International Labour Review, Vol. 140 (2001), No. 4

The Occupational Wages around the World data file

Richard B. FREEMAN* and Remco H. OOSTENDORP**

Have pay differences between comparably skilled workers increased or decreased around the world as globalization has proceeded? Have occupational wage differentials within countries widened or narrowed? How far are changes in occupational pay both between and within countries attributable to globalization and how far to other economic forces?

Such questions concerning the wages paid to workers with similar skills in different countries and the structure of wages within countries are central to ongoing debates about the operation of the labour market in a global economy. To answer them with any degree of certainty requires data on wages by occupation across countries. But there is no generally accepted data file on wages by occupation across countries. The most widely used cross-country data on wages are the figures reported for manufacturing (and some other major sectors) in the International Labour Office's annual *Yearbook of Labour Statistics*, but these do not differentiate occupations or skills. Scattered data are available from other sources. The Union Bank of Switzerland publishes prices and earnings around the world every three years or so (UBS, 2000). The World Economic Forum has gathered some data on occupations from the companies which participate in its activities (Warner, 2001). The International Metalworkers Federation publishes data on metalworkers' earnings and purchasing power around the world biennially (IMF, 2000). And there are various reports on wages in specific occupations, such as teaching (US Department of Education, 1988;

* Harvard University; Centre for Economic Policy – London School of Economics; and National Bureau of Economic Research, Cambridge, MA. ** Free University of Amsterdam and Amsterdam Institute for International Development. This article is an abridged version of *Wages around the world: Pay across occupations and countries*, NBER Working Paper No. 8058, published by the National Bureau of Economic Research in Dec. 2000.

American Federation of Teachers, 1993; Bell and Elias, 2000). However, none of these data sets cover a sufficiently long period and a sufficient number of countries to address those questions satisfactorily.

This article seeks to improve the available information system by proposing the development of an Occupational Wages around the World (OWW) file from the most far-ranging survey of wages by occupation around the world, the ILO's October Inquiry. First developed in 1924, the October Inquiry has since then expanded the number of occupations and countries it covers.[1] However, the data are difficult to handle, mainly because countries use different definitions and units when reporting data to the Inquiry, rendering the reported pay non-comparable in various ways between countries and, in some cases, within countries. In fact, only 5.7 per cent of the reported wages are estimated to be on precisely the same basis. The published data are also of varying quality, depending on the quality of the statistics provided.

In order to remedy this situation, the various statistics have been calibrated into a normalized wage rate and the Inquiry data thus transformed into usable form. The result is a file on pay in 161 occupations in over 150 countries from 1983 to 1999, in which pay is defined consistently for particular reference groups. The ILO's Bureau of Statistics has provided a coding of the quality of data, allowing differentiation between the data supplied. This article describes how the file was standardized and summarizes some of the patterns apparent in the standardized file.[2]

The October Inquiry

The ILO's October Inquiry on pay by occupation across the world has been conducted since 1924. A questionnaire is sent to national governments requesting information on wages in detailed occupations, generally within particular industries. To assure comparability of occupational definitions between countries, the ILO specifies the work involved in each occupation in great detail. To get a flavour of the degree

[1] The October Inquiry results were published chronologically by the ILO as follows: Data for 1924-45: *International Labour Review* (Geneva), Vols. 10 (No. 5, Nov. 1924) to 54 (No. 3 and 4, Sep.-Oct. 1946). Data for 1934-57: *Year Book of Labour Statistics*, First Issue (Vol. II, 1934-35) to Eighteenth Issue (1958). Data for 1951-63: *International Labour Review* (Geneva), separate monthly *Statistical Supplement*, Vols. 66 (July 1952 supplement) to 89 (July 1964 supplement). Data for 1964-90: *Bulletin of Labour Statistics* (Geneva), relevant issues: second quarter in each year, 1965-85 (1984 results); separate annual edition, *Bulletin of Labour Statistics: October Inquiry Results*, 1985 (1983-84 results) to 1991 (1989-90 results). Data for 1990+: *Statistics on occupational wages and hours of work and on food prices*, special supplement to the *Bulletin of Labour Statistics*, 1992+. Data for 1983+: Also available online on the ILO's web site, at http://laborsta.ilo.org

[2] The original October Inquiry database for 1983+ is available online on the ILO's web site, at http://laborsta.ilo.org, Tables 01 (occupational wages and hours of work) and 02 (retail food prices). The standardized file is available at http://www.nber.org/oww

Table 1. Evolution of the ILO's October Inquiry

	Number of occupations	Number of countries reporting	
		In a given year	Cumulative reporting
1924	18	15	15
1925-28	18	16-19	20
1929	30	17	20
1930-50	30	16-37	69
1951	41	19	69
1952	41	48	77
1953	48	47	82
1954-82	48	54-103	181
1983	161	56	181
1984-97	161	58-76	197
1998	161	53	197
1999	161	45	197

Sources: See footnote 1.

of detail, 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 show 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 …

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 covered. Country coverage increased fairly steadily, so that the 1983-99 Inquiry data files (on which this article focuses) had statistics on wages in 158 countries in at least one year, and on wages for up to 76 countries in any given year. The number of occupations reported on increased from 30 in 1929, to 41 in 1951, to 48 in 1953, reaching 161 in 1983.[3]

[3] The ILO actually asks for information on 159 occupations, but it differentiates three levels in occupation 139 (executives in the government): national, regional or provincial, and local government levels.

Problems of inconsistent reporting

If each country contributed information on 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. The skills of clicker cutters, accountants, economists, whatever the occupation, from Germany, Pakistan, Romania or the United States would essentially be interchangeable, so that one would truly be comparing equivalent labour.

