A Note on the U Hypothesis Relating Income Inequality and Economic Development

By SHERMAN ROBINSON*

A common empirical finding in the analysis of countries which have undergone economic development is that income distribution first became more unequal, and only in the later phase did it become more equal. This empirical observation has also been seen in modern developing countries—at least the increasing inequality phase—and has acquired the force of economic law. It has a name: the U hypothesis. A number of different economic explanations for the relationship have been presented involving a variety of factors such as productivity changes, differential savings behavior, exploitation of workers, and so forth.[1]

The purpose of this note is to demonstrate that the U hypothesis can be derived from a very simple model with a minimum of economic assumptions. One need only assume that the economy can be divided into two sectors with different sectoral income distributions and that there is a monotonic increase in the relative population of one of the sectors over time. These assumptions seem empirically unexceptional for a country undergoing economic development and are consistent with many models of development such as the Lewis-Fei-Ranis surplus labor models, dual economy models, and the more recent Harris-Todaro migration models.[2] The U hypothesis will be seen to be a necessary implication of the workings of such models.

Assume that the economy is divided into two sectors with different income distributions. The *log* mean and *log* variance of income in the two sectors are given by Y_1 and Y_2 and σ_1^2 and σ_2^2, respectively. Define the population shares of the two sectors as W_1 and W_2 with:

$$W_1 + W_2 = 1 \tag{1}$$

The overall *log* mean income is given by:

$$Y = W_1Y_1 + W_2Y_2 \tag{2}$$

and the overall *log* variance is given by:

$$\sigma^2 = W_1\sigma_1^2 + W_2\sigma_2^2 + W_1(Y_1 - Y)^2 + W_2(Y_2 - Y)^2 \tag{3}$$

The *log* variance is itself an increasing measure of income inequality. One need not use *log* means and *log* variances, but they are convenient since the *log* variance is a commonly used inequality measure. The arithmetic mean and variance would also do—the algebra is exactly the same.

Assuming that the within-sector distributions remain unchanged over time (σ_1^2, σ_2^2, Y_1, and Y_2 are constant), then from equation (3) inequality is a function only of sectoral population shares and overall *log* mean income. By equation (2), the overall *log* mean is itself a function of sectoral population shares.

Assume that sector 1 is the sector whose relative population share is increasing. Then, substituting (1) and (2) into (3) and doing a bit of algebra, one finally gets:

$$\sigma^2 = AW_1^2 + BW_1 + C \tag{4}$$

where
$$A = -(Y_1 - Y_2)^2$$
$$B = (\sigma_1^2 - \sigma_2^2) + (Y_1 - Y_2)^2$$
$$C = \sigma_2^2$$

If one assumes that the *log* mean incomes are different in the two sectors, then in-

* Assistant professor of economics, Princeton University. I am grateful to Irma Adelman for comments on an earlier draft and to Charles Frank and Richard Webb for useful discussions. They, of course, bear no responsibility for any remaining errors.

[1] See Irma Adelman and Cynthia Morris, p. 188, Simon Kuznets (1955, 1966), and Felix Paukert.

[2] See Kuznets (1966, 1971), W. Arthur Lewis, John Fei and Gustav Ranis, and John Harris and Michael Todaro.

equality is a quadratic function of W_1. Since $A<0$, the parabola has a maximum. As W_1 increases, inequality first increases, reaches a maximum, then decreases—precisely the U hypothesis.

There is one possible problem. Since by assumption $0 \leq W_1 \leq 1$, it is possible that the maximum value of σ^2 occurs for a value of W_1 outside the zero to one range. Setting the first derivative of (4) equal to zero, the maximum value of σ^2 occurs when W_1 is equal to $\hat{W}_1$:

$$\hat{W}_1 = \frac{\sigma_1^2 - \sigma_2^2}{2 \cdot (Y_1 - Y_2)^2} + \frac{1}{2} \tag{5}$$

Thus, the more equal are the *log* variances, and the more different are the *log* mean incomes, the closer is $\hat{W}_1$ to 1/2. Empirically, one would expect a much greater difference in *log* mean incomes than in *log* variances and so expect $\hat{W}_1$ to be in the range zero to one.

