unobserved) wage or earnings of the normalized data type,  $x_{i,j,o,t}$  is a row vector of data type dummies which equal one if the observation is of the corresponding type,  $\beta$  is a column vector of coefficients for each data type dummy, and  $v_{i,j,o,t}$  is an error term. The normalized data type is defined as the average monthly male wage, and data types are defined in deviation of the normalized type, and an element of the Cartesian product  $Y_1 * Y_2 * Y_3 * Y_4$ , where<sup>2</sup>

$Y_1 = \{\text{wage, earnings}\}$

$Y_2 = \{\text{male, female, both sexes}\}$

$Y_3 = \{\text{hourly, daily, monthly}\}$ <sup>3</sup>

$Y_4 = \{\text{average, median, prevailing, minimum, maximum}\}$

For the normalized data type we define  $x_{n,j,o,t}=0$ , and therefore

$$y_{n,j,o,t} = y_{n,j,o,t}^* + v_{n,j,o,t} \quad (2)$$

The error term  $v_{i,j,o,t}$  includes two components: (1) an unobservable data correction term  $\varepsilon_{i,j,o,t}$ , and (2) a measurement error term  $\eta_{i,j,o,t}$ . The unobservable data correction term reflects data type variation which is not captured by the data type correction factor  $x_{i,j,o,t}\beta$ , such as country-specific gender effects, and which equals zero for the normalized data type ( $\varepsilon_{n,j,o,t}=0$ ). In principle it would be possible to include these unobservable data correction terms in the data type correction factor  $x_{i,j,o,t}\beta$  by allowing  $\beta$  to vary with (j,o,t) (fixed effects). However, there is not enough variation in the data to estimate these country / occupation / period-specific data type fixed effects. The error term  $v_{i,j,o,t}$  can therefore be written as:

$$v_{i,j,o,t} = \varepsilon_{i,j,o,t} + \eta_{i,j,o,t} \quad (3)$$

We assume that  $\varepsilon_{i,j,o,t}$  and  $\eta_{i,j,o,t}$  are i.i.d. noise variables and mutually independent:

$$\begin{aligned} \text{cov}(\varepsilon_{i,j,o,t}, y_{n,j,o,t}) &= 0 \\ \text{cov}(\varepsilon_{i,j,o,t}, x_{n,j,o,t}) &= 0 \\ \text{cov}(\varepsilon_{i,j,o,t}, \eta_{n,j,o,t}) &= 0 \\ \text{cov}(\eta_{i,j,o,t}, y_{n,j,o,t}) &= 0 \\ \text{cov}(\eta_{i,j,o,t}, x_{n,j,o,t}) &= 0 \\ \text{var}(\eta_{i,j,o,t}) &= \sigma_{\eta}^2 \\ \text{var}(\varepsilon_{i,j,o,t}) &= \sigma_{\varepsilon_i}^2 \end{aligned} \quad (4)$$

<sup>2</sup> The original data also varies with respect to cities and regions for which data are available. We have selected for each country / occupation / period observation the most representative cities/regions. Sometimes also a range was given, with a minimum and a maximum value. In these instances the average of the range values was used.

<sup>3</sup> The original data contains observations with the following time units: {yearly, hourly, daily, monthly, fortnightly, weekly, per 12 days, per 15 days}. The yearly, fortnightly, weekly, and per 12/15 days observations can straightforwardly be transformed in monthly data. The hourly observations have been transformed into monthly data in so far information was available about the number of hours worked per week.

Suppose we have observations of multiple data types for a given country  $j$ , occupation  $o$ , and period  $t$ . What would be the best estimate of  $y_{n,j,o,t}^*$ , the normalized log wage? Suppose we can write  $y_{n,j,o,t}^*$  as a linear function of a row vector of country, occupation and time dummies (including interaction terms)  $D_{j,o,t}$ :

$$y_{n,j,o,t}^* = D_{j,o,t} \alpha \quad (5)$$

where  $\alpha$  is a vector of parameters. In this case, equations (1)-(5) imply a regression model for the observed wage and earnings rates with a heteroskedastic error structure:

$$y_{i,j,o,t} = D_{j,o,t} \alpha + x_{i,j,o,t} \beta + v_{i,j,o,t} \quad (6)$$

$$\text{var}(v_{i,j,o,t}) = \begin{cases} \sigma_{\eta}^2 & i = n \\ \sigma_{\varepsilon_i}^2 + \sigma_{\eta}^2 & i \neq n \end{cases} \quad (7)$$