However the October Inquiry data fall short of being ideal. Indeed, the problems involved are such that the Inquiry is one of the least used sources of cross-country data in the world. The main problem is that countries respond to the ILO's request for information in inconsistent ways. Recorded wages are not directly comparable either between countries or in the same country over time, or even between occupations in one country at a point of time. The recorded wages are non-comparable because countries report data from a variety of national sources rather than conducting special surveys to respond to the ILO's request. Some countries, e.g. Honduras and the Philippines, report wages paid in an occupation from an establishment survey. Other countries, e.g. India, report legislated minimum wage rates for certain occupations. Still others, e.g. Germany, report minimum wage rates based on collective agreements on hourly, daily, weekly, or monthly wage rates, depending on the occupation. Moreover, data sources change over time. For example, up to 1985 the United States reported wage rates from trade unions and earnings from the Industry Wage Surveys. From 1986 to 1997, the United States reported median usual weekly earnings from the Current Population Survey. Since 1997, the United States has reported median wage rates from an employer-based survey. Some countries give male workers' wages in certain occupations. Others report both male and female workers' wages. Still others report female workers' wages in certain occupations. And so on.

Another problem is that countries do not report consistently from year to year. In the 1983-99 period, 158 countries reported wages in at least one year, but only five countries reported wages 17 times (i.e. every year), 40 reported 10-16 times, 51 reported 5-9 times, 43 reported 2-4 times, and 19 reported just once. Looking back across the years, in 1983, 56 countries reported wages; in 1985, 71 reported wages; in 1990, 72 reported; in 1992, 60 reported; in 1995, 76 reported; in 1997, 66 reported; and finally in 1999, 45 countries reported wages. The uneven pattern in reporting makes it tricky to conduct time series and trend analyses. In addition, over time the ILO has asked for data on different numbers of occupations, which makes trend comparisons difficult, particularly those between the post-1983 period and earlier years. Moreover, some countries do not provide national data but report data from particular regions instead, e.g. major cities or urban areas.

A third problem relates to the concepts of wages used in the October Inquiry. Information is requested on average wage or salary rates and average regular gross earnings, together with the relevant hours of work, with respect to the month of October; see box for detailed definitions.

It follows from these definitions that the October Inquiry does not seek to cover all components of earnings (irregular bonuses, including such important payments as the annual or biannual bonuses paid in Japan and some other Asian countries, may represent a significant part of total gross earnings). Nor does it seek to obtain information on all supplementary labour costs. To the extent that social contributions are often proportionate to wages, this will not affect estimates of relative wage structures in countries, but it will affect, and often underestimate, inter-country differences in labour costs or living standards.

A fourth problem is that even with the ILO's detailed specification of skills, the work performed in a given occupation can vary from one country to another. Even in one country, skills differ within the narrow ILO categories. The range of skills displayed by cooks employed in restaurants and hotels (one of the ILO's specified occupations) in the United Kingdom varies considerably, depending on the size of an establishment, the type of cuisine offered, and the number of stars in the guidebook. Such differences are likely to be even greater between countries. To the extent that differences in skills within occupations are associated with education, the workers in advanced industrialized countries are likely to be more skilled than those in less advanced, developing countries.

Finally, there is the problem of the quality of the data provided to the ILO. As already mentioned, countries send the ILO data obtained from a range of different sources: government agencies; collective agreements; legally determined scales, such as minimum wage rates;

Definitions of data gathered in the October Inquiry

Wage or salary rates. The rates paid for normal time of work; they should include basic wages and salaries, cost-of-living allowances and other guaranteed and regularly paid allowances, but exclude overtime payments, bonuses and gratuities, family allowances and other social security payments made by employers directly to employees, and ex gratia payments in kind.

Average regular gross earnings. The remuneration in cash and in kind paid as a rule at regular intervals, for time worked or work done, together with remuneration for time not worked such as paid holidays and public holidays, before any deductions are made by the employer in respect of employees' taxes, contributions to social security, health insurance, pension schemes and any other obligations of employees. The following should be excluded: employers' contributions in respect of employees paid to social security and pension schemes and also the benefits received by employees under these schemes, as well as severance and termination pay, irregular bonuses such as year-end and other one-time bonuses which accrue over a period longer than a pay period.

Table 2. Quality of the wage data in the October Inquiry

Quality rating	Number (percentage) of rated sources
Not acceptable	1 (0.6)
Poor quality	25 (14.5)
Acceptable/good quality	91 (52.6)
Excellent quality	56 (32.4)
Total	173 (100)

Source: Quality rating of statistics from the ILO's Bureau of Statistics, "Sources and types of wage data in the October Inquiry", Aug. 2001, giving the distribution of all graded sources.

and surveys of varying quality. Approximately half the data are based on surveys, mostly enterprise surveys. There are potential quality problems with each of these sources, depending on the data-gathering process. To help develop the OWW, the ILO's Bureau of Statistics classified the various data sources into four quality groups, ranging from "not acceptable" to "excellent". Table 2 shows the distribution of data quality in the ILO survey classified in these four groups. The vast majority of ratings of data quality cover single countries, but in a few cases where the country used different sources or changed sources over time, the ILO provided multiple ratings. For the sake of simplicity, multiple ratings are treated as independent pieces of information, so that if a country gets a rating of excellent quality for one source and poor for another source, both ratings are entered in the tabulation. The vast bulk of the data are rated as being in the "acceptable/good" category (52.6 per cent) or the "excellent" category (32.4 per cent), but the ILO deems 14.5 per cent of the data of "poor quality" and one source as "not acceptable". In the ensuing tabulations, we have worked with the data deemed of "acceptable" quality or above, and only these data are included in our derived OWW files.

The data in the 1983-99 files

Table 3 provides a detailed description of the types of information contained in the October Inquiry files for 1983 to 1999, the period on which we focus.

Panel A gives information on the size of the sample. The first column gives data from the whole file, including observations the ILO's Bureau of Statistics regards as being of poor quality or not acceptable. We report these data because the published versions of the October Inquiry data contain this information and also because for some analyses data of dubious quality may be preferable to no data at all. The second column records the number of observations only from sources which the ILO's Bureau of Statistics regards as being acceptable/good or excellent quality.

