Dennis W. Carlton and Robert E. Hall*
The Distribution of Permanent Income

Most economists agree that the distribution of well-being is better measured by the distribution of permanent income than by the distribution of annual measured income, but there have been few attempts to put the idea into practice. Further, those attempts have made questionable assumptions about the relation between permanent income and observed variables, either by following Milton Friedman's suggestion that permanent income could be approximated by a moving average of actual income or by defining permanent income as a function of permanent observed characteristics, notably education. In this paper we take a rather different approach in which permanent income is an unobserved variable that cannot be measured at the level of the individual. Stochastic assumptions about transitory income make it possible to identify the distribution of permanent income within a population on the basis of observations on actual income in two years for each member of the population. Even with these assumptions, permanent income at the individual level remains unidentified.

The questions of greatest interest about the distribution of income relate to its tails. In the United States, the federal government has established a poverty threshold and reports the fraction of the population below the threshold each year. It is widely accepted that the fraction is biased upward by the inclusion of some individuals who are not genuinely poor but have suffered a purely temporary reduction in income from a normally high level. The bias is offset only partially by the exclusion of some of the genuinely poor on account of temporary increases in income. The central question addressed by our work is the magnitude of this bias. It is conceivable that there are almost no genuinely poor individuals and that the reported fraction in poverty is attributable entirely to temporary poverty. Our empirical results show that this is not the case. Though the lower tail of the distribution of permanent income is smaller than the lower tail of the distribution of reported income, it is by no means empty. Similarly, the observation that some fraction of the population has high incomes in a given year does not establish conclusively that there are any genuinely rich individuals, but our results rule this case out as well.

The statistical model of income underlying our work is precisely as stated in Chapter III of Friedman's *Theory of the Consumption Function*, except that we

deal throughout with logarithms of the variables. Actual log income, y, is treated as the sum of a permanent component, u, and a transitory component, v:

$$y = u + v$$

Our stochastic hypotheses are

1. The expected value of the transitory component, v, is zero. As Friedman points out, this assumption could easily fail within a population for a particular year, because there are transitory influences in the economy at large. Our technique requires only that the transitory component be drawn from a probability distribution with the same mean for all individuals, but in the empirical results presented in this paper we assume that the common mean is zero.

2. Successive transitory components are independent. This assumption is somewhat plausible when the unit of observation is the year, but not if it is the quarter or month. It is a testable assumption within the model. Positive serial correlation of the transitory component will cause us to overstate the dispersion of permanent income slightly.

3. Permanent and transitory components are independent. Friedman assumes only that the two components are uncorrelated; as he says, this assumption has »...little substantive content and can almost be regarded as simply completing or translating the definitions of transitory and permanent components; the qualitative notion that the transitory component is intended to embody is of an accidental and transient addition to or subtraction from income, which is almost equivalent to saying an addition or subtraction that is not correlated with the rest of income.« (Friedman, 1957, pp. 26 − 27). Friedman goes on to point out that zero correlation does not imply independence, and particularly that individuals with large permanent components probably have transitory components that tend to be large in magnitude. Our use of logs takes account of this tendency, perhaps more than adequately. Our technique does depend fundamentally on the full assumption of independence.

It is a remarkable fact that the three assumptions just stated are sufficient to identify the full distributions of the permanent and transitory components, given the joint distribution of observed income in a pair of years. We emphasize that no parametric assumptions about the distributions are required to achieve identification. Distributional questions of central importance such as »What fraction of the population has permanent income below the poverty threshold?« can be answered precisely from the joint frequency distribution of current income in two years for a sufficiently large body of data on individuals. Data on lifetime income are not required.

Statistical theory

Our problem can be formalized in the following way: Let u, v_1, and v_2 be independent random variables, identified as permanent income, transitory income in the first year, and transitory income in the second year, respectively. We know the joint cumulative distribution $F(Y_1, Y_2)$ of observed income in the two years,

$$y_1 = u + v_1$$
$$y_2 = u + v_2$$

104

We seek the cumulative distribution functions $G(U)$, $H_1(V_1)$, and $H_2(V_2)$ of u, v_1, and v_2. We further specify that v_1 and v_2 have means zero. Then a fundamental theorem due to C.R. Rao (1973, p. 469) assures that G, H_1, and H_2 are uniquely defined by F^1. No further information or assumptions are required.

