Working Paper No. 521

Earnings Functions and the Measurement of the Determinants of Wage Dispersion: Extending Oaxaca's Approach*

by

Joseph Deutsch
Bar-Ilan University, Israel

and

Jacques Silber
The Levy Economics Institute and
Bar-Ilan University, Israel

November 2007

An earlier version of this paper was presented at several seminars and conferences. This extended version has been presented in seminars held at the Fundación de Estudios de Economía Aplicada (FEDEA), Madrid; the University of Salamanca; the University Pablo de Olavide, Seville; the University of Barcelona; and the Levy Economics Institute of Bard College, Annandale-on-Hudson, New York.

The authors are grateful to an anonymous referee for their very useful comments and suggestions. Jacques Silber wishes to thank the Fundación de Estudios de Economía Aplicada (FEDEA), Madrid, for its support. Comments may be sent to the authors at Bar-Ilan University, Department of Economics, 52900 Ramat-Gan, Israel, or via e-mail: jdeutsch@mail.biu.ac.il or silberj@mail.biu.ac.il.

The Levy Economics Institute
P.O. Box 5000
Annandale-on-Hudson, NY 12504-5000
http://www.levy.org

ABSTRACT

This paper extends the famous Blinder (1973) and Oaxaca (1973) discrimination in several directions. First, the wage difference breakdown is not limited to two groups. Second, a decomposition technique is proposed that allows analysis of the determinants of the overall wage dispersion. The authors' approach combines two techniques. The first of these is popular in the field of income inequality measurement and concerns the breakdown of inequality by population subgroup. The second technique, very common in the literature of labor economics, uses Mincerian earnings functions to derive a decomposition of wage differences into components measuring group differences in the average values of the explanatory variables, in the coefficients of these variables in the earnings functions, and in the unobservable characteristics. This methodological novelty allows one to determine the exact impact of each of these three elements on the overall wage dispersion, on the dispersion within and between-groups, and on the degree of overlap between the wage distributions of the various groups.

However, this paper goes beyond a static analysis insofar as it succeeds in breaking down the change over time in the overall wage dispersion and its components (both between and within group dispersion and group overlapping) into elements related to changes in the value of the explanatory variables and the coefficients of those variables in the earnings functions, in the unobservable characteristics, and in the relative size of the various groups.

The authors' empirical illustration employs data obtained from income surveys conducted in Israel in 1982, 1990, and 1998, with an emphasis on the comparison between the earnings of new immigrants and those of natives or older immigrants.

Keywords: Blinder, Israel, Mean Difference, Mincerian Earnings Function, Oaxaca, Overlapping, Population Subgroups, Wage Discrimination, Wage Dispersion

JEL Classifications: J31–J71

I. INTRODUCTION

In a pathbreaking paper,[2] Oaxaca (1973) proposed a technique to decompose the relative wage gap between two population subgroups into two components, a first one measuring differences between the groups in human capital characteristics, a second one, labeled "discrimination," taking into account the impact of differences between the groups in the rates of return on these human capital characteristics. Such a distinction is, however, not correct (Polachek 1975; Borjas 2005) because the explained portion can come about because of discrimination while the unexplained portion need not be discrimination.[3]

The main goal of the present study is to extend Oaxaca's approach. While Oaxaca (1973) looked at the determinants of the wage gap between two groups, this paper not only extends the analysis to any number of groups, but also proposes a decomposition technique that permits one to analyze the determinants of the overall wage dispersion. The approach presented here combines two techniques. The first one is popular in the field of income inequality measurement and concerns the breakdown of inequality by population subgroups. The second one, very common in labor economics literature, uses Mincerian earnings functions to derive a decomposition of the average wage difference between two groups into components measuring, respectively, group differences in the average values of the explanatory variables, in the coefficients of these variables in the earnings functions, and in the unobservable characteristics. This methodological novelty allows one to determine the exact impact of each of these three elements on the overall wage dispersion, on the dispersion within and between-groups, and on the degree of overlap between the wage distributions of the various groups.[4]

---

[2] A look at the Social Sciences Citation Index indicates that Oaxaca's paper has been cited more than 500 times. It should, however, be mentioned that in the same year as Oaxaca's paper was published, Blinder (1973) proposed quite a similar framework of analysis.

[3] Further, it is not true that the coefficients of the "Mincerian" earnings function reflect differences in rates of return on human capital. Whereas the schooling coefficient can be interpreted as the average rate of return on human capital, the experience/tenure coefficient indicates both returns and amounts of investment, the constant reflects innate earnings power were there no investment, and coefficients of demographic variables depict differences in earnings potential, assuming similar human capital investment trajectories for each demographic group. We are thankful to an anonymous referee for stressing these points.

[4] In a recent paper, Fields (2003) devised also a new method that can determine how much of the income inequality in an income-generating function is accounted for by each explanatory factor. This is what Fields (2003) called "the levels decomposition." His paper also suggested a way to decompose the difference in

This paper goes beyond a static analysis in so far as it succeeds in breaking down the change over time in the overall wage dispersion and its components (between and within-groups dispersion and group overlapping) into elements related to changes in the value of the explanatory variables and the coefficients of these variables in the earnings functions, in the unobservable characteristics, and in the relative size of the various groups.

The empirical illustration of this paper looks at data obtained from income surveys conducted in Israel in 1982, 1990, and 1998, with special emphasis being put on the comparison between the earnings of new immigrants and those of natives or older immigrants.

The paper is organized as follows. Section II very briefly reviews the literature on the determinants of wage inequality, the causes of the wage gap between natives and immigrants, and the specificity of the immigration to Israel. Section III defines the mean difference of the logarithms of wages, indicates how it may be decomposed into between and within-groups inequality, and an overlapping component. It then explains how these decomposition techniques may be applied to Mincerian earnings functions to determine the respective contributions of the explanatory variables, their coefficients in the earnings functions, and unobservable characteristics to the overall wage dispersion. This decomposition technique is then applied to data from the 1982, 1990, and 1998 income surveys. Section IV extends this breakdown to an analysis of the determinants of the change over time in the overall wage dispersion and an empirical illustration of this additional decomposition based on the same three income surveys is then presented. Concluding comments are finally given in section V.

---

inequality between two groups into specific contributions of the various explanatory factors that were themselves broken down into a coefficient, a correlation, and a standard deviation effect. Fields (2003) called this procedure "the change decomposition." The technique developed in the present study therefore goes, in a way, beyond Fields' (2003) approach, first because it is not limited to an analysis of the between-groups inequality, second because it succeeds at the end in determining the respective contributions of coefficients and explanatory variables, as well as residuals to the overall inequality of the dependent variable. It should be stressed however that because the analysis includes somehow interaction effects (the overlapping component), no attempt is made to look at the specific contribution of the different explanatory variables to the overall dispersion of the dependent variable.

II. ON WAGE INEQUALITY AND IMMIGRATION

A. The Determinants of Wage Inequality

A vast literature has appeared in recent years dealing with the determinants of the increasing wage dispersion that has been observed in several Western countries during the past twenty years. Among the causes of this increasing inequality a distinction has usually been made between factors that affect the demand for labor, those that have an impact on the supply side, and institutional changes that are likely to have also played a role.

There is a general agreement among economists that during the last quarter of the twentieth century there has been an important (positive) shift in the demand for high-skilled labor. The literature has offered two main explanations for this rise in the relative demand for skilled labor. The first argument stresses the role of increased trade openness that has been observed throughout the world during the 1980s and 1990s. This trend towards "globalization" that is usually explained by a decrease in transportation and communication costs and technology transfers implies that goods may be imported at a lower price. Since many of these goods are produced by low-skilled labor, the increased degree of trade openness in developed countries will lead to a weaker demand for unskilled labor and hence, a rise in the relative demand for skilled workers (Freeman 1995; Wood 1995). Another type of explanation has emphasized the role of skill-biased technological change, a distinction being sometimes made between intensive and extensive skilled-biased technological change (Johnson 1997; Krueger 1993).

One may also think of several factors that may affect the inequality of earnings via the supply side. The immigration of low-skilled individuals that has been observed in many Western countries is a first element to be taken into account. In most countries however, the flow of immigrants does not represent an important addition to aggregate labor supply, but the effect on local labor markets may still be important if immigrants tend to stay in specific areas (see Topel 1997 for some illustrations). As a whole, the net effect of immigration seems to be small, an additional reason being that the geographic mobility of natives tends to offset the impact of immigration on local labor markets.

Variations in the size of cohorts are another type of change that may have an effect on the supply side. A baby boom may thus lead, a generation later, to an important increase in the share of young cohorts in the labor force and since younger workers have lower wages than experienced workers, this may lead to an increase in overall inequality (Welch 1979; Berger 1985). Changes in the female labor force participation rates may also play an important role since younger female cohorts have low experience, though they may have a higher level of education. Another modification on the supply side that should indeed be mentioned is the continuous upgrading of the educational composition of the labor force in the Western world, a factor which leads to a decrease in the relative wage of educated workers and hence, probably to a decrease in wage inequality.

Although a demand and supply framework can explain the rise in educational wage differentials by assuming that the rise in the relative demand for educated workers was stronger than that of their supply, some other factors of a more institutional nature should be taken into account. There may be laws that determine the minimum wage or overtime premia and thus, affect wage inequality (Fortin and Lemieux 1997). The extent of collective bargaining or the relative size of the public sector are other institutional elements that may play a role. The study of Goldin and Margo (1992) has clearly shown the impact of institutional change on wage inequality in the United States between 1935 and 1945.

This short survey of the arguments put forth to explain the increase in wage dispersion observed in several Western countries during the past decades indicates that immigration could theoretically be an important factor, acting through the supply side, but that the empirical evidence of a significant impact is not too abundant.

B. The Analysis of Wage Differences between Natives and Immigrants

Following Chiswick's (1978) pioneering work, many studies tried to analyze how immigrants' skills adapted to the host country's labor market. For a long time the consensus was that at the time of their arrival immigrants earn less than natives because they lack the specific skills rewarded in the host country's labor market. However, as these skills are acquired, the human capital stock of immigrants grows relative to that of natives and immigrants experience faster wage growth (Borjas 1994). There may even be a stage where immigrants have accumulated more human capital than natives, the argument being that there is a self-selection process in so far as immigrants are "more able and more highly motivated" than natives (Chiswick 1978). However, one has to take into account the impact of changes in the wage structure since the latter is not likely to have similar effects on natives and immigrants. Thus, in periods where rates of return to skills increase, the relative wage of immigrants may fall, even if their skills remain constant (Levy and Murnane 1992). Most of the studies looking at the earnings of immigrants refer however to countries where the annual flow of immigrants represents a small addition to the existing labor force. The case of Israel is different because at least twice during the past fifty years there have been periods of massive immigration, first during the late 1940s, then during the early 1990s.

C. On Immigration in Israel

In her comprehensive survey of the research conducted on immigration in Israel, Neuman (1999) indicates that between May 1948 and August 1951 the monthly number of immigrants was 15,000–20,000, so that in about three years the Jewish population (which included 649,500 individuals when the State of Israel was created) doubled. Forty years later there was a massive influx of immigrants from the former Soviet Union. Thus, in 1990 there were 199,500 immigrants and 176,000 in 1991. Between the beginning of 1990 and the end of 1998, 879,500 immigrants were added to the Israeli population of 4.56 million, which corresponds to a growth rate of 19.3%. It should be stressed that these immigrants, most of them from the former Soviet Union, had an exceptionally high level of education. More than half of them had academic and managerial positions before immigration. The degree and speed of assimilation in the Israeli labor market of previous immigration waves has been analyzed in several studies (Ofer, Vinokur, and Bar-Chaim 1980; Amir 1993; Beenstock 1993; Friedberg 1995; Chiswick 1997), while Eckstein and Weiss (1997) analyzed the occupational convergence and wage growth of the recent large wave of immigrants from the former Soviet Union. Using panel data they found that upon arrival immigrants receive no significant return on imported human capital, but with more time spent in Israel these returns increase, a gap remains however between the returns received by immigrants and natives. Ultimately immigrants receive the same returns on experience, but convergence is slow as is occupational convergence.

There have also been studies in the 1970s and 1980s looking at the wage differentials between immigrants from various countries, a distinction being usually made between Westerners (immigrants from Europe, America, or Australia) and Easterners (immigrants from North Africa or the Near East). Weiss, Fishelson, and Mark (1979) attributed the decrease in the wage gap between Westerners and Easterners that was observed in the early 1970s to a decrease in human capital differences. Amir (1980) however argued that the decrease in "discrimination" played a larger role.

D. Implications for the Analysis of the Link between Immigration and Wage Dispersion in Israel

The quick survey of the literature that has just been conducted should lead us to the following predictions, assuming individuals are grouped by country or continent of origin. First, given the size of the immigration waves during the past thirty years, we may expect the kind of supply-side effects mentioned earlier to be important. Second, these immigrants came mainly from the former Soviet Union and they had a relatively high level of education. It is nevertheless likely that the skills of these immigrants, who came from a country with a centralized economy, did not fit very well with the requirements of the Israeli labor market. This should imply that if human capital (education or experience) is measured only in years,[5] the rate of return on this capital should be lower for the new immigrants. For all these reasons (high level of human capital among immigrants, but low rate of return on it) one may expect between-groups gaps in wages to be more related to differences in rates of return than in human capital. These gaps are also likely to have grown over time as the share of new immigrants in the Israeli labor force became more important.[6]

It should be remembered, however, that the Israeli economy is a developed economy so that the demand-side effects mentioned earlier and affecting most Western

[5] Unfortunately the information available in the income surveys that were analyzed do not give the exact year of immigration so that it was not possible in the empirical analysis to differentiate between the impacts of the human capital accumulated abroad and in Israel.

countries must have played also a role in Israel, this being particularly true for skilled-biased technological change. Moreover, it is well-known that the Israeli economy was much more open to international trade in the 1980s and 1990s than in earlier periods. Both factors lead us to predict that wage dispersion as a whole must have increased in Israel between the early 1980s and the late 1990s. This trend is also likely to have taken place within population subgroups, if the latter are defined, for example, on the basis of the country or continent of origin, and not as a function of human capital characteristics.

Finally, given that we expect that both the between- and the within-groups dispersion increased over time, we cannot predict a priori what will happen to the degree of overlap between the wage distributions of the various population subgroups.

