

to Prof. Heckman

THE HEBREW UNIVERSITY OF JERUSALEM

DEPARTMENT OF ECONOMICS

On Using Linear Regressions in
Welfare Economics*

By

Shlomo Yitzhaki

Working Paper #217

November 1989

ABSTRACT

This paper provides a new interpretation of the slope of an OLS regression, as a weighted average of the slopes between adjacent sample points. The weights depend only on the distribution of the independent variable. When applied to a linear regression with income as the independent variable, due to the skewness of income, the regression coefficient depends heavily on the behavior of high income groups. The contribution of the highest income decile to determining the regression coefficient may well exceed the contribution of the other nine deciles of the population. This may be particularly undesirable if one seeks to use the regression for welfare purposes. An alternative criterion which enables the investigator to control the weighting scheme is suggested. The alternative criterion is based on Gini's mean difference.

* I am grateful to Joram Mayshar and Joseph Yahav for many helpful discussions and to Giora Hanoach, Robert Lerman and Ingram Olkin for their comments.

On Using Linear Regressions in Welfare Economics

Linear regressions, with income as an independent variable, are widely used in economics. The linear expenditure system, and particularly the linear Engel curve are examples of such a use. These regressions are convenient and appear in almost every Econometrics text book.¹

The paper has two aims: The first is to point out the hidden price to achieving this convenience. Assuming a linear model and using Ordinary Least Squares regression (OLS) mean that the estimated regression coefficient relies heavily on the behavior of high income groups. The contribution of the highest decile to the OLS estimator of the regression coefficient may be higher than that of the rest of the population. This property of the OLS estimator is especially disturbing in welfare economics, where typical social concerns cause people to attach more weight to low income groups. The second aim of the paper is to suggest an alternative criterion which overcomes the over weighting problem.

In the first section, the OLS estimator of the regression coefficient is presented as a weighted average of slopes defined by adjacent observations in the sample. The implications of the OLS weighting scheme in a sample, with income as an independent variable, are discussed. The second section explores

¹ For example Robert Moffitt (1989) uses linear Engel curves, (and other functional forms), in order to estimate the value of in-kind transfers. Angus Deaton (1977) assumes linear Engel curves in order to derive a simple formula for optimal tax rates. Deaton, however, does not use OLS estimation procedure. Using an additional assumption (the Atkinson's type welfare function) enables him to directly estimate the parameters by two summary statistics.

the connection between the distribution of the independent variable and the weighting scheme, leading us to the conclusion that the findings of the first section are not caused by a few outliers. We should expect them to appear in almost every cross-section sample with income as an independent variable. The third section presents alternative estimators, presents their weighting scheme and the implications for welfare economics.

I. The OLS Linear Regression.

The purpose of this section is to present the OLS estimator as a weighted average of slopes defined by adjacent observations in the sample. We start with a simple regression; later we show that the same argument can be extended to multiple regression. As an illustration it is assumed that one is interested in the marginal propensity to spend, where consumption is regressed against income. That is, the underlined model is

$$(1) \quad Y = \alpha + \beta X + \epsilon,$$

where Y is consumption and X is income.

Let y_i, x_i ($i=1, \dots, n$) be n observations. Assume that observations are ordered according to an increasing order of X , the independent variable. Let $dx_i = x_{i+1} - x_i$ ($dx_i > 0$)² and $b_i = (y_{i+1} - y_i) / (x_{i+1} - x_i)$ ($i=1, \dots, n-1$), be the income difference and the slope defined by two adjacent observations.

Proposition 1: The OLS estimator of β is a weighted average of slopes defined by adjacent observations. That is,

$$(2) \quad b_{OLS} = \sum_{i=1}^{n-1} w_i b_i \quad \text{where } w_i > 0, \quad \sum_{i=1}^{n-1} w_i = 1.$$

The weights are given by

² The assumption $dx_i > 0$ simplifies the presentation.

$$(3) \quad w_i = \left(\sum_{j=i}^{n-1} i(n-j) dx_j + \sum_{j=1}^{i-1} j(n-i) dx_j \right) dx_i / \sum_{k=1}^{n-1} \left(\sum_{j=i}^{n-1} k(n-j) dx_j + \sum_{j=1}^{n-1} j(n-k) dx_j \right) dx_k$$