Rao's theorem establishes the feasibility of determining the distribution of permanent income, but does not lead immediately to a practical technique. Under the further assumption that the various distributions have moments of all orders[2], it is fairly straightforward to calculate the moments of the distribution of permanent income. First we define a_i as the ith moment of y_1:

$$\alpha_i = E(y_1)$$

and $ß_i$ as a certain »cross-moment«:

$$ß_i = E(y_2 y_1^i).$$

Then we let μ_i be the ith moment of permanent income and λ_i be the ith moment of transitory income[3]. These can be calculated from the recursions,

$$\mu_i = ß_{i-1} - \sum_{j=1}^{i-1} B_{j-1}^{i-1} \mu_j \lambda_{i-j}$$

$$\lambda_i = a_i - \sum_{j=1}^{i} B_j^i \mu_j \lambda_{i-j}$$

Here B_j^i is the jth binomial coefficient of order i. Initial conditions for the recursion are

$$\mu_0 = 1$$
$$\lambda_0 = 1$$

Appendix A establishes the validity of the recursion.

The moments of the distribution of permanent income are not themselves very informative, especially about the tails of the distribution. Our goal is to compute values of $G(U)$ itself for a variety of interesting values of U, including the poverty threshold. The only problem is to find a good way to approximate $G(U)$ on the basis of a finite number of moments. The difficulty is that the obvious methods do not converge very rapidly as the number of moments increases and do not have a rigorous sharp bound on the magnitude of the approximation error. We have adopted a somewhat more roundabout method that provides exact sharp bounds on the approximation error.

The first N moments of G are related to G by the definiton,

$$\mu_j = \int u^j dG(u) \qquad j = 0,...,N$$

This is the classical example of what is termed a Tchebycheff system. Much is known about the information regarding the unknown distribution G that can be extracted from the observed μ_j. The earliest result of this kind is the Tchebycheff inequality which gives a lower bound on $G(\mu_1 + \delta) - G(\mu_1 - \delta)$ in the case of $N = 2$, for arbitrary values of δ. For higher values of N, a powerful theory due to Markov and Krein gives upper and lower bounds on $G(U)$. Many other bounds have been developed for special cases. A unified presentation of results in this

area appears in Karlin and Studden (1966). However, the mathematical literature does not contain results on the more general problem of upper and lower bounds $\overline{P}$ and $\underline{P}$ on $P = G(U_2) - G(U_1)$ for arbitrary limits U_1 and U_2 and arbitrary N.

This problem can be set up as one in infinite dimensional linear programming:

$$\overline{P} = \sup \int b(u)dG(u)$$

subject to

$$\mu_i = \int u^i dG(u) \qquad i = 0,...,N.$$

Here $b(u)$ is an indicator function for the interval $[U_1, U_2]$ taking the values one within the interval and zero outside it. We also seek $\underline{P}$, the inf of the same objective function subject to the same constraints. In Appendix B, we show that a modification of a technique due to Dantzig (1963) can be used to solve the infinite dimensional problem through a sequence of solutions of problems of finite dimension.

For any modest number of moments, the bounds $\underline{P}$ and $\overline{P}$ are fairly widely separated — this is the basic source of the slow convergence of methods based on approximating P with a finite set of moments. Our rigorous bounds on the approximation error are unfortunately rather wide. However, the extreme distributions for which $\underline{P}$ and $\overline{P}$ are attained have rather peculiar characteristics. They assign positive probability mass to N+1 points and zero probability elsewhere. In other words, of the many distributions of permanent income consistent with the given set of N moments, the distributions that could cause the largest approximation errors are extremely lumpy — the entire population has only N+1 different levels of permanent income.

There are two ways to reduce the bounds on the approximation error. The first is simply to use enough moments to bring $\underline{P}$ and $\overline{P}$ sufficiently close together. The extreme distributions become less and less lumpy as N increases, and there is always a large enough N to achieve any prescribed level of accuracy in calculating P.

The second approach is to impose additional prior information about the shape of the distribution of permanent income and to calculate $\underline{P}$ and $\overline{P}$ subject to the constraints imposed by this prior information.

In the empirical results presented later in this paper, we made use of two kinds of prior beliefs about the distribution. First, we required that the distribution be unimodal — the probability density must rise below the median and fall above the median. Second, we required that the distribution be smooth in a certain rather weak sense — essentially the derivative of the density must not exceed a prescribed bound. Details about the two restrictions appear in Appendix C. Together they brought about a substantial tightening of the bounds computed on the basis of a given set of moments.