Despite the relatively important number of studies dealing with the impact of immigration on the Israeli labor market, less emphasis has been given to the effect of immigration on income inequality in general, wage dispersion in particular. The next sections will first provide a methodological framework allowing to better estimate the impact of immigration on wage dispersion, and second, an empirical illustration based on income surveys that were conducted in 1982, 1990, and 1998.

III. THE STATIC ANALYSIS OF THE DETERMINANTS OF THE DISPERSION OF WAGES

A. The Decomposition of the Mean Difference of the Logarithms of Incomes by Population Subgroups

Although the standard deviation is the most popular measure of the dispersion of a distribution, there exists another index of dispersion, called the mean difference (MD), that is related to Gini's famous concentration coefficient (Gini 1912) and is defined (Kendall and Stuart 1989) as

$$MD = (1/n^2) \sum_{i=1 \text{ to } n} \sum_{j=1 \text{ to } n} |y_i - y_j| \qquad (1)$$

[6] On the other hand, over time immigrants acquire more human capital and this should lower the wage gap.

where y_i and y_j are the incomes of individuals i and j, and n is the number of individuals in the population.

Such an index may also be used when the observations are the logarithms of incomes rather than the incomes themselves, in which case the mean difference (that will be denoted here as Δ) will be defined as

$$\Delta = (1/n^2) \sum_{i=1 \text{ to } n} \sum_{j=1 \text{ to } n} | \ln y_i - \ln y_j | \quad (2)$$

Expression (2) indicates in fact that Δ measures the expected income gap (in percentage terms) between two individuals drawn (with repetition) from the sample of individuals on whom information on their income was collected.

Let now m represent the number of population subgroups. Expression (2) may then be decomposed into the sum of two terms, Δ_A and Δ_W, where Δ_A refers to what may be called the "across-groups inequality" (Dagum 1960, 1997), while Δ_W measures the "within-groups inequality," with

$$\Delta_W = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{k=h} \sum_{j\in k} |\ln y_{ih} - \ln y_{jk} | \quad (3)$$

and

$$\Delta_A = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{k\neq h} \sum_{j\in k} |\ln y_{ih} - \ln y_{jk} | \quad (4)$$

the second subindex (h or k) in (3) and (4) referring to the group to which the individual belongs.

Let us now assume that the groups are ranked by decreasing values of the average of the logarithms of income in each group so that $\ln y_{gh}$, the mean logarithm of incomes[7] in group h, is higher than $\ln y_{g,h+1}$, the mean logarithm of incomes in group (h+1). Expression (4) may then be written as

[7] y_{gh} is evidently the geometric mean of the incomes in group h.

$$\Delta_A = \Delta_d + \Delta_p \qquad (5)$$

with

$$\Delta_d = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i \in h} \sum_{k \neq h} \sum_{j \in k} (\ln y_{ih} - \ln y_{jk}) \text{ with } \ln y_{ih} \geq \ln y_{jk} \qquad (6)$$

and

$$\Delta_p = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i \in h} \sum_{k \neq h} \sum_{j \in k} (\ln y_{jk} - \ln y_{ih}) \text{ with } \ln y_{ih} < \ln y_{jk} \qquad (7)$$

Combining (6) and (7) we derive that

$$\Delta_d - \Delta_p = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i \in h} \sum_{k \neq h} \sum_{j \in k} (\ln y_{ih} - \ln y_{jk}) \qquad (8)$$

$$\leftrightarrow \Delta_d - \Delta_p = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{k \neq h} [\sum_{i \in h} (n_k (\ln y_{ih})) - \sum_{j \in k} (n_h (\ln y_{jk}))] \qquad (9)$$

$$\leftrightarrow \Delta_d - \Delta_p = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{k \neq h} [n_k n_h (\ln y_{gh} - \ln y_{gk})] \qquad (10)$$

where n_h and n_k represent, respectively, the number of individuals in groups h and k.

Since the between-groups mean difference Δ_B is obtained by giving each individual the average value of the logarithms of the incomes of the group to which he belongs, we may, using (2), define an index Δ_B as

$$\Delta_B = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{k \neq h \text{ to } m} n_h n_k \, |\ln y_{gh} - \ln y_{gk}| \qquad (11)$$

and it may be observed when comparing (10) and (11), that

$$\Delta_B = (\Delta_d - \Delta_p) \qquad (12)$$

Since expression (5) indicates that

$$\Delta_A = (\Delta_d - \Delta_p) + (2\,\Delta_p) \tag{13}$$

we conclude, using (1), (2), (3), (4), (5), (12), and (13), that

$$\Delta = \Delta_w + \Delta_B + (2\,\Delta_p) \tag{14}$$

One should note that expression (7) indicates that ($2\Delta_p$), the residual that is obtained in the traditional decomposition of the mean difference by population subgroups, is expressed as a simple function of the "transvariations"[8] which exist between all pairs of population subgroups.

In the next section it will be shown that it is also possible to compute the contribution of each population subgroup to the value of the three components of the overall wage dispersion that have just been derived.

B. Computing the Contribution of Each Population Group to the Various Components of the Breakdown

The results that have been derived in equations (3) to (14) may be used to determine the contribution of each population group to the three components of the breakdown that have been defined.

The contribution C_{wh} of group h to the within-groups inequality Δ_w may thus be expressed, using (3), as

$$C_{wh} = (1/n^2)\sum_{i\in h}\sum_{j\in h} | \ln y_{ih} - \ln y_{jk} | \tag{15}$$

[8] Following Gini (1959) we may say that there exists a "Transvariation" between two distributions $\{x_i\}$ and $\{y_j\}$ with respect to their (arithmetic, geometric,..) means m_x and m_y when among the $n_x n_y$ possible differences ($x_i - y_j$), the sign of at least one of them is different from that of the expression ($m_x - m_y$), n_x and n_y being the number of observations in these two distributions. The importance of such a "Transvariation" may be measured in several ways (Deutsch and Silber 1998). The reference here is to the moment μ_1 of order 1 which is defined as: $\mu_1 = \int_{-\infty}^{+\infty} g(y)\,dy \int_{-\infty}^{y} (y-x)\,f(x)\,dx$, where g(y) and f(x) are the densities of the variables y and x.

Similarly, using (6), (7), and (12), the contribution C_{bh} of group h to the between inequality Δ_B may be written as

$$C_{bh} = (1/n^2) \sum_{k\neq h} \sum_{i\in h} \sum_{j\in k} (\ln y_{ih} - \ln y_{jk}) \text{ with } \ln y_{ih} > \ln y_{jk}$$
$$- (1/n^2) \sum_{k\neq h} \sum_{i\in h} \sum_{j\in k} (\ln y_{jk} - \ln y_{ih}) \text{ with } \ln y_{ih} < \ln y_{jk} \quad (16)$$

Finally, the contribution C_{ph} of group h to the overlapping term ($2\,\Delta_p$) may be written, using (7) and (14), as

$$C_{ph} = (1/n^2) \sum_{k\neq h} \sum_{i\in h} \sum_{j\in k} (\ln y_{jk} - \ln y_{ih}) \text{ with } \ln y_{ih} < \ln y_{jk} \quad (17)$$

Naturally the contribution C_h of group h to the overall inequality term Δ in (14) will be expressed as

$$C_h = C_{wh} + C_{bh} + C_{ph} \quad (18)$$

and it can be easily proven that

$$\Delta = \sum_{h=1 \text{ to } m} C_h. \quad (19)$$

In the next section, the various decompositions that have been previously defined will be combined with Oaxaca's (1973) traditional breakdown of income differences into a "variable" and a "coefficient" component.[9] This will allow us to analyze the impact of these two elements (the values of the variables and that of the regression coefficients) not only on the difference between the average logarithms of incomes in two population subgroups, but also on the dispersion of these (logarithms of) incomes in each group and on the degree of overlapping between two distributions of (the logarithms of) incomes.

[9] It may be observed that Oaxaca's (1973) decomposition corresponds to what was defined earlier as the between-groups mean difference (Δ_B) as it is defined in (11), for the specific case where there are only two groups k and h, assuming their size n_h and n_k are equal. (Oaxaca's decomposition is in fact then equal to half the mean difference Δ_B).

We will see that in some of these decompositions there will be another component that will represent the impact of unobservable characteristics.

C. Estimating the Contributions of the Variables, their Coefficients in the Regressions, and of the Unobservable Characteristics to the Overall Wage Dispersion

The human capital model relates the earnings of an individual to the amount of his investment in human capital and, using the framework of analysis originally put forth by Mincer (1974), we may write that

$$\ln y_{ih} = \sum_{l=1 \text{ to } L} \beta_{lh}\, x_{lih} + u_{ih} \tag{20}$$

where the subscripts i, h, and l refer, respectively, to the individual (i), the group to which he belongs (h), and the explanatory variable (l). The coefficient β_{lh} is the regression coefficient corresponding to the explanatory variable x_{lih}, which refers to the characteristic l of individual i who belongs to group h. Finally u_{ih},which is the residual of the regression, refers evidently to factors that have not been taken into account.

Combining now expressions (3) and (20), we derive first the within-groups inequality

$$\begin{aligned}\Delta_W &= (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{k=h} \sum_{j\in k} \left|(\sum_{l=1 \text{ to } L} \beta_{lh}\, x_{lih} + u_{ih}) - (\sum_{l=1 \text{ to } L} \beta_{lk}\, x_{ljk} + u_{jk})\right| \\ &= (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{j\in h} \left|(\sum_{l=1 \text{ to } L} (\beta_{lh}\, x_{lih} - \beta_{lh}\, x_{ljh}) + (u_{ih} - u_{jh})\right| \\ &= (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{j\in h} 2\, [(\sum_{l=1 \text{ to } L} (\beta_{lh}\, x_{lih} - \beta_{lh}\, x_{ljh}) + (u_{ih} - u_{jh})]\end{aligned}$$

with $\ln y_{ih} > \ln y_{jh}$, so that finally

$$\Delta_W = A + B \tag{21}$$

where

$$A = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{j\in h} 2\, [(\sum_{l=1 \text{ to } L} \beta_{lh} (x_{lih} - x_{ljh})] \text{ with } \ln y_{ih} > \ln y_{jh} \qquad (22)$$

and

$$B = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{j\in h} 2\, (u_{ih} - u_{jh}) \text{ ,with } \ln y_{ih} > \ln y_{jh} \qquad (23)$$

Expressions (21) through (23) indicate clearly that the within-groups inequality of (the logarithms of) incomes is the sum of two elements: a first one (A) that is the consequence of differences between individuals belonging to the same group in values taken by their measured characteristics and a second one (B) that derives from differences between individuals belonging to the same group in unmeasured characteristics.

We may in a similar way derive an expression for the between-groups inequality Δ_B since by combining (11) and (20) we may write that

$$\begin{aligned}\Delta_B &= (1/n^2) \sum_{h=1 \text{ to } m} \sum_{k=1 \text{ to } m} n_h n_k \left| \sum_{l=1 \text{ to } L} \beta_{lh} x_{lgh} - \sum_{l=1 \text{ to } L} \beta_{lk} x_{lgk} \right| \\ &= (1/n^2) \sum_{h=1 \text{ to } m} \sum_{k=1 \text{ to } m} n_h n_k \left| \sum_{l=1 \text{ to } L} (\beta_{lh} x_{lgh} - \beta_{lk} x_{lgk}) \right| \\ &= (1/n^2)\, 2\, [\sum_{h=1 \text{ to } m} \sum_{k=1 \text{ to } m} n_h n_k [\sum_{l=1 \text{ to } L} (\beta_{lh} x_{lgh} - \beta_{lk} x_{lgk})]] \text{ with } \ln y_{gh} > \ln y_{gk} \qquad (24)\end{aligned}$$

where x_{lgh} and x_{lgk} are the arithmetic means of characteristic l in groups h and k, respectively.

Remembering then that (ab-cd)=((a+c)/2)(b-d) + ((b+d)/2)(a-c), we finally derive

$$\Delta_B = C + D \qquad (25)$$

where

$$C = (1/n^2) [\sum_{h=1 \text{ to } m} \sum_{k=1 \text{ to } m} n_h n_k [\sum_{l=1 \text{ to } L} (\beta_{lh} + \beta_{lk}) (x_{lgh} - x_{lgk})]] \qquad (26)$$

with $\ln y_{gh} > \ln y_{gk}$, and

$$D = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{k=1 \text{ to } m} n_h n_k [\sum_{l=1 \text{ to } L} (x_{lgh} + x_{lgk}) (\beta_{lh} - \beta_{lk})]] \quad (27)$$

with $\ln y_{gh} > \ln y_{gk}$

Expressions (25) through (27) indicate that the between-groups inequality of (the logarithms of) incomes is the sum of two elements: a first one (C) that is the consequence of differences between the groups in the average levels of the explanatory variables and a second one (D) that is explained by differences between the groups in the coefficients of these variables in the regressions.[10] Note that in (26) and (27) no reference is made to unmeasured characteristics since in measuring the between-groups inequality we assume that each individual in a group receives the average (logarithm of the) income of the group and that, by the definition of a regression, this average does not include a residual.