The parameters of equation (6) can be estimated with OLS. This is not efficient, however, and the estimated standard errors are inconsistent. GLS or Weighted Least Squares generates efficient estimates and consistent standard errors. The predicted normalized log wage  $\hat{y}_{n,j,o,t}^*$  is given by

$$\hat{y}_{n,j,o,t}^* = D_{j,o,t} \hat{\alpha} \quad (8)$$

where  $\hat{\alpha}$  are the predicted coefficients for the country, occupation, and time dummies. The estimated standard error of  $\hat{y}_{n,j,o,t}^*$  for given  $j,o,t$  is given by

$$\text{var}(\hat{y}_{n,j,o,t}^*) = (D_{j,o,t} \hat{\Sigma}_{\alpha} D_{j,o,t}') \quad (9)$$

where  $D = \{D_{j,o,t}\}$  is a matrix of the country, occupation, and time dummies (with interaction terms), and  $\hat{\Sigma}_{\alpha}$  is the estimated variance/covariance matrix of  $\hat{\alpha}$ .

Unfortunately the data exhibits not enough variation to estimate equation (6) with a full set of country, occupation, and period dummies with their interaction terms. We therefore include only interaction terms for the country and period dummies, to allow for country-specific time effects. Because many interaction terms are therefore omitted, it might be necessary to include a random country / occupation / period effect  $\theta_{j,o,t}$  to capture omitted variation. Because we are interested in wage inequality, it is important to allow for as much variation as possible in the calibration procedure to avoid data smoothing.

Equation (5) should therefore be rewritten as

$$y_{n,j,o,t}^* = D_{j,t} \alpha_{j,t} + D_o \alpha_o + \theta_{j,o,t} \quad (10)$$

where  $D_{j,t}$  is a row vector of country and time dummies (including interaction terms),  $D_o$  a row vector of occupation dummies,  $\alpha_{j,t}$  and  $\alpha_o$  are vectors of parameters, and  $\theta_{j,o,t}$  the random country / occupation / period effect. In this case, we have a regression model for the observed wage and earnings rates with a block diagonal error structure:

$$y_{n,j,o,t} = D_{j,t} \alpha_{j,t} + D_o \alpha_o + x_{i,j,o,t} \beta + \theta_{j,o,t} + v_{i,j,o,t} \quad (11)$$

$$\text{var}(\theta_{j,o,t} + v_{j,o,t}) = \begin{pmatrix} \sigma_{\theta}^2 + \sigma_{\varepsilon_1}^2 + \sigma_{\eta}^2 & \sigma_{\theta}^2 & \cdots & \sigma_{\theta}^2 \\ \sigma_{\theta}^2 & \sigma_{\theta}^2 + \sigma_{\varepsilon_2}^2 + \sigma_{\eta}^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{\theta}^2 & \cdots & \cdots & \sigma_{\theta}^2 + \sigma_{\varepsilon_t}^2 + \sigma_{\eta}^2 \end{pmatrix} \quad (12)$$

where  $v_{j,o,t}$  is a vector of error terms of  $v_{i,j,o,t}$  for given (j,o,t), and  $\sigma_{\varepsilon_n}^2 = 0$ .

The parameters of equation (11) can be estimated with OLS. This is not efficient, however, and the estimated standard errors are inconsistent. GLS generates efficient estimates and consistent standard errors. The predicted normalized log wage  $\hat{y}_{n,j,o,t}^*$  is given by

$$\hat{y}_{n,j,o,t}^* = D_{j,t} \hat{\alpha}_{j,t} + D_o \hat{\alpha}_o + \hat{\theta}_{j,o,t}^* \quad (13)$$

where  $\hat{\alpha}_{j,t}$ ,  $\hat{\alpha}_o$  are the predicted coefficients and  $\hat{\theta}_{j,o,t}^*$  the estimated random effect. The estimated standard error of  $\hat{y}_{n,j,o,t}^*$  for given (j,o,t) is given by