Empirical results

In this section we present the results of a study of the annual incomes of a sample of US families over the period 1967—72. The data were obtained from the Panel Study of Income Dynamics (1972). We restricted the sample to nonrural families with white male heads between the ages of 25 and 64 in order to make it more

homogeneous. We also eliminated families whose records of income were incomplete for the six years. Our sample contained 3 048 observations for 508 families. We defined income as total labor earnings of the head deflated by an index of income of men in the total US economy, with 1967 set to 1.000[4]. Since we had six observations for each family rather than just the two assumed in the earlier discussion in this paper, we computed the moments of the distribution of permanent income in a somewhat different way. The ith moment, μ_i, was computed as the average of all possible products of i distinct observations on actual income for each family. According to the permanent income model, the expectation of each such product is the moment, μ_i. It is possible to show that the six moments that can be computed this way are the only ones that can be derived from the data without adjustment for the moments of the distribution of transitory income[5]. We limited our analysis to the six moments because of our concern about the possibly large sampling variation in moments that were estimated by the recursive process outlined earlier. However, we have not so far carried out a formal analysis of the problem of sampling variation.

This procedure yielded the following moments:

$$\mu_1 = 9.0662; \quad \mu_2 = .1980; \quad \mu_3 = -.0067;$$
$$\mu_4 = .1926; \quad \mu_5 = -.0345; \quad \mu_6 = .3759.$$

The second and higher-order moments are centered around the mean.

The conventional components of variance analysis reveals that there is an important transitory component of income. The sample variance of the log of annual income is 0.2500, well in excess of our estimate of the variance of the log of permanent income $\mu_2 = .1980$. The difference between them, .0520, is our estimate of the variance of the log of transitory income.

Our next step is to investigate the information about the distribution of permanent income contained in the full set of six moments. Before starting a discussion based on the methods developed earlier in this paper, however, we should point out that one popular model of the income distribution, the log normal distribution, is clearly contradicted by our findings. The normal distribution with variance .1980 should have a fourth moment of .1176 and a sixth moment of .1164. Our estimates exceed these values by margins far larger than could be explained by sampling errors. The distribution of permanent income has tails that are thicker than predicted by the normal distribution.

Table 1 presents the results of applying our techniques to the problem of measuring the fraction of permanent income in each of a set of income categories. The first column shows the distribution of actual income within the sample. Presumably the tails of this distribution are fattened by the inclusion of transitory income. The next column gives the lower bound on the percent of the population having permanent incomes in the category, without imposing the restrictions that the distribution be unimodal and smooth. The third column gives the corresponding upper bound. These two columns can be interpreted as stating all of the information about the distribution across the income categories that is rigorously derivable from the first six moments of the distribution. For example, it is not possible to rule out definitively the possibility that there are no genuinely poor[6] families in this population and that all of those classified as poor on the basis of actual income have negative transitory components. Similarly, the first six

Table 1 Distribution of Observed Income and Derived Distribution of Permanent Income

Income range	Distribution of observed income (percent)	Unrestricted bounds		Restricted bounds (unimodal and smooth)	
		$\underline{P}$	$\overline{P}$	$\underline{P}$	$\overline{P}$
0–2000	1.1	0	2.1	0.3	0.6
0–3335	2.8	0	11.2	1.9	3.1
2000–8000	40.8	6.7	83.0	37.5	45.9
8000–14,000	44.6	0	86.6	39.0	51.0
14,000–20,000	9.1	0	35.3	6.8	12.0
20,000+	4.4	0	14.8	2.7	4.0

moments do not show conclusively that there is any family with a permanent income above $ 20,000. This illustrates just how weak is the distributional information contained in a small set of moments. The large gap between $\underline{P}$ and $\overline{P}$ in every category shows that the problem is the low information content of moments with respect to the class of all possible distributions. The fourth and fifth columns restrict the class of distributions to those that are unimodal and smooth. Within this class, the first six moments are much more informative. $\underline{P}$ is positive in every category, so the existence of genuinely poor and genuinely well-off families is established conditional on the hypothesis that the distribution of permanent income is unimodal and smooth. $\overline{P}$ now yields useful information — no more than 0.6 percent of all families are in the very lowest income category, and no more than 4.0 percent in the highest. The latter conclusion is particularly interesting because it clearly supports the view that the tails of the distribution of actual income are too fat — 4.4 percent of the families have observed incomes in the top category, but at least 0.4 percentage points represent the net effect of misclassification caused by transitory income. On the other hand, the results are not strong enough to demonstrate that a similar overstatement occurs in the official poverty category.

