# MODELING THE DYNAMICS OF THE NONMETRO DISTRIBUTION OF WAGE AND SALARY INCOME AS A FUNCTION OF ITS MEAN

John Angle
Economic Research Service, 1800 M Street, NW, Washington, DC 20036

**Key Words: distribution, dynamics, nonmetro, wage and salary income[1]**

The question of what would happen to the nonmetropolitan (abbreviated 'nonmetro')[2] distribution of wage and salary income should its mean fall substantially is important, because nonmetro workers have long been disadvantaged relative to metro workers. Historically, as a group they have lower levels of education and a much larger proportion employed in low wage jobs (Fuguitt, Brown, Beale, 1989; Duncan, 1992). The best data set to answer this question is the March, Current Population Survey (CPS), a survey of a large sample of households by the U.S. Bureau of the Census on behalf of the U.S. Bureau of Labor Statistics[3].

## A Model That Accounts for a Puzzling Pattern in the Dynamics of a Distribution

Forecasting an event of a magnitude greater than any seen in the database, requires a robust model of the dynamics of the phenomenon. The greatest year to year fall in mean nonmetro annual wage and salary income from 1963 through 1995 was about 5%. It would take a very good model to credibly forecast what would happen in the event of a 50% decrease in the mean year to year. Close fit and model parsimony are evidence for the validity of this paper's model but confidence in the model's ability to make an extraordinary forecast is more likely to be forthcoming if it can be shown that the model explains a puzzling pattern in the data. This paper's model accounts for a visually arresting but at first glance quite puzzling pattern in the dynamics of the nonmetro distribution of wage and salary income in the thirty-three year period, 1963 through 1995. This pattern is puzzling because while suggestive of major systemic constraints on income distribution of some kind, it is neither discussed nor explained in the economic literature.

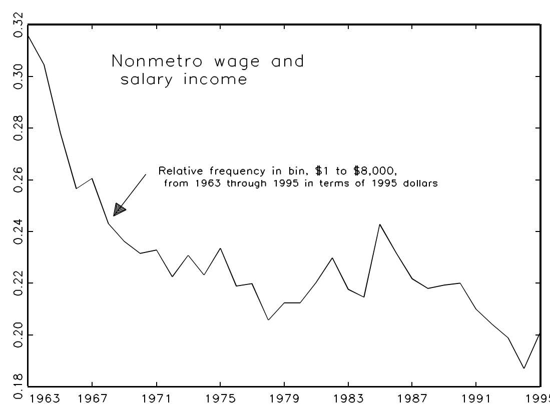


Figure 1

The puzzling pattern is the fact that when figure 1 is rotated $180^0$ around its x-axis and overlain on top of figure 2, as can be easily done when both graphs are on transparencies, the two graphs largely overlap. Figure 1 is the time-series of the relative frequency of nonmetro wage and salary incomes in the range (or bin as statisticians say) of $1 to $8,000 in terms of 1995 dollars. Figure 2 is the time-series of the relative frequency of the bin $36,001 to $44,000. This bin is to the right of the mode and median of the distribution. The two relative frequencies are correlated -.888. The relative frequency in the left tail bin is greater than the relative frequency in the right tail bin and the graphs of figures 1 and 2 could not overlap in the absence of the scaling done automatically by the graphing program. The following transformations, (1) applied to the left tail relative frequencies:

$$\frac{\max(y_l)-y_{lt}}{\max(y_l)-\min(y_l)} \tag{1}$$

and (2) applied to the right tail relative frequencies:

$$\frac{y_{rt}-\min(y_r)}{\max(y_r)-\min(y_r)} \tag{2}$$

---

[1] Thanks to Gordon Burt, Open University, and ERS/USDA colleagues Linda Ghelfi, Robert Gibbs, Charles Hallahan, Karen Hamrick, and Dean Jolliffe for their comments on previous drafts of this paper. They bear no responsibility for any error. This paper does not necessarily reflect the views and policies of the Economic Research Service or the U.S. Department of Agriculture.