Proof: The OLS estimator is

$$(4) \quad b_{OLS} = \text{Cov}(Y,X) / \text{Cov}(X,X)$$

For our purposes it is convenient to express the numerator and the denominator in an alternative way. The numerator can be rewritten as,

$$(5) \quad \text{Cov}(Y,X) = (1/2) E_1 E_2 (Y_1 - Y_2)(X_1 - X_2),$$

where (Y_i, X_i) ($i=1,2$) are i.i.d variables and E denotes expectation. Ignoring multiplicative constants, (which cancel when both the numerator and the denominator are considered), the application of this formula in the sample leads to:

$$(6) \quad \text{Cov}(y,x) = \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)(y_i - y_j),$$

and by substituting $y_i - y_j = \sum_{k=s}^{t-1} b_k dx_k$ and $x_i - x_j = \sum_{p=s}^{t-1} dx_p$, where $s = \text{Min}(i,j)$

and $t = \text{Max}(i,j)$, we get

$$\text{Cov}(y,x) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=s}^{t-1} \sum_{p=s}^{t-1} b_k dx_k dx_p,$$

and after some tedious algebra we get

$$(7) \quad \text{Cov}(y,x) = \sum_{i=1}^{n-1} \left\{ \sum_{j=i}^{n-1} i(n-j) dx_j + \sum_{j=1}^{i-1} j(n-i) dx_j \right\} dx_i b_i$$

Applying the same procedure to the denominator we get

$$(8) \quad \text{Cov}(x, x) = \sum_{i=1}^{n-1} \left(\sum_{j=i}^{n-1} i(n-j) dx_j + \sum_{j=1}^{i-1} j(n-i) dx_j \right) dx_i$$

Dividing (7) by (8) yields equation (2) and (3). QED.

The two components of Equation (2) are the slopes b_i and the weights w_i . It is important to note that the weighting scheme w_i depends only on the distribution of the independent variable X . As shown later, one can interpret the difference between the OLS estimator and alternative estimators as originating from the weighting scheme employed.

The weight w_i depends both on the rank of the observations and its distance from other observations. The weight is larger, the closer the observation is to the median and the larger is the distance between the observations. To control for the latter effect consider the case where the observations of the independent variable are equidistant, that is $dx_i = c$ for all i . In this case the weights in (3) become:

$$(9) \quad w(i) = K i(n-i), \quad \text{where } K = 6 / \{n(n-1)(n+1)\}.$$

This weighting scheme is symmetric around the median and the closer the observation to the median, the higher the weight. That is the weighting scheme is bell-shaped. This weighting scheme is presented in Column (1) of Table 1. The top decile receives less than 3 percent of the weights. If dx_i varies along the distribution, then the larger dx_i , the larger the weight.

Column (2) in Table 1 presents the empirical weighting scheme in a sample with income as an independent variable. The sample used is the Israeli Survey of Family Expenditure 1979/80. In order to reduce the variability due to family size, only observations of households with two members are used.³

³ This group was chosen because the number of observations (531) is the largest. Other groups have similar patterns of weighting schemes. By restricting the sample to one group, the variability of the sample is

As can be seen, the OLS estimator assigns around 60 percent of the weights to the richest decile; over three quarters of the total weight is given to the top three deciles. To see the implication of this empirical weighting scheme consider the following hypothetical unrealistic case. Assume that the model is misspecified, it is not linear but is rather composed of two linear sections. The marginal propensity to spend is a constant +0.1 for the poorest 90 percent of the population and -0.1 for the top decile. The OLS estimate of the marginal propensity to spend will then be -0.016, that is, the commodity will be considered as an inferior commodity, although for ninety percent of the population the marginal propensity to spend is positive.⁴

Note that the same weighting scheme is inherited to all statistical tests. For example, a misspecification test will rely on the same weighting scheme. Therefore, it can't give assurances that low income groups are not ignored. Examining for undue influence may be helpful. However, the influence of an observation is composed of two components: the weight and the deviation of the observation. If only observations with low weight deviate from the average slope then the influence of each observation may be small, some of them because of the low weight and others because of the small deviation.

Further insight on the effect of the distribution of the independent variable on the weighting scheme of the OLS regression can be gained by investigating the implied weighting scheme of specific distributions. This will enable us to establish a formal relationship between the OLS weighting scheme and the distribution of the independent variable.

reduced.