$$\text{var}(\hat{y}_{n,j,o,t}^*) = (D \hat{\Sigma}_{\alpha} D')_{j,o,t} + \hat{\sigma}_{\theta}^2 \quad (14)$$

where  $D = \{D_{j,t}, D_o\}$  is the combined matrix of the country, occupation, and time dummies,  $\hat{\Sigma}_{\alpha}$  is the estimated variance/covariance matrix of  $\hat{\alpha} = (\hat{\alpha}_{j,t}, \hat{\alpha}_o)$ , and  $\hat{\sigma}_{\theta}^2$  is the estimated variance of the random effect.

## II. Estimation Procedures

In principle we can estimate the regression model (11) taking into account the block diagonal structure of the error variance matrix. In fact this is a model which combines random effects with heteroskedasticity for an unbalanced panel, and no routine for this estimation is available in standard software packages such as STATA and SAS.

We also discuss therefore an alternative calibration procedure, which we call the direct calibration method, which is easier to implement. Suppose that  $\beta$  is known. In this case we can use the following estimate of  $y_{n,j,o,t}^*$ :

$$\bar{y}_{i,j,o,t} = y_{i,j,o,t} - x_{i,j,o,t} \beta = y_{n,j,o,t}^* + \varepsilon_{i,j,o,t} + \eta_{i,j,o,t} \quad (15)$$

This estimate is unbiased and has variance equal to

$$\text{var}(\bar{y}_{i,j,o,t}) = \sigma_{\varepsilon_i}^2 + \sigma_{\eta}^2 \quad (16)$$

For each (j,o,t) there are often multiple  $\bar{y}_{i,j,o,t}$  available. What is the best estimate of  $y_{n,j,o,t}^*$ ? Suppose we take a weighted average of  $\bar{y}_{i,j,o,t}$ :

$$\sum_i w_i \bar{y}_{i,j,o,t} \quad (17)$$

This weighted sum gives an unbiased estimate of  $y_{n,j,o,t}^*$  and has variance equal to

$$\text{var}\left(\sum_i w_i \bar{y}_{i,j,o,t}\right) = \sum_i w_i^2 (\sigma_{\varepsilon_i}^2 + \sigma_\eta^2) \quad (18)$$

This variance is minimized if  $w_i$  is chosen as

$$w_i = \frac{1/(\sigma_{\varepsilon_i}^2 + \sigma_\eta^2)}{\sum_k (1/(\sigma_{\varepsilon_k}^2 + \sigma_\eta^2))} \quad (19)$$

Equation (19) suggests that there are two extreme cases:

(a) uniform weighting

If  $\sigma_\eta^2 \rightarrow \infty$ , then  $w_i \rightarrow 1/I_{j,o,t}$ , with  $I_{j,o,t}$  the number of data types observed for given (j,o,t). In this case measurement error dominates the unobservable data correction error, and each estimated normalized wage is weighted equally.

(b) lexicographic weighting

If  $\sigma_{\varepsilon_i}^2 = \sigma_{\varepsilon_{i'}}^2 \rightarrow \infty \forall i, i' \neq n$ , then

$$\begin{cases} w_n \rightarrow 1, w_{i \neq n} \rightarrow 0 & \exists i = n \\ w_i \rightarrow 1/I & \neg \exists i = n \end{cases}$$

If the unobservable data error dominates, the normalized data type will receive the full weight if observed, otherwise the data types will be weighted uniformly.

In case  $\beta$  is unknown (as is the case here), the same extreme weighting schemes can be derived. In this case equation (11) can be estimated with OLS, and the estimated  $\hat{\beta}$  can be used in equation (15):

$$\hat{\bar{y}}_{i,j,o,t} = y_{i,j,o,t} - x_{i,j,o,y} \hat{\beta} \quad (20)$$

This gives also an unbiased estimate of the normalized wage, but now the variance depends also on the sampling variance of  $\hat{\beta}$ :