The "maximum conceivable observations" indicate the number of observations that the Inquiry would contain if each country reported a single wage statistic for each occupation yearly: over 432,000 pieces of data.[4] The actual number of observations is smaller, largely because in many years most countries do not report statistics. On average, countries report wages for 6.9 years out of 17 possible years. As a result, over a quarter million (255,990) of the potential observations are missing, because various countries did not report data in particular years. Moreover, in the years when countries did report, they did not report data for every occupation. The main point is that there are 82,543 country-year-occupation cells with wage data in the 1983-99 file. Of those observations, however, 7,449 are from sources that the ILO classifies as poor quality or not acceptable. This means that there are 75,094 actual country-year-occupation observations of acceptable/good or excellent quality.

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 sex and for both sexes. In the first column nearly half the observations (46.2 per cent) contain multiple wage figures. While having multiple figures for the same occupation will help us calibrate the data to a single form, this makes the raw data difficult to use in cross-country comparisons, particularly since countries report pay in different ways. Including multiple wages, there are 137,512 pieces of data, and 125,313 pieces of data of acceptable or better quality.

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 countries report on different numbers of occupations, which creates problems in comparing wage structures between countries. The distribution of occupations by country shows that there are wage data for many occupations in large numbers of countries, which means we can contrast the labour costs and living standards of workers in the same occupation around the world.

Panel C shows the various ways in which countries report wages. Most countries report wage rates from employer surveys or collective agreements or legislated pay schedules. However, many report earnings, some from household surveys but mostly from employer surveys.

---

[4] The maximum is the product of the number of countries (158) times the number of occupations (161) times the number of years (17).

Table 3. Types of observation contained in the October Inquiry computer files, 1983-99

	All data	Acceptable data
A. SAMPLE SIZE		
Maximum conceivable observations	432 446	369 495
Observations missing because country did not report in given year	255 990	211 393
Observations missing because occupation missing in year country reported	93 913	83 008
Actual year/country/occupation observation	82 543	75 094
Observations with multiple figures	38 107	34 117
Multiple figures	54 969	50 219
Total, including all multiple observations	137 512	125 313

B. COUNTRIES AND OCCUPATIONS WITH AT LEAST ONE REPORTED WAGE STATISTIC

Countries with at least one reported wage statistic for different numbers of occupations

No. of occupations	No. of countries (total 158)	No. of countries (total 135)
<30	8	6
30-59	21	17
60-79	21	18
80-99	21	18
100-119	32	27
120-139	20	20
140+	35	29

Occupations with at least one reported wage statistic for different numbers of countries

No. of countries reporting on occupations	No. of occupations (total 161)	
<60	15	22
60-79	25	42
80-99	38	44
100-119	47	48
120+	36	5

	All data	Acceptable data
C. ACTUAL OBSERVATIONS		
Pay concept		
Wages (142 countries)	88 453	82 251
Earnings (95 countries)	49 059	43 062
Averaging concept		
Mean	95 221	85 255
Minimum	30 060	28 902
Maximum	4 202	4 023
Average of min.-max.	336	330
Prevailing	4 491	3 748
Median	2 542	2 434

Other	13	13
Missing	647	638
Period concept		
Monthly	86 906	79 173
Hourly[1]	23 471	22 931
Weekly	15 987	13 883
Daily	7 220	6 562
Annual	2 529	2 308
Fortnight	1 253	439
Other	146	17
Sex		
Male workers	58 555	51 783
Male and female workers	49 294	46 861
Female workers	29 663	26 669

[1] The hourly figures include a small number of observations which concern hours paid for, and another small number which concern wages relating to hours worked.

Source: Tabulated from ILO October Inquiry computer files, 1983-99. *All data* includes observations where the ILO's Bureau of Statistics judged the data to be of poor or not-acceptable quality. *Acceptable data* includes only observations of acceptable/good or excellent quality; see table 2.

Most give statistics in the form of averages,[5] but 22 per cent report minimum wage rates, some from collective agreements. Some countries report maximum wage rates. Others give prevailing wages. After reporting median usual weekly earnings for most occupations from individual reports on the Current Population Survey, the United States shifted to reporting median rates from Occupational Employment Statistics, a national industry-specific occupational employment and wage survey of establishments. The 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 sex: 43 per cent of the observations relate to male workers, 36 per cent to all workers, and 21 per cent to female workers.

In view of all these variants, the vast majority of the Inquiry statistics are simply non-comparable. Just 5.7 per cent simultaneously relate to the most common pay concept (wage rates), use the most common averaging concept (mean), cover the most common time span (monthly), and concern the sex on which more data are available

[5] 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.

(men).[6] How is one to make a valid contrast of pay across country lines when, for example, the United States reports median weekly earnings for both sexes, China reports average monthly wages for female and male workers separately, and Germany gives collectively bargained minimum wage rates for different time periods?

The standardization procedure

To put the Inquiry data into a readily usable form, we undertook a massive standardization exercise. Our goal was to transform each observation, however reported, into a standard rate based on the most common form of data in the Inquiry, namely, monthly average wage rates for male workers.[7]

To see how we standardized the data, consider each observation W to be the sum of a standard rate, W^*, and an adjustment for the way the data are reported, W^a, where a reflects the deviation of the observed wage from the most common form, and an error term, v:

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

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 are 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:

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

The key to the adjustment process is finding the appropriate B coefficients by which to assess how much a given observation must be changed to reach the standard form. Given the data available, 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 using the observations in which there are data for both the standard form and the non-standard form:

---

[6] 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, e.g. 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 the use of these procedures, only 15.7 per cent of the reported figures are directly comparable.

[7] We can transform these observations into any other scale, such as average earnings for women on a weekly basis. Choosing the most common form, however, minimizes the noise introduced by the calibration procedures because it starts with a larger number of non-calibrated wages.