[2] 'Nonmetro' refers to the set of nonmetropolitan counties. A nonmetropolitan county is a county not in a Metropolitan Statistical Area (MSA) as defined by the Office of Management and Budget. MSA's include core counties containing a city of 50,000 or more people or having an urbanized area of 50,000 or more and total area population of at least 100,000. Additional contiguous counties are included in the MSA if they are economically integrated with the core county or counties. The metropolitan status of every county in the U.S. is re-evaluated following the Decennial Census. While there has been a net decline in counties classified as nonmetro over the decades, the definition of nonmetro has remained roughly constant

[3] The distribution of nonmetro wage and salary income is estimated with data from the March Current Population Surveys (CPS), a national household survey conducted by U.S. Bureau of the Census, whose data from the years 1964 through 1996 have been cleaned and recoded by Unicon inc. (1997). The mean number of individual records in the March CPS in this thirty-three period is 142,796 individuals. The March CPS has a supplement of questions on income posed on behalf of the Bureau of Labor Statistics .This supplement inquires about annual wage and salary income in the previous calendar year. The present study includes anyone earning at least $1 in wage and salary income, between the ages of 25 to 65 residing in a nonmetro county. There is an average of 13,291 individuals in each year who meet the criteria for selection.

The measurement of education changed in the CPS after the 1990 Census from a count of years of school completed to a more degree oriented measure which better measures the diversity of post-secondary education. The present study reconciles the two categorizations of educational attainment by collapsing both sets of categories to an ordinal polytomy of five categories. The crudeness of this polytomy obliterates the distinction between the two different categorizations of educational attainment. This polytomy is 1) elementary school or less, 2) some high school, 3) completed four years of high school, 4) some college, and 5) completed four or more years of post-secondary education.

All dollar amounts have been converted to constant 1995 dollars using the PCE (personal consumption expenditure) price index numbers form Table B-7 Chain-type price indexes for gross domestic product, **Economic Report to the President**, February 2002 (Council of Economic Advisers, 2002).

mimic the scaling done in figures 1 and 2 and, additionally, transformation (1) "flips" the left tail relative frequencies to make them nearly monotonic increasing instead of nearly monotonic decreasing. The result is the near overlap of the transformed relative frequencies in figure 3.

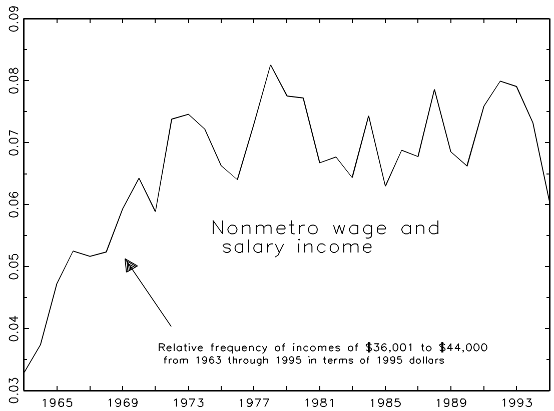


Figure 2

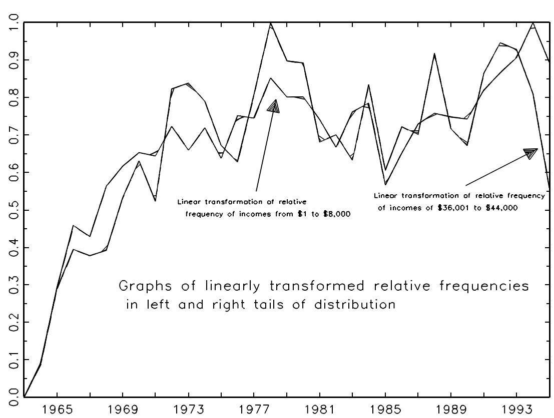


Figure 3

The forecast of this paper's model of how a greater fall in the unconditional mean of annual nonmetro wage and salary income than any seen in recent decades would change the distribution is credible because the model fits the data closely and explains the puzzling quasi-symmetry of its dynamics with great parsimony. The parsimony of the model inheres in its small number of parameters, the fact that all change in the model is from exogenous shock, mostly from one variable, change in the unconditional mean, and the fact that the model is not specialized in any way for the nonmetro population of the U.S.