The third element of the decomposition of inequality of (the logarithms of) incomes measures the degree of overlap between the distributions of the various groups and, combining (7) and (20), Δ_p may be expressed as

$$\Delta_p = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i \in h} \sum_{k \neq h} \sum_{j \in k} [(\sum_{l=1 \text{ to } L} \beta_{lk} x_{ljk} + u_{jk}) - (\sum_{l=1 \text{ to } L} \beta_{lh} x_{lih} + u_{ih})] \quad (28)$$

with $\ln y_{ih} < \ln y_{jk}$

Using similar decomposition rules as before, we derive that

$$\Delta_p = E + F + G \quad (29)$$

where

$$E = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i \in h} \sum_{k \neq h} \sum_{j \in k} [\sum_{l=1 \text{ to } L} ((\beta_{lk} + \beta_{lh})/2) (x_{ljk} - x_{lih})] \quad (30)$$

with $\ln y_{ih} < \ln y_{jk}$

[10] The decomposition given in expressions (25) to (27) corresponds to that proposed by Reimers (1983).

$$F = (1/n^2) \sum_{h=1 \text{ to } m} \sum_{i \in h} \sum_{k \neq h} \sum_{j \in k} [\sum_{l=1 \text{ to } L} ((x_{ljk} + x_{lih})/2)(\beta_{lk} - \beta_{lh})] \quad (31)$$

with $\ln y_{ih} < \ln y_{jk}$

and

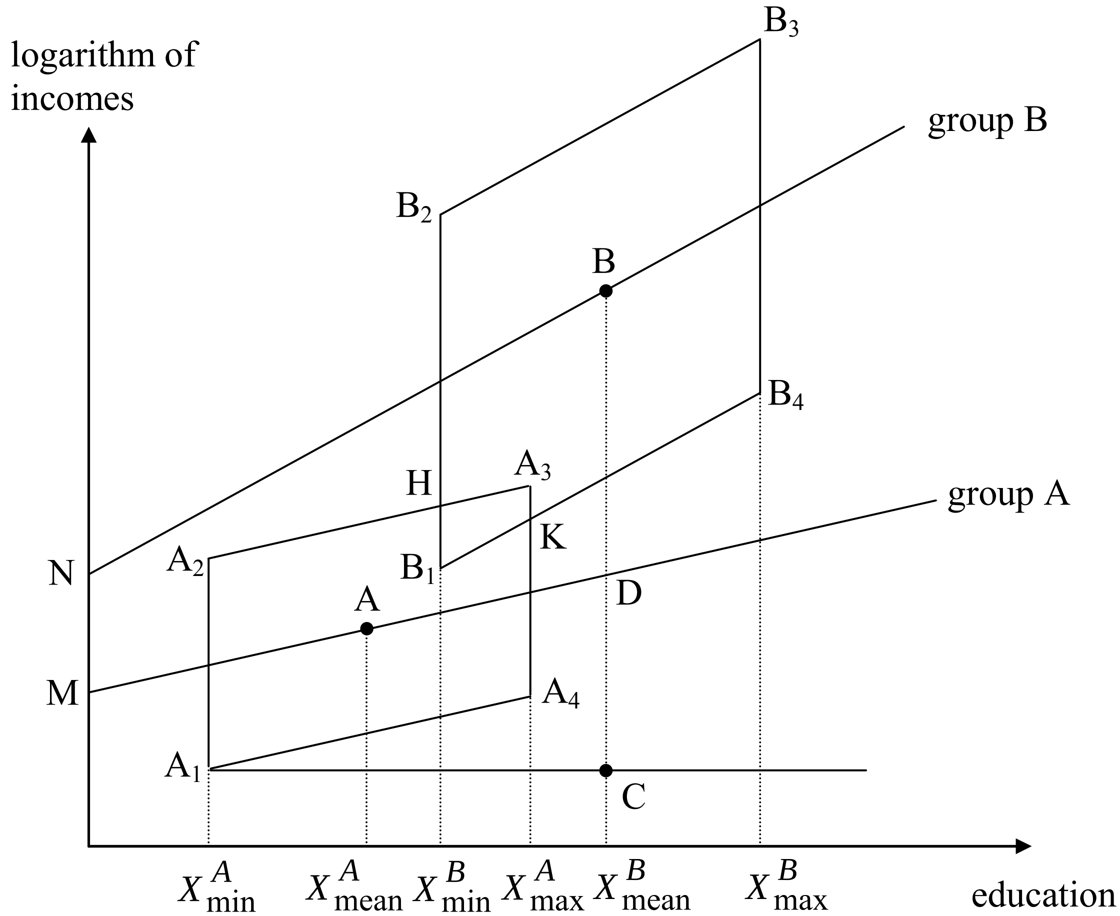
$$G = 1/n^2) \sum_{h=1 \text{ to } m} \sum_{i \in h} \sum_{k \neq h} \sum_{j \in k} [(u_{jk} - u_{ih})] \quad (32)$$

with $\ln y_{ih} < \ln y_{jk}$

Expressions (29) through (32) indicate that the degree of overlapping between the distributions corresponding to the various groups is a function of three elements: a first component (E) that reflects differences in the values taken by the explanatory variables among the individuals affected by the overlapping, a second element (F) that is explained by differences between the groups in the regression coefficients corresponding to these variables, and a third expression (G) that is due to unmeasured characteristics among the individuals affected by the overlapping.

Combining expressions (21), (25), and (29) we conclude that (A + C + E), (D + F), and (B + G) represent, respectively, the contributions of differences in the values of the explanatory variables, in the regression coefficients corresponding to these variables, and in unobservable characteristics to the overall wage dispersion.

FIGURE 1



Such a decomposition may be given a graphical interpretation that extends the usual Blinder-Oaxaca diagram. To simplify, we limit the analysis to two population subgroups and one explanatory variable, say, education. In Figure 1 the straight lines MA and NB refer, respeectively, to the earnings functions (regression lines) of the two groups A and B. Let $x_{mean,A}$ and $x_{mean,B}$ represent the mean values (mean educational level and logarithm of earnings) of groups A and B. In reality there is, however, a dispersion of educational levels (on the horizontal axis) in each group (between $x_{B,min}$ and $x_{B,max}$ for group B, and between $x_{A,min}$ and $x_{A,max}$ for group A). There is also, at the various educational levels and for each group, a dispersion of unobservable characteristics (e.g., innate ability) on the vertical axis. Assume, for simplicity, that all the observations for group B are located in the area B_1 B_2 B_3 B_4 and for group A in the area A_1 A_2 A_3 A_4. We have assumed that these two areas overlap (area B_1 H A_3 K). The between-groups dispersion is (for example) decomposed into the two elements CD and DB as in the traditional Blinder-Oaxaca diagram, and this dispersion clearly depends only on the regression coefficients and on the average value of the educational levels of the two groups. The within-groups dispersion corresponds to the two areas B_1 B_2 B_3 B_4 and A_1 A_2 A_3 A_4 and, as can be seen in the graph, this dispersion is due to within-groups differences in educational levels and in the unobservable characteristics. Finally, the overlapping component, represented by the area B_1 H A_3 K clearly depends on differences in the slopes of the lines MA and NB (differences in regression coefficients), in the educational levels, and in the value of the unobservable characteristics.

We now turn to the results of the empirical investigation.

D. Male Earnings Functions in Israel in 1982, 1990, and 1998

The empirical illustration that is presented in this section is based on the income surveys that are conducted each year in Israel. We have chosen to limit our analysis to three surveys: those of the years 1982, 1990, and 1998. Since one of the aims of this research is to look at the impact on earnings of the country of origin and of the period of immigration of the immigrants, we have limited our analysis to the Jewish male population and divided it in four groups: those born in Israel (group IL), those born in Asia or Africa (group AA), and those born in Europe or America. However, in order to take into account what happened to the most recent immigrants, we have divided the last group into two subgroups: those who immigrated to Israel before 1972 (group EA) and those who immigrated after 1971 (group NIM). It is clear that, specially for the last two surveys analyzed (1990 and 1998), most of the members of the last subgroup came from the former Soviet Union so that we will be able to focus on the earnings of this important population of immigrants.

Let us first take a look at the general characteristics of the population analyzed.

The two first columns of Tables 1-A to 1-C give for each year the means and standard deviations of the various variables that have been introduced in the regressions. The results are given each time for the whole sample.[11] It appears that the proportion of married individuals declined over the years from 87% in 1982 to 75% in 1998. The proportion of singles, on the contrary, increased during the same period from 10.7% to 20.9%. The other categories of marital status (divorced, widows, or separated) increased slightly from 2.3% in 1982 to 4.1% in 1998. The average number of years of schooling increased from 10.7 in 1982 to 12.6 in 1998, while the average number of years of experience correspondingly decreased from 26.2 years in 1982 to 21.7 in 1998. The proportion of the males having attended a Talmudic school (Yeshiva), a factor likely to have a downward effect on earnings, decreased from 2.6% in 1982 to 1.5% in 1998.

It is difficult to compare the means obtained for the various population subgroups since the younger people are more likely to be individuals born in Israel so that one would expect to observe, for these men born in Israel, a smaller proportion of married individuals and less years of schooling or experience. This is actually the case for each year. Such differences do not however prevent us from comparing regression results because then the age is kept constant (since we have defined experience in the traditional way, that is, as age minus six minus the number of years of schooling).[12]

The last two columns of Tables 1-A to 1-C give, for the whole sample, the results of the Mincerian earnings functions that have been estimated for each of the three periods analyzed. It appears, for example, that the coefficient of schooling increased throughout the period, being equal to 6.8% in 1982, 7.5% in 1990, and 8.9% in 1998. The coefficient of the experience variable[13] at the beginning of the career showed a different pattern since it rose form 2.7% to 3.9% between 1982 and 1990, but was equal to 2.9% in 1998. Individuals who were married earned, on average, 20.8% more than those who were divorced, separated, or widows in 1982, 10.7% more in 1990, and 15.5% more in 1998. Single individuals, on the contrary, earned 10.7% in 1982, 10.1% in 1990, and 4.1% in

[11] The corresponding statistics for each of the four subgroups that have been distinguished (IL, AA, EA, and NIM) and for each of the years 1982, 1990, and 1998 are given in Appendix 1.

[12] This implies that the years spent in the army are considered as part of the working experience.

[13] Since this study is based on cross sections, it is in fact impossible to make a distinction between the impact of experience and that of the business cycle, and one has to be careful in interpreting some of the changes observed, for example, between 1982 and 1990.

1998 less than those who were neither married nor singles. These data indicate therefore that the gap between married and single men decreased significantly between 1982 and 1998.

Similar regressions have been estimated for each of the four population subgroups that have been distinguished and are presented in Appendix 1. It appears, for example, that in 1982 the coefficient of schooling was much higher for those born in Israel (8.8%) than for those born in Asia or Africa (4.9%), or Europe or America (6.0% for those who came before 1972 and 5.0% for those who came after 1971). Similarly, in 1998 the coefficient of schooling was 12.5% for those born in Israel, 7.3% for those born in Asia or Africa, 11.0% for those born in Europe who immigrated before 1972, and 6.5% for those born in Europe who arrived in Israel after 1971.

E. The Components of the Overall Wage Dispersion

Table 2 gives for each of the three years (1982, 1990, and 1998) the decomposition of the overall wage dispersion into the three components mentioned in Section III: the between- and within-groups dispersions, and the overlapping term. The number that appears in the line labeled "Total" gives for each year the overall mean difference of the logarithms of income, that is the expected income difference in percentage terms between two individuals chosen (with repetition) in the sample. Whereas this mean difference only slightly increased between 1982 to 1990 (from 63.6% to 64.4%), the change was very important between 1990 and 1998 since the mean difference reached the value of 73.0% in 1998. What are the reasons for such an important increase in the overall dispersion observed during the decade 1990–1998? This is a period where in several Western countries wage dispersion increased for reasons related to technological change, increasing openness to trade, and institutional change such as the weakening of the trade unions (see the short survey of the literature in Section II). It should however be remembered that during the 1990–1998 period, 880,000 individuals immigrated to Israel, mostly from the former Soviet Union. Since one of the population subgroups includes only those who migrated from Europe or America after 1971, the analysis presented in this section enables one to determine the impact of this immigration on the overall wage dispersion. However, the decomposition techniques presented previously also give the specific impact on the overall dispersion of incomes and on its three components (the between- and within-groups dispersion, and the overlapping element) of the explanatory variables, their coefficients, and of the unobserved characteristics. All these results will now be presented and analyzed.

1. The Relative Importance of the Between- and Within-Groups Dispersion, and the Contribution of the Overlapping Component

Table 2 indicates that in absolute terms the contribution of the between-groups dispersion to the overall dispersion decreased from 11.5% to 6.4% between 1982 and 1990, but it increased between 1990 and 1998 to reach 14.7% in 1998. In percentage terms the picture is similar since the contribution of the between-groups dispersion decreased from 18.1% to 10.0% between 1982 and 1990, but was equal to 20.1% in 1998.

The within-groups dispersion increased in absolute terms during both subperiods. It was equal (in absolute terms) to 17.9% in 1982, 21.7% in 1990, and 26.5% in 1998. The picture is quite similar if one looks at the relative contribution of the within-groups to the overall wage dispersion since this contribution rose from 28.1% in 1982 to 33.8% in 1990 and 36.3% in 1998.

For the overlapping term the pattern is as follows: in absolute terms it increased from 34.2% in 1982 to 36.2% in 1990, but fell down to a level of 31.8% in 1998. In relative terms the contribution of the overlapping term rose form 53.8% in 1982 to 56.2% in 1990, and fell back to 43.6% in 1998.

The picture during the 1982–1990 period is, hence, very different from the one observed during the years 1990–1998. During the first period, the between-groups dispersion decreased while the within-groups dispersion rose, the overlapping term increasing only slightly. These conclusions are true in absolute and relative terms. During the second subperiod, on the contrary, the between- as well as the within-groups dispersion rose while the amount of overlapping decreased, this being again true in absolute and relative terms. Two factors at least may explain these patterns. First, there was at that time an increase in wage dispersion in several Western countries and this is probably also true for the within-groups dispersion. At the same time there was a specific Israeli story: the massive immigration of Jews from the former Soviet Union had increased, at least in a first stage, the degree of stratification in the Israeli society, leading thus to an increase in the between-groups dispersion. This latter effect was more important than the increase in the within-groups dispersion that was just mentioned, since the degree of overlapping decreased during this period.

To better understand these changes we now take a look at the respective role played by the explanatory variables, their coefficients, and by the unobserved characteristics.

2. The Contribution of the Explanatory Variables, their Coefficient,s and of the Unobserved Characteristics to the Wage Dispersion

Table 3 indicates that in 1982, out of a total wage dispersion of 63.6%, the explanatory variables contributed in absolute terms 16.8%, their coefficients 2.7%, and the unobserved characteristics 44.1%. The corresponding figures for 1990 when the overall dispersion was 64.4%, were 19.9%, 0.3%, and 44.2%. In 1998 the mean difference of the logarithms of wages was equal to 73.0% while the three contributions previously mentioned were, respectively, equal to 21.4%, 4.8%, and 46.7%. It appears therefore that over time the contribution of the explanatory variables increased in absolute value. The contribution of unobserved characteristics, on the contrary, did not vary much over time, while that of the regression coefficients was low and unstable.