⁴ In the equidistant case the OLS estimate is .094.

Table 1: OLS Weighting Schemes of Adjacent slopes according to income deciles.^a

Income Decile	Equidistant Weights ^b (1)	Empirical Weights (Whole Sample) (2)	Empirical Weights (Truncated Sample) ^c (3)
1 (lowest)	.028	.005	.014
2	.076	.013	.036
3	.112	.024	.058
4	.136	.032	.091
5	.148	.037	.093
6	.148	.050	.113
7	.136	.072	.150
8	.112	.080	.169
9	.076	.099	.168
10	.028	.588	.109
All	1.0	1.0	1.0

Source: Tabulation from the Israeli Survey of Family Expenditure, households with two members only. (531 observations.)

a: Income is defined as before tax income, including imputed rent on own housing.

b: The weighting scheme for a sample with observations with equal distance.

c: Only 500 observations: 31 highest observations dropped from sample

II. The Weighting Scheme and the Distribution of the Independent Variable.

This section investigates the relationship between the distribution of the independent variable and the weighting scheme of the OLS estimator. We show that the pattern of weights presented in column (2) of Table 1 is typical of all linear models with income as an independent variable.

Let Y, X be two random variables with a density function $f(y, x)$. Let f_X , F_X and μ_X denote the marginal density, the marginal cumulative distribution and the expected value of X , respectively. Assume also that first and second moments exist. Let $g(x) = E(Y|X=x)$ be the regression curve, while $g'(x)$ represents its slope.

Proposition 2: The OLS estimator of the regression coefficient b , in the simple linear regression $Y = a + bX$, is a weighted average of the slopes of the regression curve:

$$(10) \quad b_{OLS} = \int w(x)g'(x)dx, \text{ where } w(x) > 0 \text{ and } \int w(x)dx = 1.$$

The weights are:

$$(11) \quad w(x) = (1/\sigma_X^2) (\mu_X F_X(x) - \Theta_X(x)),$$

where σ_X^2 is the variance and,

$$(12) \quad \Theta_X(x) = \int_{-\infty}^x t f_X(t) dt.$$

Proof: Note that $b_{OLS} = \text{Cov}(Y, X) / \text{Cov}(X, X)$. The numerator can be expressed as:

$$(13) \quad \begin{aligned} \text{Cov}(Y, X) &= E_X E_Y \{ (Y - \mu_Y)(X - \mu_X) \} = E_X E_Y \{ (X - \mu_X)Y \} = E_X (X - \mu_X) E_Y (Y|X=x) \\ &= \int (x - \mu_X) g(x) f_X(x) dx, \end{aligned}$$

where $g(x) = E_Y(Y|X=x)$ is the conditional expectation.

Using integration by parts with $v'(t) = (x - \mu_X)f_X(t)$;

$$v(x) = \int_{-\infty}^x (t - \mu_X)f_X(t)dt \text{ and } u(x) = g(x); \quad u'(x) = g'(x),$$

$$(14) \text{Cov}(Y,X) = \left\{ \int_{-\infty}^x (t - \mu_X) f_X(t) dt \right\} g(x) \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \int_{-\infty}^x (\mu_X - t) f_X(t) dt g'(x) dx .$$

Provided that second moments exist, the first term converges to zero.⁵ Hence equation (13) can be written as:

$$(15) \text{Cov}(Y,X) = \int (\mu_X F_X(x) - \theta_X(x)) g'(x) dx .$$

To show that the sum of the weights is equal one we apply the same procedure to the denominator of the OLS regression coefficient. Then

$$(16) \sigma_X^2 = \text{Cov}(X,X) = \int (\mu_X F_X(x) - \theta_X(x)) dx . \text{ QED.}$$

Since $w(x)$ is a function of the distribution of the independent variable, there are two possible presentations of the weighting scheme. The first is the presentation of w as a function of X the independent variable. The second is the presentation of w as a function of F_X , the cumulative distribution of the independent variable.⁶ The second presentation is useful whenever the interest is in the share of the weights that is assigned to portions of the population. The transformation from one presentation to the other can be done by defining $v(F) = w[x^{-1}(F)]$ as the weighting scheme.