Some properties of these equations are interesting. The U hypothesis in no way depends on which sector has the higher income. If total income is to rise then Y_1 must be greater than Y_2, but the U hypothesis depends only on their being different. It also does not matter which sector has the more unequal distribution of within sector income. The difference between σ_1^2 and σ_2^2 affects $\hat{W}_1$, but not the existence of the U. If $\sigma_1^2 < \sigma_2^2$, then it will take longer for the distribution to start becoming more equal (for a given rate of change of sector population shares), but the turning point exists. It is interesting that even if people are moved from a sector with relatively more equality to one with less, the overall distribution will still become more equal.

A simple example will given an idea of the orders of magnitude involved. Assume the following parameter values:[3]

$$Y_1 = 6.91 \qquad Y_2 = 4.61$$
$$\sigma_1^2 = .40 \qquad \sigma_2^2 = .20$$

[3] The within-group *log* variances are based roughly on data from studies of South Korea and Turkey. The group mean incomes corresponding to the *log* means are rough guesses and the effects of varying them are explored below.

TABLE 1—DISTRIBUTION STATISTICS: VARIOUS GROUP POPULATION SHARES

W_1	Y	Geometric Means	σ^2	Gini Coefficient
.10	4.84	126.	.70	.44
.20	5.07	159.	1.09	.54
.40	5.53	252.	1.55	.62
.52	5.80	331.	1.63	.63
.60	5.99	399.	1.59	.63
.80	6.45	633.	1.21	.56
.90	6.68	796.	.86	.49

Note: The geometric mean is exp (Y). The Gini coefficient is calculated from the *log* variance assuming a lognormal distribution. The overall *log* variance σ^2 is calculated from equation (4). $(\sigma_1^2-\sigma_2^2)=.20$.

The values of Y_1 and Y_2 correspond to geometric mean incomes of 1000 and 100, respectively. If one assumes a lognormal distribution, then the arithmetic mean income is a function of the *log* mean and *log* variances and the Gini coefficient is a simple function of *log* variance.[4] The corresponding arithmetic mean incomes are 1221 and 111 and the within-sector Gini coefficients are 0.35 and 0.25 for distributions one and two, respectively. Statistics for the overall distribution given different values of W_1 are presented in Table 1. The maximum value of σ^2 occurs when $\hat{W}_1 = .52$. The U relationship between inequality and development appears very distinctly in the example. It is interesting that even when $W_1 = .10$ or $W_1 = .90$, σ^2 is significantly different from σ_2^2 or σ_1^2.

During the course of economic development, one also might expect the mean income differences of various groups to narrow, perhaps after an initial rise. Table 2 indicates how inequality varies as the ratio of group mean incomes change. If one assumes that group population shares do not change

[4] For a lognormal distribution with *log* variance σ^2, the Gini coefficient is given by $G = 2[\int_{-\infty}^{x} N(0,1)] - 1.0$ where $x = \sigma/\sqrt{2}$ and $N(0, 1)$ is the standard normal distribution. The arithmetic mean is equal to exp $(Y + 0.5\,\sigma^2)$ where Y is the *log* mean. See J. Aitchison and J. A. C. Brown. Note that the use of this formula for the overall distribution assuming that the within-group distributions are lognormal is only an approximation since the sum of two lognormal distributions is not necessarily lognormal.

TABLE 2—DISTRIBUTION STATISTICS: VARIOUS GROUP INCOME RATIOS

Income Ratio	Y_1-Y_2	$\hat{W}_1$	σ^2	Gini Coefficient
10	2.30	.52	1.63	.63
9	2.20	.52	1.51	.62
8	2.08	.52	1.38	.59
7	1.95	.53	1.25	.57
6	1.79	.53	1.11	.54
5	1.61	.54	0.95	.51
4	1.39	.55	0.79	.47
3	1.10	.58	0.61	.42
2	0.69	.71	0.44	.36

Note: Income ratio is the ratio of geometric mean incomes, exp (Y_1)/exp (Y_2). The variance σ^2 is calculated from equation (4) for the corresponding turning point value of W_1. $(\sigma_1^2-\sigma_2^2)=.20$.

but only that the spread of group mean incomes first widens and then later narrows, one also gets a significant U relationship.