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

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 Bs, W^* can be predicted for observations that did not have both types from the following equation:

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

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

To see what this means in practice, consider a situation in which the standard form is specified as male monthly mean wage rates (because this is the most frequently reported form). Then, an observation which gives mean monthly wage rates for female workers would require one adjustment, for sex. If we know the impact of sex on wage rates for the specific occupation or country, this would require one adjustment in the reported wage rate. As an example, consider the adjustment we actually make for cloth weavers in China in the 1990 data. The reported mean wage rate is 171 yuan per month for female workers. Under our base calibration, we estimate that this wage rate should be raised to 201 yuan per month to be on the same basis as the male wage rates. Fifty per cent of our calibrations involve the addition of one adjustment factor; 46 per cent involve the addition of two adjustment factors; 4 per cent require three adjustments; and there are sporadic cases in which four adjustments are necessary.

However, the extent of variation in the Inquiry data makes the standardization exercise more complicated. The main problem is the numerous different types of data. There are two types of earnings; three types of data by sex; three forms of data by time span that cannot be standardized by dimensional analysis;[8] and five distinct forms of averaging. This gives 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 jungle, 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

[8] 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.

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 19,010 observations, except that many of the data types rarely occur 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 rate/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 this standardization process is that there is no single natural way to simplify the vector of B coefficients. There is 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 sex differentials fall by level of development and we adjust 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 these difficulties is to try several standardization procedures and to examine the different B coefficients they produce and the differences between them in final predicted normalized earnings.

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. 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:

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

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.

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 0.9983 to 0.9998 for the predicted (log) wages, and from 0.9644 to 0.9987 for the dispersion measures used in this article. Details of the calculations can be found in the appendix of Freeman and Oostendorp (2000).

Using the resultant file

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, reported wage data continuously from 1984 to 1991, but not earlier or later. It gave wages for 83 occupations in 1984, for 131 in 1985, for 143 in 1986, and for 135 in 1991. Hungary gave wages for 26 occupations in 1987 and for 130 or so occupations between 1995 and 1999, but gave no figures in other years. And so on.

The new file allows us to examine occupational wage structures within countries over time, as well as comparable labour costs and the living standards of similarly skilled workers across countries. However, in order 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 have compressed the data in two ways.

First, for every country we have calculated the standard deviation of the log of pay and the ratio of the wages in the occupation in the 90th percentile of the occupation wage distribution to the wages in the occupation in the 10th percentile of the distribution for *all of the occupations* reported in a given period. This comparison uses the maximum amount of data but 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 "least common denominator" strategy would greatly reduce the sample size.

Instead, we have chosen a different way of compressing the data, namely to treat observations as samples from the distribution of occupational wages for each country, rather than as estimates of wages for a specific occupation.[9] We then estimate the decile distribution of wages by occupation and calculate measures of dispersion from this distribution. Specifically, we order occupations by their wage in each country time period; divide the ordering into deciles, and take the median wage in each decile as that decile's wage in the country.[10] This gives ten wages for each country, from which we are able to calculate measures of dispersion. For instance, if the top decile of highest-paying occupations in the United States consisted of eight occupations, we would use the median wage among those eight to represent that top decile in the United States, whereas if the top decile of highest-paying occupations in India consisted of only three occupations, we would use the median wage among those three to estimate the top decile in India. In the event, the measures of dispersion of pay based on this decile analysis and measures based on all of the data are highly correlated, which implies that the problem of the number of occupations 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 30 or more occupations, rather than from analysis of the derived decile distribution.

Table 4 presents our measures of the dispersion of wages by occupation for the middle period in the data, 1989-92, organized by countries' level of development. To obtain the maximum number of comparisons over this period, we report the figures for the year in the period which gave the most data. Columns 3 and 5 give the statistics based on all of the occupations reporting in the peak year, while columns 4 and 6 give the statistics based on our estimated decile earnings in the occupation-earnings distribution. Both calculations show that skill differences or wage inequality are smaller in the more advanced countries, and are particularly small in the (then) 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 (GDP per capita in constant PPP). While there is considerable variation in occupational wage inequality for countries with the same level of GDP per capita – for

[9] 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, it is valid to treat occupations as units of observation.

[10] 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 982 country-year data points, 949 have information on each decile.