To illustrate the effect of the distribution of the independent variable on the weighting scheme, we now consider three specific examples; the uniform, the normal and the lognormal. The first two are intended to show interesting cases, while the third resembles the distribution of income.

(a) The Uniform Distribution. Let X be uniformly distributed between $[a,b]$.

Applying (11) the weight attached to the slope at income x is

⁵ The relative weight attached to different deciles will not be affected even if the first term converges to a constant.

⁶ Since $F_X(x)$ is a monotonic increasing differentiable function, the inverse function of $F_X(x)$ is always defined.

$$(17) \quad w(x) = (b-x)(x-a) / 2(b-a).$$

This weighting scheme is similar to the one in (9) above for the equidistant sample. It is symmetric around the median and the closer the observation to the median the higher the weight. An interesting feature of the weighting scheme is that its shape is identical whether it is viewed as a function of x or as a function of F .

(b) The Normal Distribution. Let X be distributed according to the Standard Normal with $\mu=0$. Hence, by (11),

$$(18) \quad w(x) = - (1/\sqrt{2\pi}) \int_{-\infty}^x t e^{-t^2/2} dt = (1/\sqrt{2\pi}) e^{-x^2/2}.$$

The weight as a function of the independent variable is bell shaped like the density function. On the other hand, since the weight is equal to the density function, each decile of the population receives an equal weight of 10 percent.

(c) The Lognormal Distribution. Let X be lognormally distributed with μ and σ . In this case it is convenient to write (11) in a slightly different way.

$$(19) \quad w(x) = (\mu_X / \sigma_X^2) [F_X(x) - F_1(x)]$$

where μ_X , σ_X^2 are the expected value and the variance of the lognormal distribution while $F_1(x) = (1/\mu_X) \int_0^x t dF_X(t)$ is the first moment

cumulative distribution. By Theorem 2.6 in Aitchison and Brown (1957, p.12) the first moment distribution is also lognormal with parameters $\mu + \sigma^2$ and σ^2 . Hence the weight at x is the difference between two cumulative lognormal distributions. Using the usual transformation we can write the weight as:

$$(20) \quad w(x) = (\mu_X / \sigma_X^2) \{ \Phi[(\log x - \mu) / \sigma] - \Phi[(\log x - \mu - \sigma^2) / \sigma] \},$$

where $\Phi()$ is the cumulative standard normal. This term can be numerically evaluated.

Table 2 presents the weighting scheme $w(x)$, for the case where X is lognormally distributed, for different values of the parameters of μ and σ . It turns out that the weighting scheme is not sensitive to μ . To see this, notice the difference between columns 4 and 5. However, the weighting scheme is sensitive to σ . For ease of reference, the third line in Table 2 presents the Gini coefficient which corresponds to each value of σ . As can be seen from column 4, if the Gini coefficient is around 0.4, which is a typical value for before-tax income, the expected value of the weight for the top decile is around 45 percent, and for the highest quintile it would be more than 60 percent. If wealth were used as an independent variable then a Gini coefficient of 0.55 may be considered as typical. In this case the weight given to the top decile may well exceed 60 percent. Experiments with other distributions (the Pareto and the exponential) show that the share of the top decile is not lower than 30 percent.⁷

Some intuition into the causes of this weighting scheme can be gained by looking at the share of the range of the overall distribution that each decile occupies in the sample.⁸ The share of the top decile in the range of the independent variable is around 85 percent. The differences in income at the top decile are large enough to outweigh the tendency of the OLS to attach

⁷ One reason for the difference between the weighting scheme of the theoretical distributions and the sample is that the sample is a stratified one. Since the stratification is not correlated with the dependent variable, the usual procedure is to ignore the sample weights when running an OLS regression. However, the sample weights affect the weighting scheme.

⁸ Note, however, that the weighting scheme is affected by the square of the distance between the observations. Hence, the share of the range does not reveal the whole effect of the distance between observations.

a higher weight, the closer the observation to the median of the distribution. The same argument explains the share of the lowest decile. Its share in the range is 1.2 percent. However, since it is far away from the median, its share in the OLS estimator is lower than 1.2 percent.