The figures are somehow different in percentage terms (see again Table 3). It appears that over time there was also an increase in percentage terms in the contribution of the explanatory variables, at least between 1982 where it was equal to 26.4% and 1990 when it reached 30.9%. There was no important change during the 1990–1998 period. The relative contribution of unobserved characteristics decreased over time, mainly during the second subperiod (from 69.4% in 1982, to 68.6% in 1990, and 64.0% in 1998). Finally, the relative contribution of subgroup differences in the regression coefficients varied over time since it was equal to 4.2% in 1982, 0.5% in 1990, and 6.6% in 1998.

A similar analysis may be conducted at the level of each of the three components of the overall wage dispersion: the between- and within-groups dispersion, and the overlapping term. The results are presented in Table 4. For the between-groups dispersion, as was mentioned previously, only the explanatory variables and their coefficients play a role. It appears that the relative role of the explanatory variables varied strongly over time: it was equal to 35.4% in 1982, 57.0% in 1990, but only 11.8% in 1998. The picture is evidently the opposite for the relative contribution of the regression coefficients. The very important role of the latter in 1998 indicates, for example, that, ceteris paribus, the coefficient of the schooling variable is much lower among new immigrants.[14]

For the within-groups dispersion, only two factors play a role: the explanatory variables and the unobserved characteristics. Table 4 indicates that, in relative terms, the contribution of the explanatory variables steadily rose over time, from 27.6% in 1982, to 32.3% in 1990, and 36.0% in 1998. The trend, in relative terms, is evidently opposite for unobservable characteristics.

For the overlapping component, as was mentioned previously, each of the three factors (the explanatory variables, the unobserved characteristics, and the regression coefficients) plays a role. It is first interesting to note that the data indicate that the component measuring the impact of the regression coefficients had for each year a negative contribution to the degree of overlap. This implies that if there had been no differences between the individuals involved in the overlap in the value of the explanatory variables (so that the sum of all the binary comparisons of measured characteristics, as it is given in (30), would have been assumed to be nil) or in their unobservable characteristics [so that the sum of all the binary comparisons of unobserved characteristics, as it is given in (32) would also have been nil], the between-groups differences in the regression coefficients would have led to a smaller amount of overlap.

As far as the two other components are concerned, it appears that the relative importance of the explanatory variables increased over time (from 22.8% in 1982, to 25.3% in 1990, and 31.9% in 1998). The relative contribution of unobservable characteristics, on the contrary, was rather unstable (91.1% in 1982, 81.5% in 1990, and 93.4% in 1998).

3. Summarizing the Empirical Results

The various observations that have just been made could be summarized as follows.

[14] This is confirmed by the results of the regressions run separately for each of the four population subgroups which are given in Appendix 1.

First, during the two subperiods that have been analyzed, the between-groups dispersion first decreased, then increased; since the same pattern has been observed (in percentage terms) for the component reflecting the regression coefficients and given that this component contributes most to this dispersion, we may fairly assume that the regression coefficients played a central role here.

Second, the within-groups dispersion increased in both periods, a pattern that is observed also (in percentage terms) for the component corresponding to the explanatory variables. Although this component never represents more than a third of the within-groups dispersion, it is likely that its variation over time explains the increasing importance of the within-groups dispersion.

Third, the overlapping component first increased, then decreased. This is also the pattern observed for the regression coefficients, although their contribution remains negative throughout the period. We may therefore conjecture that the story of the overlap is mainly that of the regression coefficients and if the share of the overlap in the overall dispersion decreased drastically between 1990 and 1998, it seems to be a consequence of the fact that the sharp decrease in the regression coefficients observed among new immigrants led to a reduction in the amount of overlap between the income distributions of the four population subgroups.

In the next section we extend the analysis and show how it is also possible to decompose changes over time in the dispersion of wages and in its components.

TABLE 1-A

Descriptive Statistics and Regression Results for the 1982 Income Survey (whole population)

Variable	Mean	Standard Deviation	Regression Coefficients	t- values
Logarithm of Wage per Hour	4.0184	0.5723		
Married	0.8672	0.3394	0.2077	3.46
Single	0.1068	0.3088	-0.1079	-1.55
Years of Schooling	10.6754	3.7835	0.0681	24.03
Years of Experience	26.2035	14.7535	0.0274	9.66
Square of Years of Experience	904.2872	880.5513	-0.0004	-8.57
Attended Talmudic School	0.0257	0.1582	-0.3546	-5.85
Intercept			2.7680	33.76
R^2	0.2564			
Number of Observations	2725			

TABLE 1-B

Descriptive Statistics and Regression Results for the 1990 Income Survey (whole population)

Variable	Mean	Standard Deviation	Regression Coefficients	t-values
Logarithm of Wage per Hour	2.5072	0.5742		
Married	0.8140	0.3891	0.1070	2.20
Single	0.1523	0.3593	-0.1009	-1.78
Years of Schooling	11.6693	3.2367	0.0749	24.83
Years of Experience	23.3972	13.8600	0.0385	14.52
Square of Years of Experience	739.5288	787.2934	-0.0005	-12.07
Attended Talmudic School	0.0202	0.1408	-0.2695	-4.30
Intercept			1.0609	14.71
R^2	0.2810			
Number of Observations	3113			

TABLE 1-C

Descriptive Statistics and Regression Results for the 1998 Income Survey (whole population)

Variable	Mean	Standard Deviation	Regression Coefficients	t-values
Logarithm of Wage per Hour	3.4487	0.6553		
Married	0.7460	0.4353	0.1554	4.29
Single	0.2109	0.4080	-0.0415	-0.97
Years of Schooling	12.6383	2.8172	0.0886	32.23
Years of Experience	21.6922	13.0829	0.0289	12.59
Square of Years of Experience	641.7122	679.5137	-0.0004	-9.33
Attended Talmudic School	0.0153	0.1229	-0.1567	-2.63
Intercept			1.8412	31.01
R^2	0.2360			
Number of Observations	6197			

TABLE 2

Decomposition of the Wage Dispersion into a Between-Groups Dispersion, a Within-Groups Dispersion, and an Overlapping Component

Actual Results	**1982**	**1990**	**1998**
Component			
Between-Groups	11.5	6.46	14.65
Within-Groups	17.89	21.74	26.48
Overlap	34.19	36.2	31.85
Total	63.58	64.4	72.98

In Percentage Terms	**1982**	**1990**	**1998**
Component			
Between-Groups	18.08	10.02	20.07
Within-Groups	28.14	33.76	36.29
Overlap	53.78	56.21	43.64
Total	100.00	100.00	100.00

TABLE 3

Decomposition of the Wage Dispersion into Components Corresponding to the Explanatory Variables, their Coefficients, and to the Unobservable Characteristics.

Actual Results	**1982**	**1990**	**1998**
Component			
Explanatory Vaiables	16.80	19.87	21.44
Regression Coefficients	2.69	0.33	4.83
Unobservable Characteristics	44.10	44.20	46.72
Total	63.58	64.40	72.98

In Percentage Terms	**1982**	**1990**	**1998**
Component			
Explanatory Vaiables	26.42	30.85	29.37
Regression Coefficients	4.23	0.52	6.62
Unobservable Characteristics	69.36	68.63	64.01
Total	100.00	100.00	100.00

TABLE 4
Decomposition for Each Year of the Between-Groups, Within-Groups, and Overlapping Components into Elements Corresponding to the Explanatory Variables, the Regression Coefficients, and the Unobservable Characteristics Elements

Elements of the Decomposition	1982	1990	1998
Between-Groups Dispersion			
Differences in the Value of the Explanatory Variables	35.41	56.98	11.82
Differences in the Regression Coefficients	64.59	43.02	88.18
TOTAL	100.00	100.00	100.00
Within-Groups Dispersion			
Differences in the Value of the Explanatory Variables	27.61	32.35	36.00
Differences in Unobservable Characteristics	72.39	67.65	64.00
TOTAL	100.000	100.000	100.000
Overlap			
Differences in the Value of the Explanatory Variables	22.77	25.29	31.93
Differences in the Regression Coefficients	-13.86	-6.75	-25.39
Differences in Unobservable Characteristics	91.09	81.46	93.45
TOTAL	100.00	100.00	100.00

IV. THE DECOMPOSITION OF THE CHANGE OVER TIME IN THE WAGE DISPERSION: METHODOLOGY AND EMPIRICAL ILLUSTRATION

The breakdown of the change over time in the degree of income dispersion is quite complex and its details are given in Appendix 2. It may be summarized by looking at the impact of various factors on changes in the three components of the overall dispersion, the between-groups dispersion, the within-groups dispersion, and the overlapping component.

Concerning the changes in the within-groups dispersion, it is shown in Appendix 2 that three elements play a role:

- the changes that take place over time within the various groups in the dispersion of the explanatory variables
- the modification that take place over time in the different groups in the regression coefficients
- the variations over time within the various groups in the dispersion of the unobserved characteristics.

Concerning the changes in the between-groups dispersion, the following elements are distinguished (see Appendix 2):

- the changes over time in the relative size of the different groups
- the variations over time in the between-groups dispersion of the explanatory variables
- the modifications taking place over time in the average values of these variables
- the variations over time in the between-groups dispersion of the regression coefficients
- the modifications that take place over time in the average values of these regression coefficients.

Finally, concerning the overlapping component, the following factors are listed in Appendix 2:

- the changes in the dispersion of the values of the explanatory variables for those individuals involved in the overlapping in each period
- the variations over time in the average values of the explanatory variables among those same individuals
- the changes over time in some weighted dispersion of the regression coefficients, these weights depending only on those individuals involved in the overlapping in

each period

- the variation over time in some weighted average of these regression coefficients, here again the weights depending only on those individuals who are part of the overlapping in each period
- the change over time in the dispersion of the unobserved characteristics among those individuals involved in the overlapping in each period

The results of this complex breakdown[15] are given in Tables 5 through 7. Table 5 gives the respective impacts of changes in the between- and within-groups mean difference of the logarithms of wages, as well as of variations in the overlapping component. It first appears that in both periods (1982–1990 and 1990–1998) there was an increase in the mean difference of the logarithms of wages, this increase being much stronger during the period 1990–1998. It can, in fact, be observed that the variations in the overall mean difference of the logarithms of wages given in Table 5 correspond to the results given in Table 2. Given that the change in this mean difference was quite small during the period 1982–1990, we will concentrate our attention on what is observed during the period 1990–1998. As indicated in Table 2, the mean difference of the logarithms of wages increased from 64.4% to 72.98% between 1990 and 1998. In other words, the expected percentage gap in wages between two individuals drawn randomly (with repetition) from the sample increased in absolute terms by 8.6% (from 64.4% to 72.98%), which is exactly the result that appears in Tables 5 and 6.

Note that during the period 1982–1990 the increase in the between-groups dispersion would, per se, have led to a decrease in the amount of overlap, but since at the same time the within-groups dispersion increased we also end up with an increase in the amount of overlap between the wage distributions of the four groups distinguished.

[15] It should be observed that the breakdown proposed here goes beyond the one suggested by Blau and Kahn (1996, 1997, 2000), first because we are able to deal with more than two groups, second because we do not limit ourselves to changes in the between-groups dispersion, and third because we stress the impact of more than four determinants. The four determinants they identify are changes in the average value of productive characteristics (explanatory variables), in the prices of these productive characteristics (in the regression coefficients), in the dispersion of unmeasured characteristics, and finally in the relative position of the residuals of one group in the distribution of those of the other (an effect that is somehow one of the elements determining what we called change in the overlap).

Between 1990 and 1998, on the contrary, there was an increase in both the between- and the within-groups dispersion and the net result was a decrease in the amount of overlap.

Table 6 analyzes this increase in wage dispersion from another angle. There we try to identify whether the increase in wage dispersion was a consequence of changes in the value of the explanatory variables, of variations in the regression coefficients, of changes in the unobserved variables, or even of a modification in the relative size of the population subgroups. Table 6 shows, thus, that between 1990 and 1998 more than half of the increase in the overall wage dispersion was the consequence of an increase in the contribution of the regression coefficients.

Table 7 combines the results of Table 5 and 6. In addition, it makes a distinction between the contribution of changes in the average value and in the dispersion of the explanatory variables, as well as in the average value and in the dispersion of the coefficients of these variables in the earnings functions. In what follows we will try to give an intuitive interpretation to these various effects and then present and analyze the results of our empirical investigation.

A. The Determinants of the Change Over Time in the Between-Groups Dispersion

Let us start with the impact of the change over time in the between-groups dispersion. As proven in Appendix 2, three factors may play a role here: the explanatory variables, the regression coefficients, and the relative size of the population subgroups. In addition, it should be stressed that the explanatory variables have an impact on the variation over time in the wage dispersion either because their average value changes or because their dispersion varies. The intuition of this distinction is as follows.

Assume first that the average value (average computed for groups h and k together) of a given explanatory variable l increases between times 0 and 1. Then for a given gap between the coefficients β_{lh} and β_{lk} of this variable l in the earnings functions of groups h and k, the between-groups h and k wage dispersion should increase, ceteris paribus. Assume now that what increased is the gap between-groups h and k in the average value of a given explanatory variable l. Then for given coefficients β_{lh} and β_{lk} of the variable l in groups h and k, the between-groups wage dispersion between-groups h and k should increase, ceteris paribus.

Assume now that the average value over groups h and k of the coefficients β_{lh} and β_{lk} of variable l in the earnings function of groups h and k increased. Then for a given gap between-groups h and k in the value of the explanatory variable l, the wage dispersion should increase, ceteris paribus.

But if one assumes that what increased is the gap (β_{lh} - β_{lk}) between groups h and k in the value of the coefficient of variable l in the earnings functions of groups h and k, then for a given average value (over groups h and k) of variable l the wage dispersion should increase, ceteris paribus.

Finally, as far as the impact of a change in relative size of the population subgroups is concerned, the idea is, without entering into the technicalities of Appendix 2, that changes in the relative sizes of the groups should also affect the overall wage dispersion.

B. The Determinants of the Change Over Time in the Within-Groups Dispersion

The intuition for the results derived in Appendix 2 is here simpler. Given that when we analyze the within-groups dispersion we assume from the onset that the regression coefficients in the earnings functions are the same for all the individuals belonging to a given population subgroup, a change over time in these coefficients will have an impact on the within-groups wage dispersion of the explanatory variables. Similarly, assuming no change over time in these coefficients, a change in the within-groups dispersion of the explanatory variables will lead to a change in the overall within-groups dispersion. Clearly, a change in the dispersion of the unobservables will also have an effect on the within-groups dispersion. Finally, given that the overall within-groups income dispersion is a weighted average of the income dispersion within the various groups (the weight depending on the relative size of the population subgroups), a change in the relative size of the different groups will also have an impact on the overall within-groups dispersion.