In addition to the values of the bounds in Table 1, our technique yields actual distributions that have the ovserved moments and assign fractions of the population equal to the bounds in the prescribed income interval. Without imposing prior constraints on the shape of the distribution of permanent income, our algorithm gave an upper bound of 11.2 percent of the population with permanent income below the poverty line and a lower bound of zero. The distribution corresponding to the upper bound has 11.2 percent of families with permanent incomes of $ 3315, 36.6 percent with $ 8643, 42.5 percent with $ 8909, 6.5 percent with $ 23,069, and 3.2 percent with $ 23,489. In this world, an important fraction of families have levels of well being just inside the poverty line. The great bulk of the population is around the mean and a minority of around 10 percent have very high incomes. The distribution for the lower bound is very similar except that the low-income group has income just above the poverty line.

The accomplishments of this paper are, first, to draw attention to Rao's theorem and the theoretical possibility of identifying fully the distribution of permanent income from limited data, and second, to generalize the existing theory of variance components to higher moments than the second. The growing availability of longitudinal data on individual families should make it possible to refine the information available on the distribution of permanent income using the techniques suggested in this paper. Our results to date are only at the borderline of usefulness. The first step in improving them should be the development of a formal treatment of sampling variation. Then it should be possible to decide whether the recursive calculation of higher-order moments adds additional useful information. It may be possible to compute the desired bounds virtually exactly by using enough moments, in which case the rather tedious and expensive computation of bounds with linear programming could be eliminated. Another approach that we have investigated tentatively is to drop the assumption of the existence of moments and to compute P directly from the joint frequency distribution of successive observations on income. Nothing practical has emerged from our work on this approach to date, however.

Appendix A. Recursion for the moments of permanent and transitory income

First,

$$\alpha_i = E(y_1^i)$$

$$= E(\sum_{j=0}^{i} B_j^i u^j v_1^{i-j})$$

$$= \sum_{j=0}^{i} B_j^i \mu_j \lambda_{i-j} \, ,$$

where $\mu_j = Eu^j$, and $\lambda_{i-j} = Ev^{i-j}$.

Since $B_0 = 1$ and $\mu_0 = 1$,

$$\lambda_i = \alpha_i - \sum_{j=1}^{i} B_j^i \mu_j \, \lambda_{i-j}$$

Second,

$$\beta_i = E(y_2 y_1^i)$$

$$= E((u+v_2)y_1^i)$$

$$= E(uy_1^i)$$

$$= E(\sum_{j=0}^{i} B_j^i u^{j+1} v_1^{i-j})$$

$$= \sum_{j=0}^{i} B_j^i \mu_{j+1} \lambda_{i-j} \, .$$

109

Since $B_i^i = 1$ and $\lambda_0 = 1$,

$$\mu_{i+1} = \beta_i - \sum_{j=0}^{i-1} B_j^i \mu_{j+1} \lambda_{i-j}$$

Appendix B. Calculating $\underline{P}$ and $\overline{P}$

To solve the infinite dimensional LP problem we solve a sequence of finite dimensional LPs. Since the identical techniques apply to either the upper or lower bound problem, we carry out the discussion for the upper bound only.

Let $b(u)$ be an indicator function that takes on the value 1 when u is in the interval of interest, and zero otherwise. To find the upper bound $\overline{P}$, for any interval, we solve

$$\overline{P}_N = \max_{\{g_n\}} \sum_{n=1}^N b(u_n) g_n \tag{B1}$$

subject to

$$\sum_{n=1}^N u_n^j g_n = \mu_j, \quad j = 1,\dots,T, \text{ and} \tag{B2}$$

$$\sum_{n=1}^N g_n = 1,$$

$$g_n \geq 0 \text{ for } n = 1\dots N.$$

Here $\{u_n\}$ is a gird of points in the range of u and μ_j are the $j = 1\dots 6$ moments of permanent income. Next we define

$$\Delta^N = \max_u \left[b(u) - \sum_{j=0}^T \pi_j u^j \right] \tag{B3}$$

where $\pi_0,\dots,\pi_T$ are the dual variables from the LP problem. Further, let u^{*N} be a value of u where the max in (B3) is attained. With this preparation we define the

Iterative algorithm for calculating $\overline{P}$:

Start with a value of N and a grid such that the LP problem, (B1)–(B2), has a feasible solution. We assume that there exists some probability mass distribution which has the same 6 moments as our underlying continuous distribution. This was the case for all our examples. Define a new grid by appending u^{*N} to the old grid. Iterate until Δ^N is suitably small. The properties of the algorithm are established in the following

Theorem 2. The sequence of $\overline{P}^N$ converges monotonically upward to $\overline{P}$:

$$\overline{P}^N < \overline{P}^{N+1} < \dots < \overline{P} \text{ and } \lim_{N \to \infty} \overline{P}^N = \overline{P}. \text{ Further, the error at any}$$

iteration is bounded by Δ^N: $\overline{P}^N < \overline{P} < \overline{P}^N + \Delta^N$.