Before we conclude this section two additional points are worth making: (a). Should we present the OLS weighting scheme of a regression with a quadratic income variable, such as the Quadratic Expenditure System [Pollak and Wales (1978)], then the dependence of the OLS estimates on the behavior of the high income groups would be much larger than in the case of a linear model. (b) Having established the weighting scheme of the OLS estimator in the simple regression case, let me show that these properties of the OLS estimator carry through to the multiple regression case.

Let Y be the dependent variable while X_1, \dots, X_m be the independent variables. The OLS estimator is

$$b_{OLS} = (x'x)^{-1} x'y$$

where y , x denote the vector and the matrix of observations. Dividing and multiplying each term in $x'y$ by the appropriate variance enables us to write the OLS estimator as a weighted average of simple OLS estimators.⁹ The weighting scheme of each simple OLS estimator has already been established; hence, the properties of the weighting scheme carry through the multiple regression case too.

⁹. Let λ_1 be a diagonal matrix with $1/\sigma_i^2$ ($i=1, 2, \dots, m$) as the elements of the main diagonal and λ_2 a diagonal matrix with σ_i^2 as the elements. Then
$$b_{OLS} = (x'x)^{-1} \lambda_2 \lambda_1 x'y.$$
 Since $\lambda_1 x'y$ is the vector of simple regression coefficients, the Multiple regression OLS estimator is a weighted average of simple OLS estimators.

Should the OLS weighting scheme disturb us? It is clear that if the model is truly linear, it does not matter which decile has the lion's share in the weighting scheme. However, if the model is not linear, then the implication of the weighting scheme is that a welfare economist who designs optimal taxes with the intention of affecting the distribution of income, determines the distributional characteristics of the commodities mainly by the behavior of the top decile. Low income groups are ignored. The next section offers alternative estimators which enable the welfare economist to stress those segments of the income distribution he is interested in.

Table 2: The OLS Weighting Schemes for a Lognormal Distribution.

Parameters:								
μ	0.	0.	0.	0.	7.0	0.	0.	0.
σ	0.45	0.6	0.7	0.75	0.75	0.8	1.0	1.2
Gini Coefficient	0.25	0.33	0.38	0.4	0.4	0.43	0.52	0.60

Column	1	2	3	4	5	6	7	8
Decile 1	2.5	2.1	1.0	.8	.8	.7	0.3	.1
2	3.9	3.4	1.9	1.6	1.6	1.4	0.7	.3
3	5.0	4.5	2.8	2.4	2.5	2.1	1.1	.5
4	6.0	5.5	3.7	3.3	3.3	2.9	1.7	.9
5	7.1	6.6	4.8	4.3	4.4	3.9	2.4	1.3
6	8.8	8.3	6.5	6.0	6.0	5.5	3.6	2.1
7	10.2	9.9	8.3	7.8	7.9	7.3	5.2	3.3
8	12.6	12.5	11.4	11.0	11.1	10.5	8.2	5.7
9	16.1	16.4	16.9	16.7	17.0	16.5	14.4	11.2
10	27.9	30.4	42.8	46.0	45.3	49.3	62.6	74.7

III. Alternative Estimators.

Propositions 1 and 2 provide a particular formalization for the well known perception that OLS estimators are sensitive to extreme observations. A common procedure to deal with that sensitivity is to drop extreme observations from the sample. However, this procedure seems to suffer from internal inconsistency. Outlying observations are given by the OLS extremely high weights, yet when it turns out that they are strongly affecting the estimate they are arbitrarily reassigned a weight of zero. In any case, dropping a few extreme observations may not be very helpful in a sample with income as an independent variable. Experiments with the sample in Table 1 show that in order to reduce the weight of the top decile to around 10 percent we have to drop the top 5 percent of the sample. Column 3 in Table 1 presents the weighting scheme for such a truncated sample.

An alternative way is to use indices of variability that are less sensitive to extreme observations and that imply estimators of the slope of a regression line with lower (but not zero) weight attached to extreme observations. The aim of the rest of this section is to present the weighting schemes of the (extended) Gini estimators. These estimators are used whenever a decomposition of the Gini coefficient is carried out. [See Donaldson and Weymark (1983), Chakravarty (1988) and Yitzhaki (1983) for the definition and description of the properties of the extended Gini. Lerman and Yitzhaki (1985), and Yitzhaki (1988) are examples of its use in welfare economics.]