C. The Determinants of the Change Over Time in the Value of the Overlapping Component

The intuitive interpretation of the various elements of the change in this overlapping component is similar to that given previously to the components of the change in the between- and within-groups dispersion. We will therefore not examine each component of this change, but just take an example. Assume a change occurred (among those involved in the overlapping) in the dispersion of the explanatory variables. Then clearly, assuming no change in the regression coefficients, there will be a modification of the overlapping component. Similarly, assume there was no change (among those involved in the overlapping) in the dispersion of the explanatory variables, but only in the value of the regression coefficients. Then, evidently, there will be a change in the value of the overlapping component.

Similar interpretations may be given to the other components of the change in the overlapping component.

Let us now take a look at the results presented in Table 7 and concentrate on the changes that occurred in the overall wage dispersion between 1990 and 1998. It appears that the change (8.19%) in the between-groups dispersion was almost equal to that in the overall dispersion (8.59%) because the changes in the within-groups dispersion and in the amount of overlap neutralized each other. More than half this contribution of changes in the between-groups dispersion was a consequence of variations in the dispersion of the regression coefficients (L). Another important factor was the change in the dispersion of the variables themselves (H). Table 7 indicates also that, as far as changes in the within-groups dispersion are concerned, the most important contribution to this change is related to variations over time in the various groups in the dispersion of the unobserved components.

Finally, the most important contributions to changes in the overlapping component refer either to changes over time in the dispersion of the regression coefficients or to a variation in the average value of the explanatory variables among those individuals affected by the overlapping.

D. The Contribution of Different Population Subgroups to the Various Components of the Change in the Overall Wage Dispersion

In Section IIIB we explained how to compute the contribution of the different population subgroups to the three components of the overall wage dispersion (the between- and within-groups wage dispersion, and the overlapping component). A similar exercice may be implemented to compute the contribution of the various population subgroups to the change over time in the three components. It is even possible to combine such a breakdown by population subgroups with the decomposition in subcomponents (impact of explanatory variables, of regression coefficients, and of unobservable characteristics) that were given in Table 7. This complex breakdown will not be detailed,[16] but an illustration of its application to the period 1990–1998 is given in Table 8. The latter indicates clearly the important role played by the immigrants who came from Europe or America after 1972. As far as the change in the between-groups dispersion is concerned, we see, as expected, the important contribution of this group to the changes in the dispersion of the regression coefficients (component L) and in the relative size of the various groups (components G and J). We also can observe the important role this group of immigrants plays in affecting the change in the overlapping component via, for example, the change over time in the dispersion of the unobservables (Y) among those individuals involved in the overlap.

[16] It may be obtained, upon request, from the authors.

TABLE 5

Decomposition of the Overall Change between 1982 and 1998 in the Mean Difference of the Logarithms of Wages: The Role of Changes in the Between- and Within-Groups Mean Difference and of that in the Overlapping Component

Components of the Decomposition	Period 1982–1990	Period 1990–1998	Period 1982–1998
Change in Between-Groups Mean Difference	-5.04	8.19	3.15
Change in Within-Groups Mean Difference	3.85	4.74	8.59
Change in Overlapping Component	2.01	-4.35	-2.34
Total Change in Mean Difference	0.8142	8,587	9.401

TABLE 6

Decomposition of the Overall Change between 1982 and 1998 in the Mean Difference of the Logarithms of Wages: The Role of Changes in the Relative Size of the Population Subgroups, in the Value of the Explanatory Variables, in the Regression Coefficients, and in the Unobserved Variables.

Components of the Decomposition	Period 1982–1990	Period 1990–1998	Period 1982–1998
Impact of Change in the Relative Size of the Population Subgroups	-2.91	-0.28	-3.94
Impact of Changes in the Value of the Explanatory Variables	4.98	1.70	7.71
Impact of Changes in the Regression Coefficients	-1.35	4.64	3.02
Impact of Changes in Unobservables	0.10	2.52	2.62
Total Change in Mean Difference	0.8142	8.587	9.401

TABLE 7

Decomposition by Subcomponents of the Change in the Mean Difference of the Logarithms of Wages

Subcomponents 1) Change in the Between-Groups Wage Dispersion	Period 1982–1990	Period 1990–1998	Period 1982–1998
G + J	-2.91	-0.28	-3.94
H	0.07	3.11	1.77
K	0.07	-0.25	0.71
I	1.52	0.90	3.83
L	-3.79	4.71	0.78
Total Change in Between-Groups Dispersion	-5.04	8.19	3.15
2) Changes in the Within-Groups Wage Dispersion	**Period 1982–1990**	**Period 1990–1998**	**Period 1982–1998**
C	2.20	1.34	3.82
D	-0.11	1.16	0.77
B	1.76	2.24	4.00
Total Change in Within-Groups Dispersion	3.85	4.74	8.59
3) Change in the Overlapping Component	**Period 1982–1990**	**Period 1990–1998**	**Period 1982–1998**
S	1.49	0.32	3.07
U	1.15	-2.82	-1.67
T	-0.12	0.69	-0.69
V	1.15	-2.82	-1.67
Y	-1.66	0.28	-1.38
Total Change in Overlapping Component	2.01	-4.35	-2.34
4)Total Change in Wage Dispersion	0.81	8.59	9.40

Explanation of symbols:

(G + J): effect of changes over time in the relative size of the different groups
(H + K): the impact of variations over time in the value of the explanatory variables in the various groups with
H: impact of variations over time in the dispersion of these variables
K: role of modifications taking place over time in the average value of these variables
(I + L): role played by modifications over time in the regression coefficients in the different groups with
L: impact of variations over time in the dispersion over the different groups of these regression coefficients
I: the role played by the modifications that take place over time in the average value of these regression coefficients
C: impact of changes over time in the various groups in the dispersion of the explanatory variables
D: effect of the modification that took place over time in the different groups in the regression coefficients
B: role played by variations over time in the various groups in the dispersion of the unobserved components
S = (S_1+S_2+S_3): the effect of changes in the dispersion of the value of the explanatory variables among those involved in the overlapping at both periods

$U = (U_1+U_2+U_3)$: impact of variations over time in the average values of the explanatory variables among those same individuals
$V = (V_1+V_2+V_3)$: effect of changes over time in some weighted dispersion of the regression coefficients, these weights depending only on those individuals involved in the overlapping
$T = (T_1+T_2+T_3)$: impact of the variation over time in some weighted average of these regression coefficients, here again the weights depending only on those individuals who are part of the overlapping.
Y: role of change over time in the dispersion of the unobserved components

TABLE 8

Contributions of the Different Population Subgroups to the Various Components of the Changes in Wage Dispersion during the 1990–1998 Period.

Component of Change	**Contribution of Those Born in Israel**	**Contribution of Those Born in Asia or Africa**	**Contribution of Those Born in Europe or America Who Immigrated before 1972**	**Contribution of Those Born in Europe or America Who Immigrated after 1972**	**All Groups Together**
Change in Relative Size of Population	0.21	-0.58	-1.41	1.50	-0.28
Change in Value of Explanatory Variables	2.19	-1.24	-0.14	0.90	1.70
Change in Regression Coefficients	1.89	0.79	0.63	1.34	4.64
Change in Unobserved Variables	3.08	-4.36	-3.08	6.88	2.52
Total Change in Wage Dispersion	7.37	-5.40	-4.01	10.62	8.59
Change in Between-Groups Dispersion	2.85	0.22	0.08	5.05	8.19
G+J	0.21	-0.58	-1.41	1.50	-0.28
H	1.22	0.81	0.81	0.28	3.11
K	-0.26	-0.26	0.01	0.25	-0.25
I	0.48	-0.07	0.57	-0.08	0.90
L	1.19	0.32	0.10	3.10	4.71
Change in Within-Groups Dispersion	5.30	-2.42	-1.12	2.99	4.74
C	1.81	-0.58	-0.38	0.49	1.34
D	1.05	0.09	0.03	-0.02	1.16
B	2.44	-1.93	-0.78	2.51	2.24
Change in Overlap	-0.78	-3.19	-2.96	2.58	-4.35
S	0.48	-1.18	-0.68	1.70	0.32
U	-1.06	-0.03	0.10	-1.83	-2.82
T	0.22	0.48	-0.17	0.16	0.69
V	-1.06	-0.03	0.10	-1.83	-2.82
Y	0.64	-2.43	-2.31	4.37	0.28
Total Change in Wage Dispersion	7.37	-5.40	-4.01	10.62	8.59

V. CONCLUDING COMMENTS

This paper extends Oaxaca's (1973) original approach by proposing a methodology for analyzing the respective impact of explanatory variables, the coefficients of these variables in the earnings functions, and unobservable characteristics on the overall wage dispersion and its components (between- and within-groups dispersion, as well as the degree of overlap between the groups' wage distribution when the individuals are also characterized by the groups to which they belong). An illustration based on income surveys conducted in Israel in 1982, 1990, and 1998 and making a distinction between four population subgroups, one of them including immigrants from Europe or America who came after 1971, indicated that the approach proposed here sheds some interesting light on the evolution over time of the wage dispersion in Israel. The paper proposed also a methodology to decompose changes over time in the amount of wage dispersion. When applied to the same Israeli data this approach allowed us to show, for example, that the increase in the overall wage dispersion between 1990 and 1998 was strongly connected to the increase in the between-groups dispersion, which was itself related to an increase during this period in the dispersion of the regression coefficients which followed the massive immigration that took place in the early 1990s in Israel.

REFERENCES

Amir, S. 1980. "Changes in the Wage Function for Israeli Jewish Male Employees between 1968–69 and 1975–76." *Bank of Israel Review* 52(August): 5–29 (in Hebrew).

————. 1993. "The Absorption Process of Academic Immigrants from the USSR in Israel: 1978–1984." Report of the Israeli International Institute for Applied Economic Policy Review (in Hebrew).

Beenstock, M. 1993. "Learning Hebrew and Finding a Job: Econometric Analysis of Immigration Absorption in Israel." Discussion Paper No. 93.05. Jerusalem: The Maurice Falk Institute for Economic Research in Israel.

Berger, M. 1985. "The Effect of Cohort Size on Earnings Growth: A Reexamination of the Evidence." *Journal of Political Economy* 93(3): 561–573.

Blinder, A. S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8(4): 436–55.

Blau, F. D. and L. M. Kahn. 1996. "Wage Structure and Gender Earnings Differentials: An International Comparison." *Economica* 63(Supp): S29–S62.

————. 1997. "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s." *Journal of Labor Economics* 15(1): 1–42.

————. 2000. "Gender Differences in Pay." *Journal of Economic Perspectives* 14(4): 75–99.

Borjas, G. J. 1985. "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants." *Journal of Labor Economics* 3(4): 463–89.

————. 1994. "The Economics of Immigration." *Journal of Economic Literature* XXXII(4): 1667–1717.

————. 2005. *Labor Economics*, Third edition. Columbus, OH: McGraw-Hill.

Chiswick, B. R. 1978. "The Effect of Americanization on the Earnings of Foreign-Born Men." *Journal of Political Economy* 86(5): 897–921.

————. 1997. "Hebrew Language Usage: Determinants and Effects on Earnings among Immigrants in Israel." Discussion paper No. 97.09. Jerusalem: The Maurice Falk Institute for Economic Research in Israel.

Dagum, C. 1960. “Teoria de la transvariacion—Sus aplicaciones a la economia.” *Metron* XX.

————. 1997. “A New Approach to the Decomposition of the Gini Income Inequality Ratio.” *Empirical Economics* 22(4): 515–531.

Deutsch, J., and J. Silber. 1998. “The Overlapping of Distributions: Alternative Measures and their Application to the Analysis of Consumption Patterns.” in S. R. Chakravarty, D. Coondoo, and R. Mukherjee (eds.) *Quantitative Economics: Theory and Practice.* New Delhi: Allied Publishers.

Eckstein, Z., and Y. Weiss. 1997.“The Absorption of Highly Skilled Immigrants: Israel, 1991–1995.” Paper presented at the CEPR conference, “European Migration: What Do We Know.” Munich, Germany. November.

Fields, G. S. 2003. “Accounting for Income Inequality and its Change: A New Method, with Application to the Distribution of Earnings in the United States.” *Research in Labor Economics*. 22: 1–38.

Fortin, N. M., and T. Lemieux. 1997. “Institutional Change and Rising Wage Inequality: Is There a Linkage?” *Journal of Economic Perspectives* 11(2): 75–96.

Freeman, R. 1995. “Are Your Wages Set in Beijing.” *Journal of Economic Perspectives* 9(3): 15–32.

Friedberg, R. M. 1995. “You Can’t Take it with You? Immigration Assimilation and the Portability of Human Capital: Evidence from Israel.” Discussion paper No. 95.02. Jerusalem: The Maurice Falk Institute for Economic Research in Israel.

Gini, C. 1912. “Variabilita e Mutabilita: Contributo allo Studio delle Distribuzioni e Relazioni Statistiche.” *Studio Economico-Giuridice dell’Univ. di Cagliari* 3: 1–158.

————. 1959. *Memorie de Metodologia Statistica: Volume Secondo—Transvariazione.* Rome: Libreria Goliardica.

Goldin, C., and R. Margo. 1992. “The Great Compression: The Wage Structure in the United States at Mid-Century.” *Quarterly Journal of Economics* 107(1): 1–34.

Johnson, G. E. 1997. “Changes in Earnings Inequality: The Role of Demand Shifts.” *Journal of Economic Perspectives* 11(2): 41–54.

Kendall, M. G., and A. Stuart. 1989. *The Advanced Theory of Statistics.* London: Charles Griffin and Company Limited.

Krueger, A. B. 1993. “How Computers Have Changed the Wage Structure: Evidence from Micro-data, 1984–1989.” *Quarterly Journal of Economics* 108(1): 33–60.

Levy, F., and R. J. Murnane. 1992. “U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations.” *Journal of Economic Literature* 30(3): 1333–81.

Mincer, J. 1974. *Schooling, Experience, and Earnings*. New York: Columbia University Press.

Neuman, S. 1999. “Alyah to Israel: Immigration under Conditions of Adversity.” Discussion Paper No. 89. IZA: Bonn.