110

The proof parallels Dantzig (1963, ch. 24). Our practical experience with the algorithm has been entirely favorable — with T = 6, we have achieved convergence in fewer than 4 iterations in most cases, where the convergence criterion is $|\Delta^{n*}|$ < .0001.

Appendix C. Restrictions on the shape of the distribution of permanent income

We imposed two restrictions on the distribution of permanent income. First, we required the distribution to be unimodal at the median, increasing before the median, and decreasing beyond the median.

The second restriction imposed smoothness by limiting the slope of the density function. We regard these assumptions about the distribution of permanent income as plausible and very weak. They are designed to rule out the unlikely lumpy distributions that correspond to the bounds reported in the fourth and fifth column of Table 1.

The first constraints are of the form $g_{n+1} - g_n > 0$ when n represents a point to the left of the median income and $g_{n+1} - g_n < 0$ when n represents a point to the right of the median income. The growth constraints are of the form $\dfrac{|g_{n+1} - g_n|}{|g_n|} \leq 1.3$. Essentially, we are imposing a finite rate at which the probability mass can increase or decrease between adjacent intervals of the log of income.

To obtain an idea of what 1.3 means as a growth rate, consider the normal distribution. For a comparable grid size, the growth rate of the normal would be at most 1.1–1.15. Therefore, our growth rate of 1.3 allows considerable departure from normality in the set of feasible distributions for the log of permanent income.

Once the constraints are imposed, the computation of $\bar{P}$ and $\underline{P}$ becomes more complicated than before. The techniques of Appendix B cannot be used, because whenever an additional point is added to the grid the constraints also change. To calculate $\bar{P}$, we choose a very fine grid size for the log of income. Each interval was .06 units on the log of income scale. For this grid size, N = 200. Using this grid size, we solved the following LP:

$$\bar{P} = \max_{\{g_n\}} \sum_{n=1}^{N} b_n g_n$$

s.t.

$$\sum_{n=1}^{N} g_n u_n = \mu_i \qquad i = 1, \cdots 6,$$

$$\sum_{n=1}^{N} g_n = 1$$

$$g_{n+1} - g_n > 0 \qquad \text{if } u_n < \text{median income}$$

$$g_{n+1} - g_n < 0 \qquad \text{if } u_n > \text{median income}$$

$$\frac{|g_{n+1} - g_n|}{|g_n|} < 1.3$$

$$g_n > 0,$$

111

where all notation has been previously defined in Appendix B.

The solution to such large LP programs can be formidable. The IBM MPSX routine was used to solve this LP.

Except for the obvious modificaions, the solution for the lower bound $\underline{P}$ is identical to that just given for $\overline{P}$.

Notes

[1] Gary Chamberlain called our attention to Rao's theorem.

[2] This is not an entirely innocuous assumption, even with respect to the distribution of the log of income.

[3] For this discussion, we assume the transitory components v_1 and v_2 have the same distribution. Our later empirical results do not rest on this assumption.

[4] This index was constructed from Table 2, Current Population Reports P–60, No. 92, March 1974.

[5] Because such an adjustment is not needed, our results do not require any assumption that the transitory components are identically distributed.

[6] The official poverty line in 1967 for a family of four was $3335. Source: Current Population Reports, P–23, No. 28, August 1969.

References

Dantzig, G., *Linear Programming and Extensions*, Princeton, N.J., Princeton University Press, 1963

Friedman, M., *A Theory of the Consumption Function*, Princeton, N.J., Princeton University Press, 1957

Kagan, A., Y Linnik, C. Rao, *Characterization Problems in Mathematical Statistics*, New York, Wiley, 1973

Karlin, M. and J. Studden, *Tchebycheff Systems with Applications in Analysis and Statistics*, New York, Interscience Publishers, 1966

Survey Research Center – A Panel Study of Income Dynamics, Institute for Social Research, The University of Michigan, Ann Arbor, Michigan, 1972

Offprint from

Income Distribution
and Economic Inequality

Edited by Zvi Griliches, Wilhelm Krelle,
Hans-Jürgen Krupp, and Oldrich Kyn