The (extended) Gini regression coefficient is defined as:¹⁰

¹⁰ See Yitzhaki and Olkin (1988) for the definition, and Olkin and Yitzhaki (1988) for investigations of the properties of one member of the family.

$$(21) \quad b_{NP}(\nu) = \text{Cov}(Y, [1-F_X(X)]^{\nu-1}) / \text{Cov}(X, [1-F_X(X)]^{\nu-1}) \quad \nu > 0, \nu \neq 1.$$

where ν is a parameter determined by the investigator. In the sample F_X is estimated by the empirical cumulative distribution. All estimators defined in (21) are based on the extended Gini variability index. The denominator is the extended Gini variability index while the numerator is the extended Gini covariance.

Proposition 3: The extended Gini estimators of the regression coefficient have the following properties:

(a) All estimators are weighted averages of the slopes of the regression curve. That is,

$$(22) \quad b_{NP}(\nu) = \int v(x, \nu) g'(x) dx, \quad \text{with } v(x, \nu) > 0 \text{ and } \int v(x, \nu) dx = 1, \text{ where}$$

$$(23) \quad v(x, \nu) = \{(1-F_X(x)) - [1-F_X(x)]^\nu\} / \int \{(1-F_X(t)) - [1-F_X(t)]^\nu\} dt.$$

(b) In the sample, all estimators are weighted averages of slopes defined by adjacent observations.

$$(22)' \quad b_{NP}(\nu) = \sum_{i=1}^{n-1} v_i b_i \quad v_i > 0, \sum v_i = 1, \text{ where}$$

$$(23)' \quad v_i = \{n^{\nu-1}(n-i) - (n-i)^\nu\} dx_i / \sum_{k=1}^{n-1} \{n^{\nu-1}(n-k) - (n-k)^\nu\} dx_k.$$

(c) Given the assumptions of the OLS model hold, namely that in the population $Y = \alpha + \beta X + \epsilon$, where X and ϵ are independent and $\sigma_\epsilon^2 < \infty$, all extended Gini estimators are consistent estimators of β .

Proof:

(a): It is convenient to present the numerator by the following presentation:

$$(24) \quad \nu \text{ cov}(Y, [1-F_X(X)]^{\nu-1}) = \int \{(1-F_X(x)) - [1-F_X(x)]^\nu\} g'(x) dx,$$

where $g(x) = E_Y(Y|X=x)$ and $g'(x)$ is the derivative with respect to x . The

denominator can be presented in a similar form, the only difference is that $g'(x) \equiv 1$.

To derive (24) note that $E\{[1-F_X(x)]^{\nu-1}\} = 1/\nu$.

Hence

$$\begin{aligned} \nu \operatorname{cov}(Y, [1-F_X(X)]^{\nu-1}) &= E_X E_Y \{(Y - \mu_Y) ([1-F_X(X)]^{\nu-1} - (1/\nu))\} = \\ &= E_X \{([1-F_X(X)]^{\nu-1} - (1/\nu)) g(x)\} = \int ([1-F_X(x)]^{\nu-1} - (1/\nu)) g(x) f_X(x) dx. \end{aligned}$$

By using integration by parts, where

$$u(x) = g(x)$$

$$v'(x) = (1/\nu) \{ \nu [1-F_X(x)]^{\nu-1} - 1 \} f_X(x), \quad v(x) = (-1/\nu) \{ [1-F_X(x)]^\nu - (1-F_X(x)) \},$$

we get

$$\begin{aligned} \nu \operatorname{cov}(Y, [1-F_X(X)]^{\nu-1}) &= (1/\nu) g(x) \{ (1-F_X(x)) - [1-F_X(x)]^\nu \} \Big|_{-\infty}^{\infty} \\ &\quad - (1/\nu) \int \{ (1-F_X(x)) - [1-F_X(x)]^\nu \} g'(x) dx. \end{aligned}$$

The existence of the covariance insures that the first term equals zero. The final term is therefore:

$$\nu \operatorname{cov}(Y, [1-F_X(X)]^{\nu-1}) = (-1/\nu) \int \{ (1-F_X(x)) - [1-F_X(x)]^\nu \} g'(x) dx.$$

Applying to same procedure to the denominator of (21), we get,

$$(25) \quad \nu \operatorname{cov}(X, [1-F_X(X)]^{\nu-1}) = (-1/\nu) \int \{ (1-F_X(x)) - [1-F_X(x)]^\nu \} dx.$$

The final step is to divide (24) by (25). QED.