This model, $f_{it}(x)$, is one of the distribution of annual wage and salary income of nonmetro workers at a particular level of education, level i, in year t. The unconditional nonmetro distribution is the sum of these partial[4] distributions by level of education, weighted by the number of workers at each level of education. ‘ $f_{it}(x)$’ names the mathematical model rather than the empirical partial distribution of which it is a model.

The model $f_{it}(x)$ can be expressed in three propositions:

1) Each partial[4] distribution is a gamma probability density function (pdf):

$$f_{it}(x) \equiv \frac{\lambda_t^{\alpha_i}}{\Gamma(\alpha_i)} x^{\alpha_i - 1} e^{-\lambda_t x} \tag{3}$$

2) Each gamma pdf modeling a partial distribution has an unchanging shape:

$$\begin{aligned} \alpha_i \ &\equiv \ \text{shape parameter of people at ith level of education} \\ &> 0 \end{aligned} \tag{4}$$

3) All $f_{it}(x)$, i = 1,...,5 have the same scale parameter:

$$\begin{aligned} \lambda_t \ &= \ \frac{\bar{\alpha}_t}{\bar{x}_t} \\ &> 0 \end{aligned} \tag{5}$$

where,

$$\begin{aligned} f_{it}(x) \ &\equiv \ \text{pdf model of the distribution at the ith level of education at time t} \\ x \ &\equiv \ \text{income} \ > \ 0 \\ \lambda_t \ &\equiv \ \text{scale parameter at time t} \\ \bar{x}_t \ &\equiv \ \text{unconditional mean of income at time t} \\ \bar{\alpha}_t \ &= \ w_{1t}\alpha_1 + \ldots + w_{it}\alpha_i + \ldots + w_{5t}\alpha_I \\ w_{it} \ &\equiv \ \text{proportion of population at ith level of education at time t} \end{aligned}$$

and $\bar{x}_t$ and the $w_{it}$’s are exogenous and the sole source of change. If they do not change, the model is static. The $w_{it}$’s vary more slowly proportionally than $\bar{x}_t$. While making allowance for variation in $w_{it}$, the model takes $\Delta\bar{x}_t$ where:

$$\Delta\bar{x}_t \equiv \bar{x}_t - \bar{x}_{(t-1)} \tag{6}$$

as the main source of change from year to year.

## The Dynamics of the Model, $f_{it}(x)$, as a function of $\Delta\bar{x}_t$

Seeing how this paper’s model of a partial distribution of the conditional distribution, nonmetro wage and salary income conditioned on education, $f_{it}(x)$, is driven by change in the unconditional mean of nonmetro wage and salary income, $\bar{x}_t$, requires taking the partial derivative:

$$\frac{\partial f_{it}(x)}{\partial \bar{x}_t} \tag{7}$$

Appendix A shows that:

[4] ‘Partial distribution’ is used in the following sense. Where f(x|y) is the distribution of income, x, conditioned on education level, y, and y is a discrete variable, $y_i$, i = 1,2,3,4,5, then $f(x|y_i)$ is the partial distribution of income conditioned on education level i.

$$\frac{\partial f_{it}(x)}{\partial \bar{x}_t} = f_{it}(x) \cdot \left( \frac{\bar{\alpha}_t \left(x - \bar{x}_{it}\right)}{\bar{x}_t^2} \right) = f_{it}(x)\, a_{it}$$

(8)

where:

$$a_{it} \equiv \left[ \frac{\bar{\alpha}_t \left(x - \bar{x}_{it}\right)}{\bar{x}_t^2} \right]$$

(9)

and $\bar{x}_{it}$ is the mean of $f_{it}(x)$, the conditional mean as distinguished from $\bar{x}_t$, the unconditional mean. As can readily be seen in (8), as $\bar{x}_t$ increases, the partial derivative (7) falls to the left of $\bar{x}_{it}$ and rises to the right. the bigger $|x - \bar{x}_{it}|$ is, the greater is the absolute value of the partial derivative (7), holding $f_{it}(x)$ constant. In other words, as the unconditional mean increases, the left tail (more precisely, all incomes $x_0 | x_0 < \bar{x}_{it}$) thins and the right tail (all $x_0 | x_0 > \bar{x}_{it}$) thickens, meaning that, in the model, there are fewer small incomes and more larger incomes. Holding $f_{it}(x)$ constant, this effect is greater in absolute value, the farther from the conditional mean, $\bar{x}_{it}$, a particular income, $x_0$, is. When the unconditional mean falls, the reverse happens: the left tail thickens and the right tail thins. While the partial derivative (7) becomes large in absolute value over the smallest incomes, $a_{it}$ acquires its largest absolute values in the extreme right tail, the largest income. Consequently the ratio $f_{i(t+1)}(x_0)/f_{it}(x_0)$ sees its largest values over $x_0$ where $x_0$ is the largest wage and salary income allowed by the economy.