Oaxaca, R. 1973. “Male-Female Wage Differentials in Urban Labor Markets.” *International Economic Review* 14(3): 693–709.

Ofer, G., A. Vinokur, and Y. Bar-Chaim. 1980. “Absorption in Israel and Economic Contribution of Immigrants from the Soviet Union.” The Maurice Falk Institute of Economic Research (in Hebrew), Jerusalem.

Polachek, S. W. 1975. “Differences in Expected Post-School Investment as a Determinant of Market Wage Differentials.” *International Economic Review* 16(2): 451–70.

Reimers, C. 1983. “Labor Market Discrimination Against Hispanic and Black Men.” *Review of Economics and Statistics* 65(4): 570–9.

Topel, R. H. 1997. “Factor Proportions and Relative Wages: The Supply-Side Determinants of Wage Inequality.” *Journal of Economic Perspectives* 11(2): 55–74.

Welch, F. 1979. “The Effects of Cohort Size on Earnings: The Baby Boom Babies’ Financial Bust.” *Journal of Political Economy* 87(5): 65–97.

Weiss, Y., G. Fishelson, and N. Mark. 1979. “Income Gaps Between Men by Continent of Origin: Israel, 1969–76.” The Foerder Economic Research Institute, Tel-Aviv University, June.

Wood, A. 1995. “How Trade Hurt Unskilled Workers.” *Journal of Economic Perspectives* 9(3): 57–80.

APPENDIX 1-A

Descriptive Statistics and Regression Results for the 1982 Income Survey for Each of the Four Subpopulations

1) Individuals Born in Israel (I)

Summary Statistics for the Regression				
$R^2 = 0.3601$	SE = 0.48927	F = 82.32	N = 868	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	3.9943	0.6113		
Intercept			2.1859	12.70
Married	0.7465	0.4350	0.4441	3.17
Single	0.2385	0.4262	0.1503	1.03
Years of Schooling	11.5657	3.0633	0.0882	14.73
Years of Experience	15.7247	10.4861	0.0364	6.23
Square of Years of Experience	357.2224	501.4416	-0.0004	-3.38
Attended Talmudic School	0.0184	0.1345	-0.6728	-5.36

2) Individuals Born in Asia or Africa (AA)

Summary Statistics for the Regression				
$R^2 = 0.1700$	SE = 0.45953	F = 30.53	N = 866	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	3.9148	0.5041		
Intercept			3.2380	20.73
Married	0.9226	0.2672	0.0327	0.29
Single	0.0566	0.2310	-0.1657	-1.24
Years of Schooling	8.9885	3.8689	0.0491	9.99
Years of Experience	29.0381	12.5786	0.0184	3.15
Square of Years of Experience	1001.4319	787.3419	-0.0003	-3.46
Attended Talmudic School	0.0173	0.1305	-0.1877	-1.55

3) Individuals Born in Europe or America Who Immigrated before 1972 (EA)

Summary Statistics for the Regression				
$R^2 = 0.2152$	SE= 0.50667	F= 35.69	N = 760	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	4.1880	0.5716		
Intercept			3.2318	20.41
Married	0.9263	0.2613	0.2038	2.17
Single	0.0329	0.1784	-0.1363	-0.97
Years of Schooling	11.3132	3.7253	0.0595	10.13
Years of Experience	35.0921	14.1397	0.0131	2.29
Square of Years of Experience	1431.3862	971.1232	-0.0002	-2.88
Attended Talmudic School	0.0474	0.2124	-0.3545	-4.04

4) Individuals Born in Europe or America Who Immigrated after 1971 (NIM)

Summary Statistics for the Regression				
$R^2 = 0.2607$	SE= 0.46280	F= 14.52	N = 231	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	3.9398	0.5371		
Intercept			2.6320	11.14
Married	0.9177	0.2747	0.2433	1.53
Single	0.0433	0.2035	0.0186	0.08
Years of Schooling	11.5563	4.0091	0.0499	5.85
Years of Experience	25.7078	14.1658	0.0480	4.94
Square of Years of Experience	861.5595	836.2168	-0.0008	-5.24
Attended Talmudic School	0.0130	0.1132	-0.1072	-0.39

APPENDIX 1-B

Descriptive Statistics and Regression Results for the 1990 Income Survey for Each of the Four Subpopulations

1) Individuals Born in Israel (I)

Summary Statistics for the Regression				
R^2 = 0.3373	SE= 0.45880	F=132.33	N = 1549	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	2.4700	0.5624		
Intercept			0.8836	8.26
Married	0.7289	0.4445	0.0592	0.79
Single	0.2460	0.4307	-0.1149	-1.41
Years of Schooling	12.1911	2.6088	0.0891	18.98
Years of Experience	16.3499	10.3207	0.0435	10.61
Square of Years of Experience	373.8360	472.8218	-0.0006	-7.10
Attended Talmudic School	0.0232	0.1507	-0.3045	-3.87

2) Individuals Born in Asia or Africa (AA)

Summary Statistics for the Regression				
R^2 = 0.2653	SE= 0.52880	F= 29.53	N = 475	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	2.4929	0.5411		
Intercept			1.3383	8.75
Married	0.9178	0.2746	0.0487	0.53
Single	0.0455	0.2084	-0.1684	-1.33
Years of Schooling	10.1043	3.6764	0.0594	10.77
Years of Experience	30.3951	12.1008	0.0353	5.66
Square of Years of Experience	1070.2892	802.5076	-0.0005	-5.56
Attended Talmudic School	0.0126	0.1117	-0.4711	-3.04

3) Individuals Born in Europe or America Who Immigrated before 1972 (EA)

Summary Statistics for the Regression				
R^2 = 0.2653	SE= 0.52880	F= 29.53	N = 475	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	2.6875	0.6163		
Intercept			1.0617	4.87
Married	0.9137	0.2808	0.1419	1.21
Single	0.0400	0.1960	-0.2130	-1.25
Years of Schooling	11.8347	3.4586	0.0773	9.26
Years of Experience	34.9189	13.1700	0.0402	4.64
Square of Years of Experience	1392.7826	936.3921	-0.0006	-4.70
Attended Talmudic School	0.0274	0.1632	-0.1852	-1.23

4) Individuals Born in Europe or America Who Immigrated after 1971 (NIM)

Summary Statistics for the Regression				
R^2 = 0.2027	SE= 0.53214	F= 13.58	N = 298	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	2.4516	0.5949		
Intercept			1.0504	4.42
Married	0.8221	0.3824	0.2392	1.68
Single	0.1275	0.3336	-0.0334	-0.18
Years of Schooling	12.8473	3.1253	0.0733	6.84
Years of Experience	23.0889	13.7142	0.0187	1.80
Square of Years of Experience	721.1770	734.6558	-0.0002	-1.24
Attended Talmudic School	0.0134	0.1151	-0.1618	-0.60

APPENDIX 1-C

Descriptive Statistics and Regression Results for the 1998 Income Survey for Each of the Four Subpopulations

1) Individuals Born in Israel (I)

Summary Statistics for the Regression				
R^2 = 0.3819	SE= 0.50891	F=343.88	N = 3331	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	3.4884	0.6472		
Intercept			1.3567	16.96
Married	0.6593	0.4740	0.1301	2.61
Single	0.3074	0.4614	-0.338	-0.61
Years of Schooling	12.7199	2.4466	0.1249	33.33
Years of Experience	16.8355	10.7915	0.0378	11.76
Square of Years of Experience	399.8894	469.3617	-0.0004	-6.12
Attended Talmudic School	0.0186	0.1352	-0.2884	-4.38

2) Individuals Born in Asia or Africa (AA)

Summary Statistics for the Regression				
R^2 = 0.2119	SE= 0.56595	F= 41.90	N = 914	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	3.4912	0.6371		
Intercept			1.9982	12.06
Married	0.8928	0.3094	-0.0397	-0.47
Single	0.0547	0.2274	-0.1357	-1.07
Years of Schooling	11.0717	3.5468	0.0727	11.95
Years of Experience	32.0957	11.4843	0.0467	6.11
Square of Years of Experience	1162.0249	742.8596	-0.0007	-5.72
Attended Talmudic School	0.0098	0.0987	-0.0205	-0.11

3) Individuals Born in Europe or America Who Immigrated before 1972 (EA)

Summary Statistics for the Regression				
R^2 = 0.2508	SE= 0.59123	F=28.34	N = 491	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	3.8053	0.6824		
Intercept			1.4726	5.36
Married	0.9206	0.2704	0.1637	1.36
Single	0.0265	0.1605	-0.1148	-0.56
Years of Schooling	13.0356	3.0756	0.1098	10.88
Years of Experience	34.9868	11.5728	0.0380	3.40
Square of Years of Experience	1358.0025	879.0703	-0.0004	-2.83
Attended Talmudic School	0.0224	0.1480	-0.2439	-1.34

4) Individuals Born in Europe or America Who Immigrated after 1971 (NIM)

Summary Statistics for the Regression				
R^2 = 0.1568	SE= 0.54518	F= 46.25	N = 1461	
Summary Statistics for the Variables and Regression Results				
Variable	Mean	Standard Deviation	Coefficients	t-values
Logarithm of Wage per Hour	3.2118	0.5935		
Intercept			2.1244	19.02
Married	0.7933	0.4049	0.2093	3.35
Single	0.1506	0.3576	-0.0976	-1.24
Years of Schooling	13.2988	2.6319	0.0651	11.77
Years of Experience	21.7888	12.3316	0.0127	2.66
Square of Years of Experience	626.8232	613.1620	-0.0003	-3.61
Attended Talmudic School	0.0089	0.0939	-0.3121	-2.04

APPENDIX 2: DECOMPOSING THE CHANGE OVER TIME IN THE WAGE DISPERSION

We will successively derive a breakdown of the three components of expression (14) as they have been given in expressions (21) through (23), (25) through (27), and (29) through (32).

1) Decomposing the Change Over Time in the Within-Groups Dispersion

Let the subscripts 0 and 1 refer, respectively, to time 0 and 1. Assume n_0 and n_1 refer, respectively, to the size of the population at times 0 and 1. Using (21) we can now express the change over time in the within-groups dispersion as

$$\Delta_{W,1} - \Delta_{W,0} = A + B \qquad \text{(B-1)}$$

where

$$A = A1 - A0 \qquad \text{(B-2)}$$

with

$$A1 = (1/n_1^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{j\in h} 2 \left[\left(\sum_{l=1 \text{ to } L} \beta_{lh,1} (x_{lih,1} - x_{ljh,1}\right)\right]$$

$$\leftrightarrow A1=(1/n_1^2)\ 2 \sum_{h=1 \text{ to } m} \sum_{l=1 \text{ to } L} \left[\sum_{i\in h} \sum_{j\in h} \beta_{lh,1} (x_{lih,1} - x_{ljh,1})\right] \qquad \text{(B-3)}$$

with $\ln y_{ih,1} > \ln y_{jh,1}$

and

$$A0 =(1/n_0^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{j\in h} 2 \left[\left(\sum_{l=1 \text{ to } L} \beta_{lh,0} (x_{lih,0} - x_{ljh,0}\right)\right]$$

$$\leftrightarrow A0 =(1/n_0^2)\ 2 \sum_{h=1 \text{ to } m} \sum_{l=1 \text{ to } L} \left[\sum_{i\in h} \sum_{j\in h} \beta_{lh,0} (x_{lih,0} - x_{ljh,0})\right] \qquad \text{(B-4)}$$

with $\ln y_{ih,0} > \ln y_{jh,0}$,while

$$B = B1 - B0 \qquad \text{(B-5)}$$

with

$$B1 = (1/n_1^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{j\in h} 2 (u_{ih,1} - u_{jh,1}), \text{ with } \ln y_{ih,1} > \ln y_{jh,1} \qquad \text{(B-6)}$$

and

$$B0 = (1/n_0^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{j\in h} 2 (u_{ih,0} - u_{jh,0}), \text{ with } \ln y_{ih,0} > \ln y_{jh,0} \qquad \text{(B-7)}$$

Combining (B-3) and (B-4) we derive that

$$A1\text{–}A0 = (1/n_1^2)\ 2\sum_{h=1 \text{ to } m} \sum_{l=1 \text{ to } L} \sum_{i\in h} \sum_{j\in h} [\beta_{lh,1} (x_{lih,1} - x_{ljh,1})] - (1/n_0^2)\ 2\sum_{h=1 \text{ to } m} \sum_{l=1 \text{ to } L} \sum_{i\in h} \sum_{j\in h} [\beta_{lh,0} (x_{lih,0} - x_{ljh,0})] \quad \text{(B-8)}$$

where for each period 0 and 1, in each group and for each explanatory variable, the individuals are ranked by decreasing value of their income (i.e., of the logarithm of their income).

Expression (B-8) may then be written as

$$A1 - A0 = CC + DD \quad \text{(B-9)}$$

where

$$CC=[\sum_{h=1 \text{ to } m} \sum_{l=1 \text{ to } L}(\beta_{lh,1}+\beta_{lh,0})[\sum_{i\in h}\sum_{j\in h}((1/n_1^2)(x_{lih,1}- x_{ljh,1}) - (1/n_0^2)(x_{lih,0} - x_{ljh,0}))]\ ] \quad \text{(B-10)}$$

and

$$DD=[\sum_{h=1 \text{ to } m} \sum_{l=1 \text{ to } L} (\beta_{lh,1}-\beta_{lh,0})[\sum_{i\in h}\sum_{j\in h}((1/n_1^2)(x_{lih,1} - x_{ljh,1}) + (1/n_0^2)(x_{lih,0} - x_{ljh,0}))]] \quad \text{(B-11)}$$

Similarly, combining (B-5), (B-6), and (B-7), we may express B as

$$B = BB = 2 \sum_{h=1 \text{ to } m} [\sum_{i\in h} \sum_{j\in h} (1/n_1^2)(u_{ih,1} - u_{jh,1})- (1/n_0^2)(u_{ih,0} - u_{jh,0})] \quad \text{(B-12)}$$

where the unobserved elements u_{ih} are, in each period, ranked by decreasing values of the individual incomes (i.e., of the logarithms of the individual incomes).