$$\text{var}(\hat{\bar{y}}_{i,j,o,t}) = x_{i,j,o,t} \text{var}(\hat{\beta}) x'_{i,j,o,t} + (1 + x_{i,j,o,t} \begin{pmatrix} D'D & D'X \\ XD & XX \end{pmatrix}_X^{-1} (D_{j,o,t} \ x_{i,j,o,t}))' (\sigma_{\varepsilon_i}^2 + \sigma_\eta^2) \quad (21)$$

where  $X = \{x_{i,j,o,t}\}$  is a matrix of the row vectors of data dummies, and the subscript X of the inverted matrix stands for the rows corresponding to the X matrix. Given that  $\text{var}(\hat{\bar{y}}_{i,j,o,t})$  becomes unbounded  $\forall i$  if  $\sigma_\eta^2 \rightarrow \infty$ , uniform weighting is optimal with infinite measurement error. On the other hand, given that  $\text{var}(\hat{\bar{y}}_{i,j,o,t})$  becomes unbounded  $\forall i \neq n$  if  $\sigma_{\varepsilon_i}^2 \rightarrow \infty$ , but equals  $\sigma_\eta^2$  for  $i=n$  (because  $x_{n,j,o,t}=0$ ), lexicographic weighting is optimal if unobservable data type error becomes unbounded. Hence, the uniform and lexicographic weights are extreme cases.<sup>4</sup>

<sup>4</sup> In fact uniform or lexicographic weighting may be *too* extreme if we take into account the fact that with unknown  $\beta$  generally  $\text{cov}(\hat{\bar{y}}_{i,j,o,t}, \hat{\bar{y}}_{i',j,o,t}) \neq 0$  and  $\text{var}(\hat{\bar{y}}_{i,j,o,t}) \neq \text{var}(\hat{\bar{y}}_{i',j,o,t})$  for  $\sigma_{\varepsilon_i}^2 = \sigma_{\varepsilon_{i'}}^2 \rightarrow \infty, i \neq i'$ .

### III Data Calibration

Estimation of the regression model (11), taking into account the block diagonal structure of the error variance matrix, is the most efficient calibration method. On the other hand, the direct calibration method of equation (20) allows a simple calibration which is unbiased, and which does not assume as much knowledge about the variance-covariance structure of the error terms (except if one wants to use optimal weights). For this reason, we have chosen to implement both procedures, and to check whether they give very different results.

Full GLS estimation of the regression model (11) turned out to be impossible because of the large number of regressors. Instead we have used a 2-step procedure, where we first estimated the variance structure of the error term  $v_{i,j,o,t}$ ,  $\hat{\sigma}_{\varepsilon_i}^2 + \hat{\sigma}_{\eta}^2$ , using a fixed effects regression, and next the variance of the random effect term  $\theta_{j,o,t}$ ,  $\hat{\sigma}_{\theta}^2$ ,

using a between regression.<sup>5</sup> These variance estimates allow us to calculate the predicted random effects  $\hat{\theta}_{i,j,o,t}$ .

Using the between estimates  $\hat{\alpha}_{j,t}$  and  $\hat{\alpha}_0$  and the predicted random effects, we can derive the predicted normalized log wage according to formula (13).

The next table summarizes the different calibration regressions. Columns (1) and (2) correspond to the random effects and direct calibration method respectively. We have also done different calibrations using different simplifications of the vector of coefficients  $\beta$  for the direct calibration method. We have tried the following:

- (2) dummies for pay concept ( $Y_1$ ), gender ( $Y_2$ ), period concept ( $Y_3$ ), and statistical concept ( $Y_4$ ).
- (3) same as (2), but with interaction terms with income rank of country (4 groups)
- (4) same as (2), but with interaction terms with region (9 groups).
- (5) same as (2), but with country-specific gender dummies.