## How the Model Reproduces The Puzzling Pattern

This paper's model, $f_{it}(x)$, [(3),(4),(5)], implies that for every point in the left tail of $f_{it}(x)$, i.e., for every $x_l$, $x_l < \bar{x}_{it}$, there is a corresponding point in the right tail of $f_{it}(x)$, an $x_r$, $x_r > \bar{x}_{it}$, such that after transformation (1) of $f_{it}(x_l)$ and transformation (2) of $f_{it}(x_r)$, the two transformed variables approximate each other. This result follows from the definition of the model, from Newton's approximation:

$$\begin{aligned} f_{i(t+1)}(x_0) &\approx f_{it}(x_0) + \left[ \frac{\partial f_{it}(x_0)}{\partial \bar{x}_t} \right] \Delta \bar{x}_{(t+1)} \\ &\approx f_{it}(x_0) + f_{it}(x_0)\, a_{i0t}\, \Delta \bar{x}_{(t+1)} \\ &\approx f_{it}(x_0) \left(1 + a_{i0t}\, \Delta \bar{x}_{(t+1)}\right) \end{aligned} \tag{10}$$

and from expressing the time-series of the relative frequencies in the left tail bin, $y_{il0}$, $y_{il1}$, $y_{il2}$, ...., $y_{ilT}$, [subscript order, a) education level, b)left or right tail, c) time] and the relative frequencies in the right tail bin, $y_{ir0}$, $y_{ir1}$, ...., $y_{irT}$, in terms of their initial terms, $y_{il0}$ or $y_{ir0}$ respectively. Several assumptions approximately met by the data are required too. For the relative frequencies in the left tail bin, $y_{ilt}$, their expression in terms of $y_{il0}$ is:

$$\begin{aligned} y_{il1} &\approx y_{il0} + y_{il0} a_{il0} \Delta \bar{x}_1 = y_{il0}\left(1 + a_{il0} \Delta \bar{x}_1\right) \\ y_{il2} &\approx y_{il0}\left(1 + a_{il0} \Delta \bar{x}_1\right) + y_{il0}\left(1 + a_{il0} \Delta \bar{x}_1\right) a_{il1} \Delta \bar{x}_2 \\ &\approx y_{il0}\left(1 + a_{il0} \Delta \bar{x}_1\right)\left(1 + a_{il1} \Delta \bar{x}_2\right) \\ &\ldots\ldots\ldots\ldots \\ y_{ilt} &\approx y_{il0}\left(1 + a_{il0} \Delta \bar{x}_1\right)\left(1 + a_{il1} \Delta \bar{x}_2\right)\ldots\ldots\left(1 + a_{il(t-1)} \Delta \bar{x}_t\right) \end{aligned} \tag{11}$$

and similarly for the right tail bin, $y_{irt}$:

$$\begin{aligned} y_{ir1} &\approx y_{ir0} + y_{ir0} a_{ir0} \Delta \bar{x}_1 = y_{ir0}\left(1 + a_{ir0} \Delta \bar{x}_1\right) \\ y_{ir2} &\approx y_{ir0}\left(1 + a_{ir0} \Delta \bar{x}_1\right) + y_{ir0}\left(1 + a_{ir0} \Delta \bar{x}_1\right) a_{ir1} \Delta \bar{x}_2 \\ &\approx y_{ir0}\left(1 + a_{ir0} \Delta \bar{x}_1\right)\left(1 + a_{ir1} \Delta \bar{x}_2\right) \\ &\ldots\ldots\ldots\ldots \\ y_{irt} &\approx y_{ir0}\left(1 + a_{ir0} \Delta \bar{x}_1\right)\left(1 + a_{ir1} \Delta \bar{x}_2\right)\ldots\ldots\left(1 + a_{ir(t-1)} \Delta \bar{x}_t\right) \end{aligned} \tag{12}$$