Table 4. Measures of the occupational wage structure within countries, 1989-92

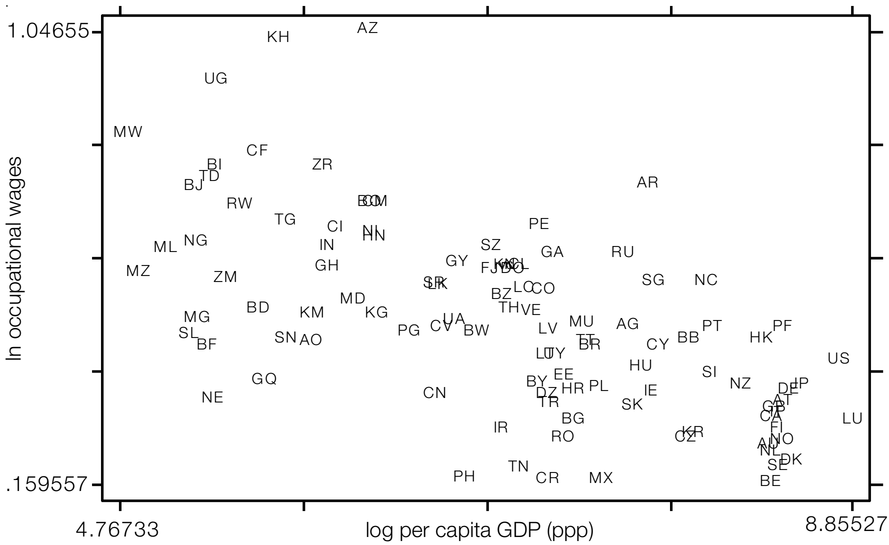
Category of country	Number of occupations	SD ln wage		p90/p10	
		All	Decile	All	Decile
High-income countries					
AT – Austria	144	0.33	0.33	2.18	2.47
AU – Australia	149	0.24	0.26	1.80	1.99
BE – Belgium	42	0.17	0.17	1.49	1.52
DE – Germany	159	0.37	0.35	2.52	2.60
DK – Denmark	57	0.19	0.19	1.57	1.67
FI – Finland	124	0.25	0.26	1.83	1.98
GB – United Kingdom	50	0.28	0.26	2.04	2.01
HK – Hong Kong	52	0.44	0.41	2.53	2.96
IT – Italy	144	0.27	0.25	1.76	1.92
JP – Japan	43	0.35	0.31	1.96	2.31
NL – Netherlands	70	0.22	0.22	1.62	1.75
NO – Norway	31	0.18	0.17	1.69	1.55
NZ – New Zealand	136	0.35	0.31	2.25	2.30
SE – Sweden	130	0.20	0.17	1.53	1.57
SG – Singapore	121	0.54	0.51	3.66	3.94
US – United States	86	0.34	0.35	2.31	2.46
Upper-middle-income countries					
AR – Argentina	133	0.74	0.70	7.32	6.82
GA – Gabon	60	0.61	0.60	4.92	4.80
KR – Korea, Republic of	129	0.37	0.37	2.37	2.67
MU – Mauritius	88	0.48	0.47	3.87	3.58
MX – Mexico	46	0.18	0.11	1.44	1.35
PR – Puerto Rico	48	0.33	0.30	2.20	2.18
TT – Trinidad and Tobago	118	0.56	0.50	3.33	3.87
UY – Uruguay	45	0.50	0.50	2.70	3.79
VE – Venezuela	142	0.42	0.38	2.66	2.70
Lower-middle-income countries					
BO – Bolivia	117	0.67	0.65	4.88	6.03
CO – Colombia	41	0.65	0.58	5.03	4.78
DZ – Algeria	135	0.33	0.31	2.57	2.37
HN – Honduras	109	0.58	0.58	4.94	4.97
PE – Peru	34	0.60	0.51	3.46	4.19
PH – Philippines	36	0.14	0.15	1.33	1.46
TH – Thailand	125	0.50	0.50	3.82	3.78
TN – Tunisia	66	0.16	0.12	1.44	1.40
TR – Turkey	45	0.31	0.32	2.23	2.35
Low-income countries					
BD – Bangladesh	133	0.50	0.46	3.34	3.52
BF – Burkina Faso	110	0.42	0.42	2.92	3.07
BI – Burundi	69	0.76	0.76	6.17	7.54
BJ – Benin	72	0.74	0.71	6.64	6.63

Table 4. Measures of the occupational wage structure within countries, 1989-92 *(cont.)*

Category of country	Number of occupations	SD ln wage		p90/p10	
		All	Decile	All	Decile
Low-income countries (cont.)					
CF – Central African Republic	92	0.78	0.81	6.05	8.45
CI – Côte d'Ivoire	134	0.66	0.67	5.41	6.41
CM – Cameroon	51	0.71	0.74	6.72	6.49
IN – India	92	0.58	0.50	4.51	3.47
ML – Mali	112	0.62	0.59	4.74	4.94
MM – Myanmar	138	0.29	0.27	2.08	2.13
MZ – Mozambique	117	0.45	0.45	3.19	3.48
RW – Rwanda	126	0.70	0.72	6.85	6.96
SD – Sudan	130	0.42	0.35	2.28	2.59
SL – Sierra Leone	102	0.59	0.58	3.17	4.90
SN – Senegal	73	0.44	0.40	2.87	3.01
TD – Chad	91	0.76	0.76	5.66	7.24
TG – Togo	39	0.52	0.50	3.40	3.60
ZM – Zambia	132	0.56	0.53	3.84	4.28
Countries with small populations					
AG – Antigua and Barbuda	64	0.41	0.42	2.70	3.05
AN – Netherlands Antilles	47	0.40	0.37	2.70	2.82
BB – Barbados	97	0.44	0.45	3.11	3.35
BM – Bermuda	46	0.41	0.39	2.10	2.60
BZ – Belize	100	0.53	0.53	3.28	4.17
CY – Cyprus	112	0.44	0.40	2.94	2.85
FK – Falkland Islands (Malvinas)	65	0.38	0.31	2.07	2.31
GI – Gibraltar	32	0.28	0.26	1.90	2.00
IM – Isle of Man	58	0.35	0.33	2.41	2.44
KM – Comoros	76	0.54	0.53	3.75	4.18
LC – St. Lucia	97	0.54	0.56	3.59	4.48
PF – French Polynesia	87	0.46	0.44	3.40	3.22
PM – Saint-Pierre-et-Miquelon	67	0.30	0.32	2.10	2.40
SC – Seychelles	58	0.38	0.39	3.06	2.98
SR – Suriname	57	0.55	0.52	3.89	4.08
VI – Virgin Islands	70	0.40	0.39	3.15	2.93
Communist and ex-communist countries					
BG – Bulgaria	112	0.28	0.25	1.88	1.97
CN – China	82	0.28	0.27	1.90	2.07
CS – Czechoslovakia	110	0.21	0.20	1.61	1.73
CU – Cuba	129	0.31	0.30	2.08	2.24
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	0.35	0.35	2.93	2.65
YU – Yugoslavia	159	0.36	0.36	2.40	2.69

Note. In some tabulations, the countries with small populations or the communist countries are allocated to their appropriate income class.