Combining finally expressions (B-1), (B-2), (B-9), (B-10), (B-11), and (B-12) we conclude that

$$\Delta_{W,1} - \Delta_{W,0} = CC + DD + BB \quad \text{(B-13)}$$

where CC measures the impact of changes over time in the various groups in the dispersion of the explanatory variables, DD the effect of the modification that took place over time in the different groups in the regression coefficients, BB corresponds to the role played by variations over time in the various groups in the dispersion of the unobserved components, and AA measures the impact of the change over time in the relative size of the population subgroups.

2) Decomposing the Change Over Time in the Between-Groups Dispersion

On the basis of (25) and using again the subscripts 1 and 0 to denote the periods 1 and 0, we may express the change over time in the between-groups dispersion as

$$\Delta_{B,1} - \Delta_{B,0} = E + F \qquad \text{(B-14)}$$

where

$$E = (E1 - E0) \qquad \text{(B-15)}$$

with

$$E1=\{\sum_{h=1 \text{ to } m} \sum_{k=1 \text{ to } m} ((n_{h,1}\, n_{k,1})/n_1^2) \; [\sum_{l=1 \text{ to } L} (\beta_{lh,1} + \beta_{lk,1})(x_{lgh,1} - x_{lgk,1})]\} \qquad \text{(B-16)}$$

for $\ln y_{gh1} > \ln y_{gk1}$, with

$$E0 =\{\sum_{h=1 \text{ to } m} \sum_{k=1 \text{ to } m} ((n_{h,0}\, n_{k,0})/n_0^2) \; [\sum_{l=1 \text{ to } L} (\beta_{lh,0} + \beta_{lk,0})(x_{lgh,0} - x_{lgk,0})]\} \qquad \text{(B-17)}$$

for $\ln y_{gh0} > \ln y_{gk0}$, while

$$F = (F1 - F0) \qquad \text{(B-18)}$$

with

$$F1=\{[\sum_{h=1 \text{ to } m}\sum_{k=1 \text{ to } m} ((n_{h,1}\, n_{k,1})/n_1^2) \; [\sum_{l=1 \text{ to } L} (x_{lgh,1} + x_{lgk,1})(\beta_{lh,1} - \beta_{lk,1})]]\} \qquad \text{(B-19)}$$

for $\ln y_{gh1} > \ln y_{gk1}$

$$F0 =\{[\sum_{h=1 \text{ to } m}\sum_{k=1 \text{ to } m} ((n_{h,0}\, n_{k,0})/n_0^2) \; [\sum_{l=1 \text{ to } L} (x_{lgh,0} + x_{lgk,0})(\beta_{lh,0} - \beta_{lk,0})]]\} \qquad \text{(B-20)}$$

for $\ln y_{gh0} > \ln y_{gk0}$.

Note that in expressions (B-16) through (B-20), the groups are classified by decreasing values of the means of the logarithms of the incomes during the relevant year, that is, by decreasing values of the geometric means of the incomes in the different groups.[17]

Combining (B-16) and (B-17) we now derive that the difference E1 – E0 may be expressed as

---

[17]Thus, in expression (B-19), for example, the difference ($x_{lgh,1} - x_{lgk,1}$) could be negative despite the fact that the difference ($\ln y_{gh,1} - \ln y_{gk,1}$) is, by assumption, positive. Similarly, in expression (B-23) the difference ($\beta_{lh,0} - \beta_{lk,0}$) may be negative although, by assumption, the gap ($\ln y_{gh,0} - \ln y_{gk,0}$) is positive.

$$(E1 - E0) = G + H + I \tag{B-21}$$

with

$$G = BG_1 + BG_2 + BG_3 \tag{B-22}$$

$$H = BH_1 + BH_2 + BH_3 \tag{B-23}$$

$$I = BI_1 + BI_2 + BI_3 \tag{B-24}$$

where

$$BG_1=\{(1/2)\sum_l \sum_h \sum_k ((\beta_{lh,1}+\beta_{lk,1})(x_{lgh,1}-x_{lgk,1}) \\ +(\beta_{lh,0}+\beta_{lk,0})(x_{lgh,0}-x_{lgk,0}))\ [(n_{h,1}n_{k,1})/n_1^2)\ -((n_{h,0}\,n_{k,0})/n_0^2)]\} \tag{B-25}$$

for the cases where $\ln y_{gh1} > \ln y_{gk1}$ and $\ln y_{gh0} > \ln y_{gk0}$

$$BG_2=\{(1/2)\sum_l \sum_h \sum_k ((\beta_{lh,1}+\beta_{lk,1})(x_{lgh,1}-x_{lgk,1}) \\ +(\beta_{lh,0}+\beta_{lk,0})(x_{lgk,0}-x_{lgh,0}))[((n_{h,1}n_{k,1})/n_1^2)\ -((n_{h,0}\,n_{k,0})/n_0^2)]\} \tag{B-26}$$

for the cases where $\ln y_{gh1} > \ln y_{gk1}$ and $\ln y_{gh0} < \ln y_{gk0}$

$$BG_3=\{(1/2)\sum_l \sum_h \sum_k ((\beta_{lh,1}+\beta_{lk,1})(x_{lgk,1}-x_{lgh,1}) \\ +(\beta_{lh,0}+\beta_{lk,0})(x_{lgh,0}-x_{lgk,0}))[\ ((n_{h,1}\,n_{k,1})/n_1^2)-\ ((n_{h,0}\,n_{k,0})/n_0^2)]\} \tag{B-27}$$

for the cases where $\ln y_{gh1} < \ln y_{gk1}$ and $\ln y_{gh0} > \ln y_{gk0}$

$$BH_1=(1/4)\sum_l\sum_h\sum_k(((n_{h,1}\,n_{k,1})/n_1^2)+\ ((n_{h,0}\,n_{k,0})/n_0^2))((\beta_{lh,1}+\beta_{lk,1}) \\ +(\beta_{lh,0}+\beta_{lk,0}))[(x_{lgh,1}-x_{lgk,1})-(x_{lgh,0}-x_{lgk,0})]\} \tag{B-28}$$

for the cases where $\ln y_{gh1} > \ln y_{gk1}$ and $\ln y_{gh0} > \ln y_{gk0}$

$$BH_2=\{(1/4)\sum_l\sum_h\sum_k(((n_{h,1}\,n_{k,1})/n_1^2)+((n_{h,0}n_{k,0})/n_0^2))((\beta_{lh,1}+\beta_{lk,1}) \\ +(\beta_{lh,0}+\beta_{lk,0}))[(x_{lgh,1}-x_{lgk,1})-(x_{lgk,0}-x_{lgh,0})]\} \tag{B-29}$$

for the cases where $\ln y_{gh1} > \ln y_{gk1}$ and $\ln y_{gh0} < \ln y_{gk0}$

$$BH_3=\{(1/4)\sum_l\sum_h\sum_k(((n_{h,1}n_{k,1})/n_1^2)+((n_{h,0}\,n_{k,0})/n_0^2))((\beta_{lh,1}+\beta_{lk,1}) \\ +(\beta_{lh,0}+\beta_{lk,0}))[(x_{lgk,1}-x_{lgh,1})-(x_{lgh,0}-x_{lgk,0})]\} \tag{B-30}$$

for the cases where $\ln y_{gh1} < \ln y_{gk1}$ and $\ln y_{gh0} > \ln y_{gk0}$

$$BI_1=\{(1/4)\sum_l \sum_h\sum_k(((n_{h,1}n_{k,1})/n_1^2)+((n_{h,0}n_{k,0})/n_0^2))((x_{lgh,1}-x_{lgk,1}) \\ +(x_{lgh,0}-x_{lgk,0}))[(\beta_{lh,1}+\beta_{lk,1})-(\beta_{lh,0}+\beta_{lk,0})]\} \tag{B-31}$$

for the cases where $\ln y_{gh1} > \ln y_{gk1}$ and $\ln y_{gh0} > \ln y_{gk0}$

$$BI_2=\{(1/4)\sum_l \sum_h\sum_k(((n_{h,1}n_{k,1})/n_1^2)+((n_{h,0}n_{k,0})/n_0^2))((x_{lgh,1}-x_{lgk,1}) +(x_{lgk,0}-x_{lgh,0}))[(\beta_{lh,1}+\beta_{lk,1})-(\beta_{lh,0}+\beta_{lk,0})]\} \tag{B-32}$$

for the cases where $\ln y_{gh1} > \ln y_{gk1}$ and $\ln y_{gh0} < \ln y_{gk0}$

$$BI_3=\{(1/4)\sum_l \sum_h\sum_k(((n_{h,1}n_{k,1})/n_1^2)+((n_{h,0}\, n_{k,0})/n_0^2))((x_{lgk,1}-x_{lgh,1}) +(x_{lgh,0}-x_{lgk,0}))[(\beta_{lh,1}+\beta_{lk,1})-(\beta_{lh,0}+\beta_{lk,0})]\} \tag{B-33}$$

for the cases where $\ln y_{gh1} < \ln y_{gk1}$ and $\ln y_{gh0} > \ln y_{gk0}$.

Similarly, combining (B-22) and (B-23) we derive that the difference (F1-F0) in (B-18) may be expressed as

$$(F1 - F0)=\{\sum_{l=1 \text{to } L}\{\sum_{h=1 \text{ to } m}\sum_{k=1 \text{to } m}[((n_{h,1}\, n_{k,1})/n_1^2)(\beta_{lh,1}- \beta_{lk,1})(x_{lgh,1}+x_{lgk,1})] - [((n_{h,0}\, n_{k,0})/n_0^2)(\beta_{lh,0}-\beta_{lk,0})(x_{lgh,0}+x_{lgk,0})]\} \tag{B-34}$$

or, after some algebraic manipulations, as

$$(F1 - F0) = J + K + L \tag{B-35}$$

where

$$J = BJ_1 + BJ_2 + BJ_3 \tag{B-36}$$

$$K = BK_1 + BK_2 + BK_3$$
$$L = BL_1 + BL_2 + BL_3 \tag{B-37}$$

with

$$BJ_1=\{(1/2)\sum_l \sum_h\sum_k((\beta_{lh,1}-\beta_{lk,1})(x_{lgh,1}+x_{lgk,1})+ (\beta_{lh,0}-\beta_{lk,0})(x_{lgh,0}+x_{lgk,0}))[((n_{h,1}n_{k,1})/n_1^2)-((n_{h,0}n_{k,0})/n_0^2)]\} \tag{B-38}$$

for the cases where $\ln y_{gh1} > \ln y_{gk1}$ and $\ln y_{gh0} > \ln y_{gk0}$

$$BJ_2=\{(1/2)\sum_l \sum_h\sum_k ((\beta_{lh,1}-\beta_{lk,1})(x_{lgh,1}+x_{lgk,1}) +(\beta_{lk,0}-\beta_{lh,0})(x_{lgh,0}+x_{lgk,0}))[((n_{h,1}n_{k,1})/n_1^2)-((n_{h,0}n_{k,0})/n_0^2)]\} \tag{B-39}$$

for the cases where $\ln y_{gh1} > \ln y_{gk1}$ and $\ln y_{gh0} < \ln y_{gk0}$

$$BJ_3=\{(1/2)\sum_l \sum_h\sum_k ((\beta_{lk,1}-\beta_{lh,1})(x_{lgh,1}+x_{lgk,1}) +(\beta_{lh,0}-\beta_{lk,0})(x_{lgh,0}+x_{lgk,0}))[((n_{h,1}n_{k,1})/n_1^2)-((n_{h,0}n_{k,0})/n_0^2)]\} \tag{B-40}$$

for the cases where ln y_{gh1} < ln y_{gk1} and ln y_{gh0} > ln y_{gk0}

$$BK_1=\{(1/4)\sum_l\sum_h\sum_k(((n_{h,1}\,n_{k,1})/n_1^2)+((n_{h,0}\,n_{k,0})/n_0^2))((\beta_{lh,1}-\beta_{lk,1}) +(\beta_{lh,0}-\beta_{lk,0}))[((x_{lgh,1}+x_{lgk,1})-(x_{lgh,0}+x_{lgk,0}))]\} \quad \text{(B-41)}$$

for the cases where ln y_{gh1} > ln y_{gk1} and ln y_{gh0} > ln y_{gk0}

$$BK_2=\{(1/4)\sum_l\sum_h\sum_k(((n_{h,1}\,n_{k,1})/n_1^2)+((n_{h,0}\,n_{k,0})/n_0^2))((\beta_{lh,1}-\beta_{lk,1}) +(\beta_{lk,0}-\beta_{lh,0}))[(x_{lgh,1}+x_{lgk,1})-(x_{lgh,0}+x_{lgk,0})]\} \quad \text{(B-42)}$$

for the cases where ln y_{gh1} > ln y_{gk1} and ln y_{gh0} < ln y_{gk0}

$$BK_3=\{(1/4)\sum_l\sum_h\sum_k(((n_{h,1}\,n_{k,1})/n_1^2)+((n_{h,0}\,n_{k,0})/n_0^2))((\beta_{lk,1}-\beta_{lh,1}) +(\beta_{lh,0}-\beta_{lk,0}))[(x_{lgh,1}+x_{lgk,1})-(x_{lgh,0}+x_{lgk,0})]\} \quad \text{(B-43)}$$

for the cases where ln y_{gh1} < ln y_{gk1} and ln y_{gh0} > ln y_{gk0}

$$BL_1=\{(1/4)\sum_l\sum_h\sum_k(((n_{h,1}\,n_{k,1})/n_1^2)+((n_{h,0}\,n_{k,0})/n_0^2))((x_{lgh,1}+x_{lgk,1}) +(x_{lgh,0}+x_{lgk,0}))[(\beta_{lh,1}-\beta_{lk,1})-(\beta_{lh,0}-\beta_{lk,0})]\} \quad \text{(B-44)}$$

for the cases where ln y_{gh1} > ln y_{gk1} and ln y_{gh0} > ln y_{gk0}

$$BL_2=\{(1/4)\sum_l\sum_h\sum_k(((n_{h,1}\,n_{k,1})/n_1^2)+((n_{h,0}\,n_{k,0})/n_0^2))((x_{lgh,1}+x_{lgk,1}) +(x_{lgh,0}+x_{lgk,0}))[(\beta_{lh,1}-\beta_{lk,1})-(\beta_{lk,0}-\beta_{lh,0})]\} \quad \text{(B-45)}$$

for the cases where ln y_{gh1} > ln y_{gk1} and ln y_{gh0} < ln y_{gk0}

$$BL_3=\{(1/4)\sum_l\sum_h\sum_k(((n_{h,1}\,n_{k,1})/n_1^2)+((n_{h,0}\,n_{k,0})/n_0^2))((x_{lgh,1}+x_{lgk,1}) +(x_{lgh,0}+x_{lgk,0}))[(\beta_{lk,1}-\beta_{lh,1})-(\beta_{lh,0}-\beta_{lk,0})]\} \quad \text{(B-46)}$$

for the cases where ln y_{gh1} < ln y_{gk1} and ln y_{gh0} > ln y_{gk0}.