It is possible to compare the relative magnitudes of the effect on inequality of changes in group mean incomes and of changes in group population shares. Define $R=Y_1-Y_2$. The logarithm of the ratio of group mean incomes is R. If $Y_1>Y_2$, then $R>0$. Equation (4) can be rewritten as:

$$(6) \quad \sigma^2 = -R^2W_1^2 + [(\sigma_1^2-\sigma_2^2)+R^2]W_1 + \sigma_2^2$$

Take the total derivative of (6) with respect to changes in R and W_1:

$$(7) \quad d\sigma^2 = D\,dR + E\,dW_1$$

where
$$D = 2RW_1W_2$$
$$E = R^2(1-2W_1)+(\sigma_1^2-\sigma_2^2)$$

At the turning point $\hat{W}_1$ in equation (5), $E=0$ and only changes in relative mean incomes affect inequality. If $W_1=1/2$, $E=(\sigma_1^2-\sigma_2^2)$ and does not depend on R. For a given R, E is relatively small when W_1 is in the neighborhood of 1/2. Also, given R, D is largest when $W_1=W_2=1/2$.

Table 3 gives some values of D and E for various values of W_1 and R. A change in R of 0.1 corresponds roughly to a decrease of 1.0 in the ratio of mean incomes (say, from 10 to 9). A change in W_1 of 0.01 (one percentage point) seems roughly comparable to a change in R of 0.1. "Comparable" means that a country undergoing rapid structural change might achieve such a change in a year or two, although clearly a change in the income ratio from 10 to 9 is far easier than a change from 3 to 2.[5]

TABLE 3—COEFFICIENTS OF THE TOTAL DERIVATIVE $d\sigma^2=DdR+EdW_1$

W_1	R	$0.1D$	$0.01E$
.10	2.30	.041	.044
.10	1.61	.029	.023
.10	0.69	.012	.006
.20	2.30	.074	.034
.20	1.61	.052	.018
.20	0.69	.022	.005
.40	2.30	.111	.013
.40	1.61	.077	.007
.40	0.69	.033	.003

Note: R corresponds to income ratios of 10 ($R=2.30$), 5 ($R=1.61$), and 2 ($R=.69$). $(\sigma_1^2-\sigma_2^2)=.20$.

It is clear from both Tables 1 and 3 that given R, the U relation is very flat around $W_1=1/2$. Changes in W_1 from 0.40 to 0.60 have almost no effect on inequality. At the ends where $W_1=0.2$ or 0.1, the effect of changes in W_1 are comparable to changes in R. In the development process as it has actually occurred, R has typically first increased and then later decreased. The two effects have thus probably reinforced one another in generating the U relationship. If R stays relatively constant during the "middle" period of development, one would find that a country might spend a long time in the trough of the U.

The implication of this exercise is that if the two-sector models that many development economists use are valid, then one should expect that a developing country, in the absence of explicit countervailing policies, will have increasing or unchanged income inequality for a relatively long period. The result is also consistent with the em-

[5] For evidence on the rate of structural change in less developed countries, see Kuznets (1966, 1971) and the author.

pirical findings of Adelman and Morris. Of course, the model is very simple and ignores any systematic changes in the within-group distributions over time. It does, however, trace out the effects of simple structural shifts on income inequality, effects which are significant and should not be neglected in more complicated and more theoretically satisfying models.

REFERENCES

I. Adelman and C. T. Morris, *Economic Growth and Social Equity in Developing Countries,* Stanford 1973.

J. Aitchison and J. A. C. Brown, *The Lognormal Distribution,* Cambridge 1957.

J. C. H. Fei and G. Ranis, *Development of the Labor Surplus Economy: Theory and Policy,* Homewood 1964.

J. R. Harris and M. Todaro, "Migration, Unemployment, and Development: A Two Sector Analysis," *Amer. Econ. Rev.,* Mar. 1970, *60,* 126–43.

S. Kuznets, *Modern Economic Growth: Rate, Structure and Spread,* New Haven 1966.

———, *Economic Growth of Nations,* Cambridge, Mass. 1971.

———, "Economic Growth and Income Inequality," *Amer. Econ. Rev.,* Mar. 1955, *45,* 1–28.

W. A. Lewis, "Economic Development with Unlimited Supplies of Labour," *Manchester Sch. Econ. Soc. Stud.,* May 1954, *22,* 139–91.

F. Paukert, "Income Distribution at Different Levels of Development," *Int. Lab. Rev.,* Aug.-Sept. 1973, *108,* 97–125.

S. Robinson, "Sources of Growth in Less Developed Countries: A Cross-Section Study," *Quart. J. Econ.,* Aug. 1971, *85,* 391–408.