**Table A.1 Calibration variants used**

	(1)	(2)	(3)	(4)	(5)
country x year dummies	X	X	X	X	X
occupation dummies	X	X	X	X	X
pay concept dummies	X	X	X	X	X
gender dummies	X	X	X	X	
period dummies	X	X	X	X	X
averaging dummies	X	X	X	X	X
interaction terms income rank			X		
interaction terms region				X	
gender x country dummies					X
estimation method	two-step	direct	direct	direct	direct

The following table gives the correlations of the predicted wage for all calibrations across all countries / occupations / years. We have also experimented with two weighting schemes for the direct calibration method, viz. uniform weighting (wu) and lexicographic weighting (wl)

**Table A.2. Correlations of predicted wages across calibrations**

method	(1)	(2) + wu	(3) + wu	(4) + wu	(5) + wu	(2) + wl	(3) + wl	(4) + wl	(5) + wl
(1)	1								
(2) + wu	0.9935	1							
(3) + wu	0.9928	0.9998	1						
(4) + wu	0.9928	0.9995	0.9996	1					
(5) + wu	0.9854	0.9851	0.9855	0.9830	1				
(2) + wl	0.9878	0.9953	0.9947	0.9950	0.9770	1			
(3) + wl	0.9866	0.9950	0.9947	0.9950	0.9762	0.9999	1		
(4) + wl	0.9870	0.9947	0.9943	0.9953	0.9745	0.9995	0.9997	1	
(5) + wl	0.9854	0.9879	0.9878	0.9859	0.9946	0.9890	0.9884	0.9868	1

<sup>5</sup> These are the first 2 steps in the STATA xtreg, re procedure.

The last 2 tables give the correlations of the standard deviation of log predicted wages and the 90<sup>th</sup>/10<sup>th</sup> percentile ratio of predicted wages across all countries / years for the different calibration methods. Here we have limited the sample to those country / year periods for which at least 30 occupations were reported. Limiting the sample to those countries which report a sufficient number of occupations increases the correlation among the dispersion measures for the different calibrations.

**Table A.3. Correlations of standard deviation of log predicted wages across calibrations**

method	(1)	(2) + wu	(3) + wu	(4) + wu	(5) + wu	(2) + wl	(3) + wl	(4) + wl	(5) + wl
(1)	1								
(2) + wu	0.9901	1							
(3) + wu	0.9898	0.9987	1						
(4) + wu	0.9895	0.9970	0.9978	1					
(5) + wu	0.9903	0.9963	0.9968	0.9968	1				
(2) + wl	0.9880	0.9984	0.9972	0.9958	0.9938	1			
(3) + wl	0.9875	0.9968	0.9982	0.9964	0.9939	0.9987	1		
(4) + wl	0.9869	0.9948	0.9957	0.9982	0.9936	0.9970	0.9979	1	
(5) + wl	0.9891	0.9952	0.9959	0.9962	0.9985	0.9957	0.9960	0.9962	1

**Table A.4. Correlations of 90<sup>th</sup>/10<sup>th</sup> percentile of predicted wages across calibrations**

method	(1)	(2) + wu	(3) + wu	(4) + wu	(5) + wu	(2) + wl	(3) + wl	(4) + wl	(5) + wl
(1)	1								
(2) + wu	0.9733	1							
(3) + wu	0.9716	0.9953	1						
(4) + wu	0.9732	0.9930	0.9942	1					
(5) + wu	0.9716	0.9872	0.9897	0.9928	1				
(2) + wl	0.9681	0.9943	0.9895	0.9893	0.9837	1			
(3) + wl	0.9664	0.9904	0.9937	0.9903	0.9853	0.9957	1		
(4) + wl	0.9662	0.9863	0.9871	0.9940	0.9869	0.9928	0.9942	1	
(5) + wl	0.9650	0.9812	0.9830	0.9878	0.9942	0.9879	0.9893	0.9925	1

It is clear from the above tables that the different calibration methods produce very similar results. For this reason we use the most straightforward calibration in the paper for further analysis, viz. calibration (2) + wu.

## ENDNOTES

<sup>1</sup> Melvin Reder, "Industrial Wages" in Aspects of Labor Economics Universities\_National Bureau Conference Series No. 14 University Microfilms 1962.

<sup>2</sup> Since the supply of workers to education depends on the returns to schooling, however, this analysis requires some consideration of the determinants of those returns and such factors as public provision of schooling and technological change in education and the rest of the economy.

<sup>3</sup> Jan Tinbergen, "Substitution of Graduates by Other Labour," Kyklos 27:2 (1974): 217-26.