Combining finally expressions (B-15) to (B-46), we derive that the variation ($\Delta_{B,1} - \Delta_{B,0}$) that occurs over time in the between-groups dispersion may be written as

$$\Delta_{B,1} - \Delta_{B,0} = (G + J) + (H + K) + (I + L) \quad \text{(B-47)}$$

where (G + J), (H + K), (I + L) measure, respectively, the effect of changes over time in the relative size of the different groups, the impact of variations over time in the average value of the different explanatory variables in the various groups, and the role played by modifications over time in the regression coefficients in the different groups.

Note that the effect (H + K) of changes in the value of the explanatory variables may be decomposed into two parts and interpreted as follows: the impact (H) of variations over time in the dispersion of these explanatory variables and the role (K) of modifications taking place over time in the average values of these variables.

Similarly, the effect (I + L) of changes in the regression coefficients may be decomposed into two parts and interpreted as follows: the impact (L) of variations over time in the dispersion over the different groups of these regression coefficients and the role (I) played by the modifications that take place over time in the average values of these coefficients.

3) Decomposing the Change Over Time in the Overlapping Component

It is important at this stage to remember what this component refers to. Let us assume that the mean ln y_{gh} of the logarithms of the incomes in group h is higher than the corresponding mean ln y_{gk} in group k. Despite the fact that on average the incomes in group h are higher than those in group k, one cannot exclude the possibility that some individuals belonging to group k have a higher income than some other individuals who are members of group h. This is the essence of the concept of overlapping between two income distributions. In giving the correct expression for the change in the degree of overlapping we will try to determine the role played by changes in the value of the explanatory variables, in the regression coefficients, and in the unobserved components.

Using expressions (28) through (33), the change M in the degree of overlapping may be expressed as

$$M = \Delta_{p,1} - \Delta_{p,0} = P + Q + R \qquad \text{(B-48)}$$

where

$$P=P_1+P_2+P_3 \qquad \text{(B-49)}$$

$$Q=Q_1+Q_2+Q_3 \qquad \text{(B-50)}$$

$$R=R_1+R_2+R_3 \qquad \text{(B-51)}$$

and

$$P_1=(1/2)\,(1/n_1^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,1}+\beta_{lh,1})(x_{ljk,1}-x_{lih,1})]$$
$$-(1/2)\,(1/n_0^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,0}+\beta_{lh,0})(x_{ljk,0}-x_{lih,0})] \quad \text{(B-52)}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{jk1} > \ln y_{ih1}$;

$$P_2= -\,(1/2)\,(1/n_1^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,1}+\beta_{lh,1})(x_{ljk,1}-x_{lih,1})]$$
$$-(1/2)\,(1/n_0^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,0}+\beta_{lh,0})(x_{ljk,0}-x_{lih,0})] \quad \text{(B-53)}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} < \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{ih1} > \ln y_{jk1}$;

$$P_3= (1/2)\,(1/n_1^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,1}+\beta_{lh,1})(x_{ljk,1}-x_{lih,1})]$$
$$+\,(1/2)\,(1/n_0^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,0}+\beta_{lh,0})(x_{ljk,0}-x_{lih,0})] \quad \text{(B-54)}$$

for $\ln y_{gh0} < \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{ih0} > \ln y_{jk0}$, $\ln y_{jk1} > \ln y_{ih1}$;

and

$$Q_1=(1/2)\,(1/n_1^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,1}-\beta_{lh,1})(x_{ljk,1}+x_{lih,1})]$$
$$-(1/2)\,(1/n_0^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,0}-\beta_{lh,0})(x_{ljk,0}+x_{lih,0})] \quad \text{(B-55)}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{jk1} > \ln y_{ih1}$;

$$Q_2= -\,(1/2)\,(1/n_1^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,1}-\beta_{lh,1})(x_{ljk,1}+x_{lih,1})]$$
$$-(1/2)\,(1/n_0^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,0}-\beta_{lh,0})(x_{ljk,0}+x_{lih,0})] \quad \text{(B-56)}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} < \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{ih1} > \ln y_{jk1}$;

$$Q_3=(1/2)\,(1/n_1^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,1}-\beta_{lh,1})(x_{ljk,1}+x_{lih,1})]$$
$$+\,(1/2)\,(1/n_0^2)\,\sum_{l=1\text{ to }L}\sum_{h=1\text{ to }m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,0}-\beta_{lh,0})(x_{ljk,0}+x_{lih,0})] \quad \text{(B-57)}$$

for $\ln y_{gh0} < \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{ih0} > \ln y_{jk0}$, $\ln y_{jk1} > \ln y_{ih1}$;

and

$$R_1=(1/n_1^2)\,\sum_{h=1\text{ to }m}\,\sum_{i\in h}\,\sum_{k\neq h}\,\sum_{j\in k}[(\,u_{jk,1}-u_{ih,1}\,)]$$
$$-\,(1/n_0^2)\,\sum_{h=1\text{ to }m}\,\sum_{i\in h}\,\sum_{k\neq h}\,\sum_{j\in k}[(\,u_{jk,0}-u_{ih,0}\,)] \quad \text{(B-58)}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{jk1} > \ln y_{ih1}$;

$$R_2= (1/n_1^2)\,\sum_{h=1\text{ to }m}\,\sum_{i\in h}\,\sum_{k\neq h}\,\sum_{j\in k}[(\,u_{jk,1}-u_{ih,1}\,)]$$
$$-(1/n_0^2)\,\sum_{h=1\text{ to }m}\,\sum_{i\in h}\,\sum_{k\neq h}\,\sum_{j\in k}[(\,u_{jk,0}-u_{ih,0}\,)] \quad \text{(B-59)}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} < \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{ih1} > \ln y_{jk1}$;

$$R_3=(1/n_1^2)\,\sum_{h=1\text{ to }m}\,\sum_{i\in h}\,\sum_{k\neq h}\,\sum_{j\in k}[(\,u_{jk,1}-u_{ih,1}\,)] \quad \text{(B-60)}$$

$+ (1/n_0^2) \sum_{h=1 \text{ to } m} \sum_{i\in h} \sum_{k\neq h} \sum_{j\in k} [(u_{jk,0} - u_{ih,0})]$

for $\ln y_{gh0} < \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{ih0} > \ln y_{jk0}$, $\ln y_{jk1} > \ln y_{ih1}$.

Here again expressions (B-52) through (B-60) may be further decomposed by writing that

$$P_1 = S_1 + T_1 + TT_1 \quad \text{(B-61)}$$

$$P_2 = S_2 + T_2 + TT_2 \quad \text{(B-62)}$$

$$P_3 = S_3 + T_3 + TT_3 \quad \text{(B-63)}$$

and

$$Q_1 = U_1 + V_1 + VV_1 \quad \text{(B-64)}$$

$$Q_2 = U_2 + V_2 + VV_2 \quad \text{(B-65)}$$

$$Q_3 = U_3 + V_3 + VV_3 \quad \text{(B-66)}$$

where

$$S_1=\{(1/4)\sum_{l=1\text{to }L} \sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k} [(\beta_{lk,1}+\beta_{lh,1}) +(\beta_{lk,0}+\beta_{lh,0})] [(1/n_1^2)(x_{ljk,1}-x_{lih,1}) -(1/n_0^2)((x_{ljk,0}-x_{lih,0})]\} \quad \text{(B-67)}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{jk1} > \ln y_{ih1}$;

$$S_2= \{(1/4)\sum_{l=1\text{to }L} \sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k} [(\beta_{lk,1}+\beta_{lh,1}) +(\beta_{lk,0}+\beta_{lh,0})][(1/n_1^2)(x_{lih,1}-x_{ljk,1}) - (1/n_0^2)(x_{ljk,0}-x_{lih,0})]\} \quad \text{(B-68)}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} < \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{ih1} > \ln y_{jk1}$;

$$S_3=\{(1/4)\sum_{l=1\text{to }L} \sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k} [(\beta_{lk,1}+\beta_{lh,1}) +(\beta_{lk,0}+\beta_{lh,0})][(1/n_1^2) (x_{ljk,1}-x_{lih,1}) -(1/n_0^2)(x_{lih,0}-x_{ljk,0})]\} \quad \text{(B-69)}$$

for $\ln y_{gh0} < \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{ih0} > \ln y_{jk0}$, $\ln y_{jk1} > \ln y_{ih1}$;

and

$$T_1=\{(1/4)\sum_l\sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(1/n_1^2)(x_{ljk,1}-x_{lih,1}) + (1/n_0^2)(x_{ljk,0}-x_{lih,0})] [(\beta_{lk,1}+\beta_{lh,1})-(\beta_{lk,0}+\beta_{lh,0})]\} \quad \text{(B-70)}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{jk1} > \ln y_{ih1}$;

$$T_2=\{(1/4)\sum_l\sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(1/n_1^2)(x_{lih,1}-x_{ljk,1}) + (1/n_0^2)(x_{ljk,0}-x_{lih,0})]\,[(\beta_{lk,1}+\beta_{lh,1})-(\beta_{lk,0}+\beta_{lh,0})]\} \tag{B-71}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} < \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{ih1} > \ln y_{jk1}$;

$$T_3=\{(1/4)\sum_l\sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(1/n_1^2)(x_{lih,1}-x_{ljk,1}) + (1/n_0^2)((x_{lih,0}-x_{ljk,0})]\,[(\beta_{lk,1}+\beta_{lh,1})-(\beta_{lk,0}+\beta_{lh,0})]\} \tag{B-72}$$

for $\ln y_{gh0} < \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{ih0} > \ln y_{jk0}$, $\ln y_{jk1} > \ln y_{ih1}$.

Similarly we have

$$U_1=\{(1/4)\sum_l\sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,1}-\beta_{lh,1})+(\beta_{lk,0}-\beta_{lh,0})][\,(1/n_1^2)(x_{ljk,1}+x_{lih,1}) -(1/n_0^2)(x_{ljk,0}+x_{lih,0})]\} \tag{B-73}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{jk1} > \ln y_{ih1}$;

$$U_2=\{(1/4)\sum_l\sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lh,1}-\beta_{lk,1})+(\beta_{lk,0}-\beta_{lh,0})][\,(1/n_1^2)(\,(x_{ljk,1}+x_{lih,1}) - (1/n_0^2)(\,(x_{ljk,0}+x_{lih,0})]\} \tag{B-74}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} < \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{ih1} > \ln y_{jk1}$;

$$U_3=\{(1/4)\sum_l\sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(\beta_{lk,1}-\beta_{lh,1})+(\beta_{lh,0}-\beta_{lk,0})][\,(1/n_1^2)(x_{ljk,1}+x_{lih,1}) - (1/n_0^2)((x_{ljk,0}+x_{lih,0})]\} \tag{B-75}$$

for $\ln y_{gh0} < \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{ih0} > \ln y_{jk0}$, $\ln y_{jk1} > \ln y_{ih1}$;

and

$$V_1=\{(1/4)\sum_l\sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(1/n_1^2)(x_{ljk,1}+x_{lih,1}) +(1/n_0^2)(x_{ljk,0}+x_{lih,0})]\,[(\beta_{lk,1}-\beta_{lh,1})-(\beta_{lk,0}-\beta_{lh,0})]\} \tag{B-76}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{jk1} > \ln y_{ih1}$;

$$V_2=\{(1/4)\sum_l\sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(1/n_1^2)(x_{ljk,1}+x_{lih,1}) + (1/n_0^2)(x_{ljk,0}+x_{lih,0})]\,[(\beta_{lh,1}-\beta_{lk,1})-(\beta_{lk,0}-\beta_{lh,0})]\} \tag{B-77}$$

for $\ln y_{gh0} > \ln y_{gk0}$, $\ln y_{gh1} < \ln y_{gk1}$, $\ln y_{jk0} > \ln y_{ih0}$, $\ln y_{ih1} > \ln y_{jk1}$;

$$V_3=\{(1/4)\sum_l\sum_{h=1 \text{ to } m}\sum_{i\in h}\sum_{k\neq h}\sum_{j\in k}[(1/n_1^2)(x_{ljk,1}+x_{lih,1}) + (1/n_0^2)(x_{ljk,0}+x_{lih,0})]\,[(\beta_{lk,1}-\beta_{lh,1})-(\beta_{lh,0}-\beta_{lk,0})]\} \tag{B-78}$$

for $\ln y_{gh0} < \ln y_{gk0}$, $\ln y_{gh1} > \ln y_{gk1}$, $\ln y_{ih0} > \ln y_{jk0}$, $\ln y_{jk1} > \ln y_{ih1}$.

Combining expressions (B-48) through (B-78), we finally derive that

$$M = W + X + R \tag{B-79}$$

where

$$W = (S_1+S_2+S_3) + (U_1+U_2+U_3) \tag{B-80}$$

$$X = (T_1+T_2+T_3) + (V_1+V_2+V_3) \tag{B-81}$$

$$R = (R_1+R_2+R_3) \tag{B-82}$$

where W, X, and R measure, respectively, the effect of changes in the value of the explanatory variables, in the regression coefficients, and in the unobserved components. Additional distinctions may be made by noting that $(S_1+S_2+S_3)$ measures the effect of changes in the dispersion of the values of the explanatory variables among those involved in the overlapping in each periods, while $(U_1+U_2+U_3)$ corresponds somehow to the impact of variations over time in the average values of these explanatory variables among those same individuals.

Similarly, $(V_1+V_2+V_3)$ measures the effect of changes over time in some weighted dispersion of the regression coefficients, these weights depending only on those individuals involved in the overlapping, while $(T_1+T_2+T_3)$ corresponds to the impact of the variation over time in some weighted average of these regression coefficients, here again the weights depending only on those individuals who are part of the overlapping. Finally, R measures the change over time in the dispersion of the unobserved components among those individuals involved in the overlapping in both periods.