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



Country key:
AO–Angola; AG–Antigua and Barbuda; AR–Argentina; AT–Austria; AU–Australia; AZ–Azerbaijan; BB–Barbados; BD–Bangladesh; BE–Belgium; BF–Burkina Faso; BG–Bulgaria; BI–Burundi; BJ–Benin; BO–Bolivia; BR–Brazil; BW–Botswana; BY–Belarus; BZ–Belize; CA–Canada; CF–Central African Republic; CI–Côte d'Ivoire; CL–Chile; CM–Cameroon; CN–China; CO–Colombia; CR–Costa Rica; CV–Cape Verde; CY–Cyprus; CZ–Czech Republic; DE–Germany; DK–Denmark; DO–Dominican Republic; DZ–Algeria; EE–Estonia; ER–Eritrea; FI–Finland; FJ–Fiji; GA–Gabon; GB–United Kingdom; GH–Ghana; GQ–Equatorial Guinea; GY–Guyana; HK–Hong Kong; HN–Honduras; HR–Croatia; HU–Hungary; IE–Ireland; IN–India; IR–Iran, Islamic Rep.of; IT–Italy; JP–Japan; KG–Kyrgyzstan; KH–Cambodia; KM–Comoros; KN–Saint Kitts and Nevis; KR–Korea, Republic of; LC–Saint Lucia; LK–Sri Lanka; LT–Lithuania; LU–Luxembourg; LV–Latvia; MD–Moldova, Republic of; MG–Madagascar; ML–Mali; MU–Mauritius; MW–Malawi; MX–Mexico; MZ–Mozambique; NC–New Caledonia; NE–Niger; NG–Nigeria; NI–Nicaragua; NL–Netherlands; NO–Norway; NZ–New Zealand; PE–Peru; PF–French Polynesia; PG–Papua New Guinea; PH–Philippines; PK–Pakistan; PL–Poland; PT–Portugal; RO–Romania; RU–Russian Federation; RW–Rwanda; SE–Sweden; SG–Singapore; SI–Slovenia; SK–Slovakia; SL–Sierra Leone; SN–Senegal; SR–Suriname; SZ–Swaziland; TD–Chad; TG–Togo; TH–Thailand; TN–Tunisia; TR–Turkey; TT–Trinidad and Tobago; UA–Ukraine; UG–Uganda; US–United States; UY–Uruguay; VC–Saint Vincent and the Grenadines; VE–Venezuela; ZM–Zambia; ZR–Zaire.

instance, the United States and the United Kingdom have high levels of inequality for advanced countries – the scatter diagram (figure 1) shows clearly the inverse relation between income per head and the spread of wages among occupations.

Table 5 examines inter-country skill differentials and trends in skill differentials within countries in our sample. Regression 1 records the result of regressing the level of dispersion in occupational wages on ln GDP per capita (constant PPP), on a dummy variable for countries ruled by a communist regime in the year covered by the observation, and on year dummies so that the coefficients simply reflect the cross-sectional variation. The significant coefficient on log of GDP per capita

Table 5. Ordinary Least Squares (OLS) regression estimates of the effect of per capita income and trend on standard deviation of log occupational wages (standard errors in parentheses)

	Cross section	Time series	
	Regression 1	Regression 2	Regression 3
Log GDP per capita	–.104 (0.11)		–.045 (0.33)
Communist (1 = communist years)	–.17 (.02)	–.21 (0.3)	–.18 (.04)
Trend		–.0013 (.0008)	
Country dummies		yes	yes
Year dummies	yes		yes
Constant	yes	yes	yes
Observations	704	831	704
R^2	.440	.873	.794

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.

shows that the pattern in figure 1 is statistically significant: higher levels of GDP reduce occupational pay differences. Regression 2 estimates the time trend in differentials. In these calculations we regress the level of dispersion of pay on country dummies, to subsume cross-section patterns, on the communist dummy (which differs from country dummies because of the transition to market economies) and a linear trend to identify the direction of change over time. The coefficient on trend in regression 2 is negative but small and lacking statistical significance, implying that globally there was little change in occupational wage structures in the 1980s and 1990s. This calculation does not, however, explain how changes in GDP per capita in a given country affect skill differentials. To identify the effect of changes in GDP over time on inequality within a country, we regressed the level of dispersion of pay on country dummies and year dummies and on the log of GDP per capita (constant PPP). The resultant coefficient on GDP per capita in regression 3 is negative, though smaller than in the cross-section regression, and not statistically significant at normal levels. Since this calculation includes controls for both country and time, it implies that within the same country, greater growth of GDP was associated with only slightly reduced inequality of wages. Institutional differences associated with GDP per capita in the cross section, notably democratic governments and trade union strength, may help explain the stronger cross section than time-series result.

Inter-country differences in labour costs

To analyse inter-country differences in the cost of skills, we have deflated the wages for each country-year-occupation by exchange rates for the US dollar. The resulting dollar measure shows great cross-country variation in labour costs. In 1989-92, the average monthly earnings of a carpenter in construction, for example, were US$52 in India, US$2,474 in Sweden and US$223 in Argentina. Even among the advanced countries, there were considerable differences in the cost of labour: over the 1989-92 period, the average monthly earnings of a kindergarten teacher were US$1,775 in Italy and US$1,536 in the United States; while those of a teacher were US$1,256 in Japan and US$2,468 in Germany.

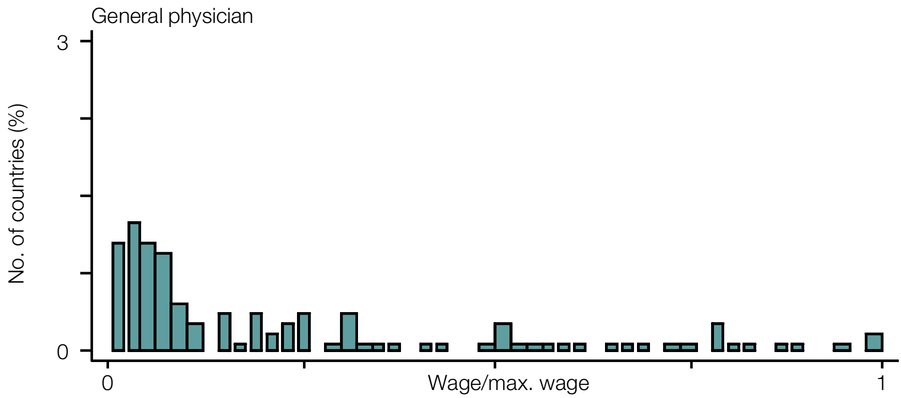
Figure 2 displays the variation in pay across countries for five occupations in the 1983-99 period: a high-wage occupation (general physician); a low-wage occupation (logger); a white-collar occupation (insurance agent); a blue-collar occupation (clicker cutter); and a high-tech occupation (computer 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 which reported wages relative to the highest-paying country in each occupation.