<sup>4</sup> George Psacharopoulos, 1994 "Returns to Investment in Education: A Global Update", World Development 22: 1325-43. For an analysis of some of the data problems with these estimates see Bennell, Paul, 1996 "Rates of Return to Education: Does the Conventional Pattern Prevail in sub-Saharan Africa?" World Development 24: 183-199.

<sup>5</sup> The ILO actually asks for information on 159 occupations but it differentiates occupation 139, executives in the government into three sectors; national, regional or provincial, and local governments.

<sup>6</sup> The maximum is the multiplicand of the number of countries (156) times the number of occupations (161) times the number of years (16).

<sup>7</sup> In a few cases the wages are in the form of ranges. We found the midpoint of the range and report it as the wage for the category.

<sup>8</sup> The situation is not quite as dire as this statistic indicates because we can obtain some greater comparability by taking account of the natural time dimensions, such as turning yearly earnings into monthly figures by dividing by 12, turning weekly wages into monthly by multiplying by 4.3, or by multiplying hourly pay by hours worked reported on the survey. But even if we standardize using these procedures, only 16.9 percent of the reported figures are directly comparable.

<sup>9</sup> We can transform into any other scale, such as average earnings for women on a weekly basis. Choosing the most common, however, minimizes the noise introduced by the calibration procedures because it starts with a larger number of non-calibrated wages.

<sup>10</sup> By dimensional analysis we mean simply changing the time units in well-determined ways, such as obtaining weekly pay by dividing annual pay by 52.

<sup>11</sup> For countries which reported at least 30 occupations in a given year.

<sup>12</sup> As long as occupations have different numbers of employees, the distribution of occupational wages will differ from the distribution of individual wages. But as long as we are concerned with the *structure* of wages, treating occupations as units of observation is valid.

<sup>13</sup> This procedure keeps the vast bulk of our country-year data since virtually all countries report on some occupations in a given grouping. Out of 990 country-occupation-year data points, 718 have information on each decile and an additional 104 have information on all but one decile.

<sup>14</sup> The results with our decile measure shows a somewhat different pattern: no change in the more advanced countries and an increase in inequality in the lowest income countries, so the patterns in the text should be viewed cautiously. Regression analysis for both sets of measures of inequality are consistent with the pattern in the figure.

<sup>15</sup> We have compared the Czech Republic and Slovakia with Czechoslovakia when it was a single country.

<sup>16</sup> The statistics of extremes can readily be applied to this analysis, as it deals with the relation between extreme values and the central tendency and dispersion of a distribution. See Gumbel, E. J. Statistics of Extremes 1958. Columbia University Press, New York.

<sup>17</sup> The reason is that Germany reported on almost all of the occupations. For the two occupations for which it did not include wages, we interpolated from the distribution of occupational wages in Australia.

<sup>18</sup> The Penn World Tables' purchasing power parity prices are derived from data obtained by the International Comparison Project (ICP) of the UN Statistical Office, extrapolated to additional countries and time periods. The ICP divides GDP into 151 detailed categories, of which 110 are consumption items, 35 are investment goods, and 5 are government spending. Ideally, countries report national prices for 400 to 700 specific items within these categories, based on ICP detailed specifications of over 1500 commodities, services, or inputs. The item prices are expressed as ratios to US prices and then averaged to yield 151 separate price parities. Prices for each of the three main parts of GDP are then estimated by taking weighted averages of the country basic parity prices, and used to calculate an international price for GDP.

<sup>19</sup> In preliminary calculations we created a separate unionism variable, but because there were too few communist or ex-communist countries in our sample to get a reliable coefficient, we dropped the separate union variable from our final calculations. The calculations gave a negative coefficient to the separate communist union variable similar to that on unionism in market economies.

<sup>20</sup> Edward Leamer has argued forcefully that only prices matter while Paul Krugman has argued that counterfactuals based upon factor content quantity calculations also have important implications for assessing the effect of trade.

<sup>21</sup> Countries with more compressed structures of wages may end up with different distributions of occupations or industries. See Steven Davis and Magnus Henrekson ("Wage-Setting Institutions as Industrial Policy: Swedish



---

Experience Seen from a U.S. Perspective,” NBER Summer Institute Workshop in Labor Studies, Comparative Labor Markets Session, July 28, 1999) for studies of the effect of pay compression on employment structures.