(b): The proof is similar to that in Proposition 1.

(c): Assume that in the population $Y = \alpha + \beta X + \epsilon$ and X and ϵ are independent and $\sigma_\epsilon^2 < \infty$. Then

$$\lim_{n \rightarrow \infty} b_{NP}(\nu) = \beta + \operatorname{cov}(\epsilon, [1-F_X(X)]^{\nu-1}) / \operatorname{cov}(X, [1-F_X(X)]^{\nu-1}) = \beta. \quad \text{QED.}$$

The weights are determined by the parameter ν and, of course, by the distribution of the independent variable. By determining ν the investigator introduces his preferences into the estimation procedure. To investigate the

properties of the weighting scheme for a given ν , we may ignore the denominator as a normalizing constant and concentrate on the numerator as a function of F .

$$(26) \quad w(F) = c(\nu) \{ (1-F) - (1-F)^\nu \},$$

where c is a constant determined by ν . By looking at the derivatives of w with respect to F we may learn something about on the behavior of the weighting scheme:

$$\begin{aligned} \partial w / \partial F &= c(\nu) [\nu(1-F)^{\nu-1} - 1] \\ \partial^2 w / \partial F^2 &= c(\nu) \nu (\nu-1) (1-F)^{\nu-2}. \end{aligned}$$

As can be seen for $\nu > 1$, the weighting scheme is increasing for low values of F , reaches a maximum and then declines. If $\nu < 1$, then the weighting scheme is increasing with F .¹¹

The effect of ν on the weighting scheme is more complicated since both the numerator and the denominator are affected. However, for each ν the weighting scheme can be numerically calculated.

The cases where $\nu \rightarrow 1$ and $\nu = 2$ present interesting special weighting schemes. If $\nu \rightarrow 1$, then the weights are simply the distances between adjacent observations. The weight attached to each decile is its share of the range. The estimated regression coefficient is the slope of the line connecting the first and last observations. If $\nu > 1$, higher weights are given to the lower segments of the distribution.

If $\nu = 2$, the denominator is (one half) of the Gini mean difference (GMD) while the numerator is Gini covariance. Olkin and Yitzhaki (1988) show that this estimator is closely related to Sievers's (1978) robust estimator of the regression coefficient and Scholz's (1978) weighted median regression

¹¹. The numerator and the denominator are negative.

estimator. Scholz, Sievers and Olkin and Yitzhaki showed that the sample distribution of the GMD estimators converge to the normal distribution, and they suggested estimators for the variance of the estimators.

The weighting scheme is symmetric in F , and the closer the observation to the median, the higher the weight. If the distance between observations of the independent variable is a constant, then the weighting scheme of the GMD is identical to that given by (9) above for the OLS estimator. For other distributions of the independent variable it can be shown that each decile receives its share in the income distribution as its weights.

Table 3 presents alternative weighting schemes from the sample of the Israeli income distribution. Column 1 presents the weighting scheme for $\nu \rightarrow 1$, that is when the weight is equal to the share of overall range that each decile occupies. Column 2 presents the weighting scheme of the GMD ($\nu=2$). The OLS weighting scheme is presented in Column 7. Comparison of OLS weighting scheme with the GMD weighting scheme shows that by using the Gini, the weight assigned to the behavior of the top decile declines from around 60 percent under OLS to around 25 percent. However, the weight of low income deciles is still low. Using $\nu > 2$ enables the user to increase the weight at the bottom of the distribution. The rest of Table 3 presents the weighting schemes for higher values of ν . Experiments with other samples show similar weighting schemes. As can be seen, the higher ν , the higher the weight attached to lower portions of the income distribution. The weighting schemes are bell shaped, and the higher ν the lower the decile receiving the highest weights. However, even for relatively large values of ν , the share of the highest income decile continues to be larger than its share in the population.