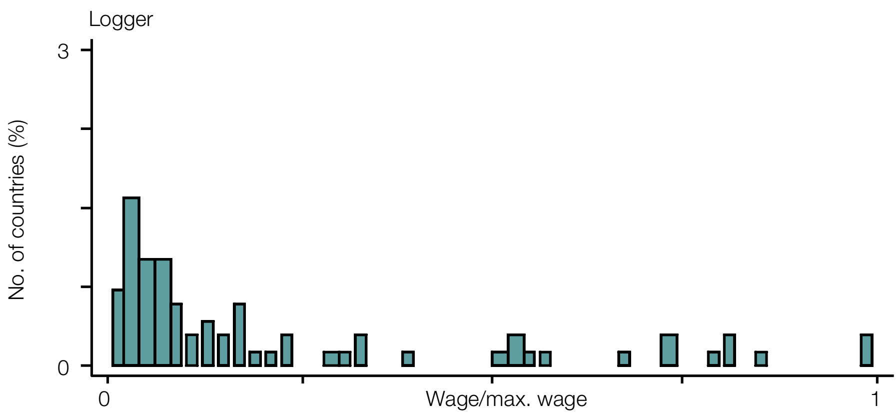
As a summary of the variation in each occupation, we also calculated the ratio of the wage in the country at the median of the distribution relative to the wage in the country with the highest wage. This is not a common measure of dispersion, but it is an easily interpreted statistic: the smaller the median wage relative to the maximum wage, the further from the maximum wage are occupational wages in other countries.[11] The median wage/maximum wage for all five occupations is relatively low – it ranges from .098 to .210 – implying huge differences in the cost of nominally comparable labour across the world. We also calculated the standard deviation of ln pay for each occupation across the reporting countries. The standard deviation is also large compared, say, with the standard deviation of ln wages among occupations within a country.

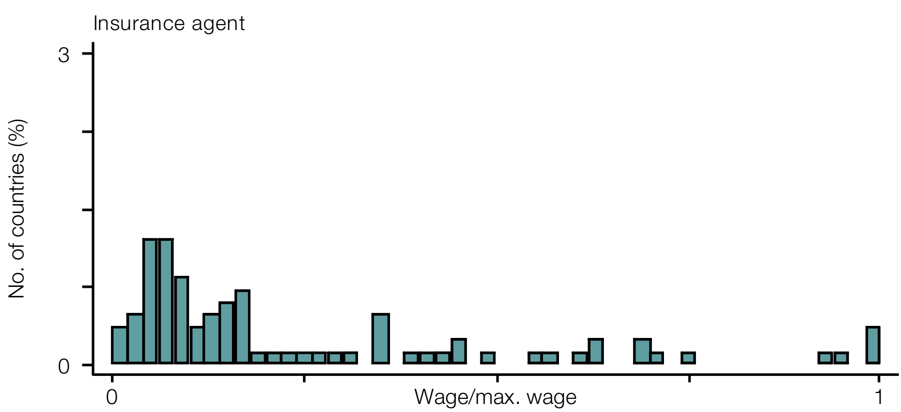
To what extent, if at all, did the economic developments of the 1980s and 1990s, such as increased globalization and the spread of technology, reduce differences in wages across occupations? One way of

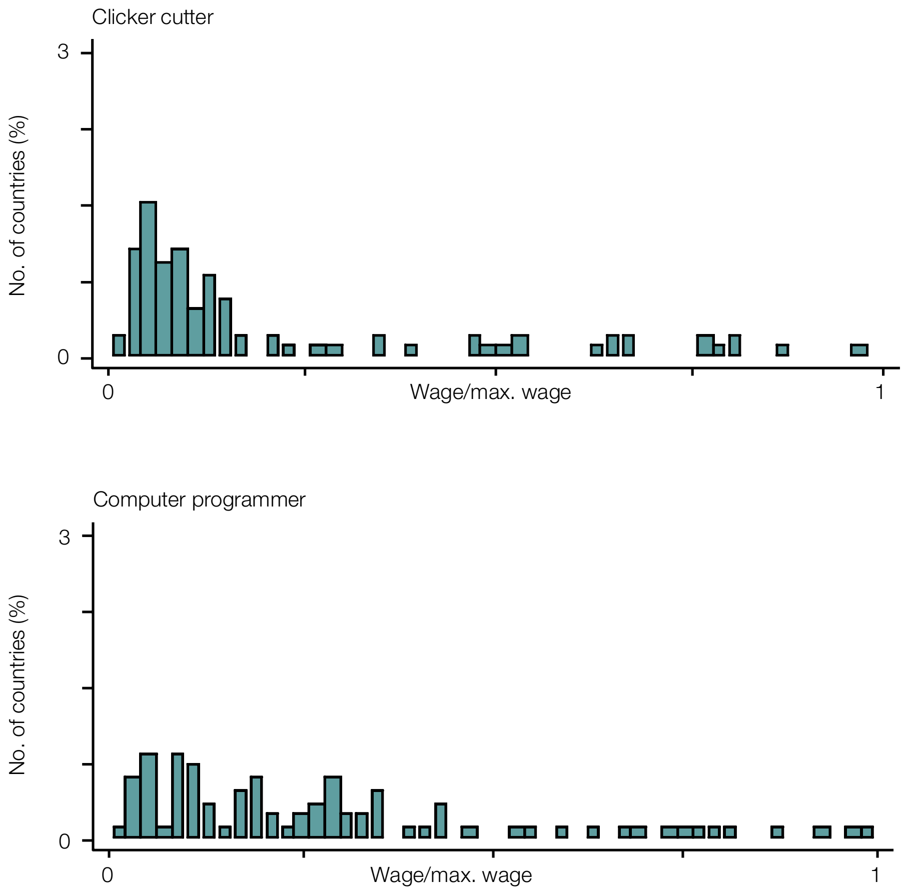
[11] The statistics of extremes can readily be applied to this analysis, as it deals with the relation between extreme values and the central tendency and dispersion of a distribution; see Gumbel (1958).

Figure 2. Distribution of wages in five occupations around the world, 1983-99, in US$









answering this question is to compare measures of the dispersion of occupational pay across countries for the same countries over time. With countries reporting observations in some years and not others, the only way to do this is to group the data into "early" and "late" years. We have done this in two ways. First, we compare the dispersion of occupational pay across countries for 101 occupations in 1983-88 (an early period) and in 1992-99 (a late period). Second, to obtain observations on more occupations (137), we expand the periods to 1983-89 and 1990-99. We then regress the measures of dispersion of pay on a time dummy variable for the most recent period and individual occupation dummy variables, to eliminate cross-occupation variation in dispersion. The coefficients on the dummy variable measure the average change in the inequality over time.

Table 6. Estimates of differences in wages in occupations across countries in two different time periods (using exchange rates)

	1983-88 vs. 1992-99		1983-89 vs. 1990-99	
	SD ln wage	Median/max. wage	SD ln wage	Median/max. wage
Dummy on later period	.130 (.006)	–.023 (.003)	.099 (.007)	–.005 (.004)
Occupational dummies	yes	yes	yes	yes
Constant	yes	yes	yes	yes
R^2	.92	.86	.85	.82
Number of occupations	101	101	137	137

Note. Standard errors are corrected for clustering: with 101 occupations over two periods, we have 202 observations in the first two columns; with 137 occupations over two periods, we have 274 observations in the last two columns.

Table 6 presents estimates of these trend effects.[12] The second and fourth columns record regressions in which the dependent variable is the standard deviation of log wages. A positive coefficient on the trend term implies a rise in inequality of occupational wages across countries. The other two columns record regressions in which the dependent variable is the ratio of the wage of the country in the middle of the distribution of wages for an occupation relative to the wage of the country with the highest wage in the occupation. A negative coefficient on the trend term shows a rise in inequality: the median country wage is further behind the maximum country wage. The regressions in columns 2 and 3 show that for both the median/max. pay and standard deviation of ln pay, variation in wages across countries **rose** over this period. The regressions in columns 4 and 5, with the larger sample of occupations but without the gap in the time periods, show similar results, but in this case while the standard deviation indicator of inequality obtains a highly significant coefficient, the measure of the median country wage to the maximum country wage is not statistically significant, suggesting that it is a less robust measure of inequality than the standard deviation. However, the key result in table 6 is that inequality of wages across countries in the same occupation increased over this period despite globalization, which should have reduced the inequality.

[12] Regression results available on request from the authors. In order to get a reasonable estimate for the median averaging concept, we excluded the median wage observations for Angola. These median pay occupations were all exceptionally highly paid occupations in the years for which data with other averaging concepts were also available in Angola. Identification of the averaging concepts only occurs in country/year pairs for which multiple averaging concepts are reported because of the inclusion of country x year dummies. Inclusion of the Angola median pay occupations led to the implausible result that median wages are significantly (much) higher than average wages.

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 article 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 1999. We developed the OWW by standardizing the diverse data in the ILO's annual October Inquiry into wage rates 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 used can be extended back in time to at least the 1950s, though with less occupational detail.

Our analysis of the OWW shows that, in the period studied, economic development was associated with smaller skill differentials, but the disparity of pay for similar work across countries widened in exchange rate terms.

References

American Federation of Teachers. 1993. *How US teachers measure up internationally: A comparative study of teacher pay, training and conditions of service.* Washington, DC.

Bell, D.; Elias, P. 2000. "Analysis of pay trends: A report for the McCrone Inquiry", in *A teaching profession for the 21st century: The Report of the Committee of Inquiry into Professional Conditions of Service for Teachers.* London, Stationery Office.

Freeman, Richard B.; Oostendorp, Remco H. 2000. *Wages around the world: Pay across occupations and countries.* NBER Working Paper No. 8058. Cambridge, MA, National Bureau of Economic Research, Dec. Available on http://papers.nber.org/papers/w8058 (visited 6 Dec. 2001).

Gumbel, E.J. 1958. *Statistics of extremes.* New York, NY, Columbia University Press.

ILO. 2001. *Statistics on occupational wages and hours of work and on food prices: October Inquiry results, 1999 and 2000.* Special supplement to the *Bulletin of Labour Statistics*. Geneva.

—. Various years. *Yearbook of Labour Statistics.* Geneva.

International Metalworkers Federation (IMF). 2000. *The purchasing power of working time: An international comparison.* Geneva.

National Bureau of Economic Research (NBER). Web site: http://www.nber.org/oww

Union Bank of Switzerland (UBS). 2000. *Prices and earnings around the globe.* Available on http://www.ubs.com/e/index/about/research/publications.html (visited 6 Dec. 2001).

US Department of Education, National Center for Education Statistics. 1988. *International comparisons of teachers' salaries: An exploratory study.* Washington, DC.

Warner, Andrew M. 2001. *International wage determination and globalization.* Paper presented at the NBER Universities Research Conference on Labor in the Global Economy, held in Cambridge, MA.