By choosing ν the investigator affects the weighting scheme. There may be several arguments which justify such a procedure. The first can be traced to Atkinson's (1970) argument, that preferences regarding the distribution of income, should be stated in advance when measuring income inequality. The argument in this paper is that in some applications, especially in the field of welfare economics, social preferences should affect the estimation procedure too. If, for example, one is interested in determining income elasticities of commodities, so that poverty alleviation subsidies can be optimally designed, then he should stress the lower portion of the income distribution in his estimation procedure. However, distributional weights are important only when the regression is not linear. If the regression is linear, i. e., the slope is a constant, then it does not matter which section of the income distribution is accentuated.

The second argument which justifies a specific weighting scheme may come from the application which follows the regression. If in the analysis which follows the regression, each income is expected to change in a proportional rate, then the GMD weighting scheme seems appropriate. The reason is that under the GMD estimation each deciles receives his share in income as its weight. This case is typical in demand analysis, when the investigator may be interested in forecasting the effect of an increase in income on the demand for a particular commodity. The expected increase in demand is the sum of each marginal propensity to spend multiplied by income.

The third argument is purely statistical. If one is interested in robust estimators then he may choose his weighting scheme according to the distribution of the independent variable.

Conclusion:

We have shown that OLS estimator of the regression coefficient can be viewed as a weighted average of the slopes defined by adjacent observations. The weights are sensitive to the distribution of the independent variable. In certain fields, such as welfare economics, the OLS weighting scheme seems to stress those sections of the distribution in which the investigator is the least interested. It is argued that in these cases one should substitute the OLS estimators by other estimators which will permit stressing the portion of the distribution the investigator is interested in.

Finally, it is worth noting that a similar situation may occur in finance. Assuming that the investor is risk averse and at the same time using the OLS method may result in a similar contradiction between economic theory and estimation methodology.

Table 3: Weighting Schemes according to Deciles^a

$\nu =$ Decile	1	2 (GMD)	4	8	12	16	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	.012	.014	.026	.045	.060	.071	.005
2	.009	.035	.059	.087	.101	.108	.013
3	.010	.059	.090	.113	.118	.118	.024
4	.010	.075	.103	.114	.113	.110	.032
5	.011	.082	.102	.102	.098	.096	.037
6	.013	.101	.110	.102	.097	.094	.050
7	.019	.132	.129	.114	.108	.105	.072
8	.022	.127	.110	.095	.090	.087	.080
9	.033	.129	.101	.085	.081	.079	.099
10	.744	.247	.170	.143	.136	.132	.588

Source: Tabulation from the Israeli Survey of Family Expenditure, households with two members only.

a: Income is defined as before tax income, including imputed rent on own housing.

References

- Aitchison J. and J. A. C. Brown (1957). The Lognormal Distribution, Cambridge University Press, Cambridge, Great Britain.
- Atkinson, A. B. (1970). "On the measurement of inequality," Journal of Economic Theory, 2, 244-63.
- Chakravarty S. (1988). "Extended Gini Indices of Inequality," International Economic Review, (February), 147-56.
- Donaldson D. and J. A. Weymark (1983). "Ethically Flexible Gini Indices for Income Distribution in the Continuum," Journal of Economic Theory, 29, 2, 353-58.
- Moffitt, Robert (1989). "Estimating the Value of An In-Kind Transfer: The Case of Food Stamps," Econometrica, 57, No. 2, 385-409.
- Olkin I. and S. Yitzhaki (1988). "Gini Regression Analysis," Working Paper No. 188, Dept. of Economics, Hebrew University.
- Pollak, R. A. and T. J. Wales (1978). "Estimation of Complete Demand Systems from household budget data," American Economic Review, 68, 348-359.
- Scholz Friedrich-Wilhelm (1978). "Weighted Median Regression Estimates," The Annals of Statistics, 6, No. 3, 603-9.
- Sievers, Gerald L. (1978). "Weighted Rank Statistics for Simple Linear Regression," Journal of the American Statistical Association, 73, 628-31.
- Yitzhaki, S. (1983). "On an Extension of Gini Inequality Index," International Economic Review, 24, No. 3, 617-28.
- Yitzhaki, S. (1988). "On the Progressivity of Commodity Taxation" Working Paper No. 255, Dept. Of Economics, Hebrew University.

Yitzhaki, S. and I. Olkin (1988). "Concentration Curves," Working Paper
No. 179, Dept. of Economics, The Hebrew University.