Since the maximum of $y_{ilt}$ is $y_{il0}$ and its minimum is either $y_{ilT}$ or close to it, the approximation of $y_{ilt}$ can be substituted into transformation (1) as:

$$\frac{y_{il0} - y_{il0}\left(1 + a_{il0} \Delta \bar{x}_1\right)\left(1 + a_{il1} \Delta \bar{x}_2\right)\ldots\ldots\left(1 + a_{il(t-1)} \Delta \bar{x}_t\right)}{y_{il0} - y_{il0}\left(1 + a_{il0} \Delta \bar{x}_1\right)\left(1 + a_{il1} \Delta \bar{x}_2\right)\ldots\ldots\left(1 + a_{il(T-1)} \Delta \bar{x}_T\right)} \tag{13}$$

Since the minimum of $y_{irt}$ is $y_{ir0}$ and its maximum is either $y_{irT}$ or close to it, the approximation of $y_{irt}$ can be substituted into transformation (2) as:

$$\frac{y_{ir0}\left(1 + a_{ir0} \Delta \bar{x}_1\right)\left(1 + a_{ir1} \Delta \bar{x}_2\right)\ldots\ldots\left(1 + a_{ir(t-1)} \Delta \bar{x}_t\right) - y_{ir0}}{y_{ir0}\left(1 + a_{ir0} \Delta \bar{x}_1\right)\left(1 + a_{ir1} \Delta \bar{x}_2\right)\ldots\ldots\left(1 + a_{ir(T-1)} \Delta \bar{x}_T\right) - y_{ir0}} \tag{14}$$

The $y_{il0}$'s cancel out of (13) while the $y_{ir0}$'s cancel out of (14). Where $|x_{ilt} - x_{it}| = |x_{irt} - x_{it}|$, $a_{ilt} = -a_{irt}$. Given this condition, $-a_{irt}$ can be substituted for $a_{ilt}$ in (13), yielding:

$$\frac{1 - \left(1 - a_{ir0} \Delta \bar{x}_1\right)\left(1 - a_{ir1} \Delta \bar{x}_2\right)\ldots\ldots\left(1 - a_{ir(t-1)} \Delta \bar{x}_t\right)}{1 - \left(1 - a_{ir0} \Delta \bar{x}_1\right)\left(1 - a_{ir1} \Delta \bar{x}_2\right)\ldots\ldots\left(1 - a_{ir(T-1)} \Delta \bar{x}_T\right)} \tag{15}$$

(15) does not equal (14), but (14) and (15) differ only in the sign of products with even numbers of $a_{irt}$ terms in both their numerators and denominators. If $0 < a_{irt} << 1.0$ , products with more than one $a_{irt}$ term are negligible. By assuming $0 < a_{irt} << 1.0$ and approximating the higher order terms in $a_{irt}$ by 0.0, (15) becomes (16):

$$\frac{a_{ir0} \Delta \bar{x}_1 + a_{ir1} \Delta \bar{x}_2 + \ldots + a_{ir(t-1)} \Delta \bar{x}_t}{a_{ir0} \Delta \bar{x}_1 + a_{ir1} \Delta \bar{x}_2 + \ldots + \ldots + a_{ir(T-1)} \Delta \bar{x}_T} \tag{16}$$

which equals (14), after the cancelling out of $y_{ir0}$, and the assumption that products with more than one $a_{irt}$ are zero. (16) can be further simplified. Where $|x_{irt} - \bar{x}_{it}| = |x_{ilt} - \bar{x}_{it}|$, and $x_{ilt}$ =

\$4,000, $x_{irt} = \overline{x}_{it} + (\overline{x}_{it} - \$4{,}000)$, and:

$$\bar{x}_{it} = \alpha_i \left( \frac{\bar{x}_t}{\bar{\alpha}_t} \right)$$

(from Appendix A), it follows from (9) that:

$$a_{irt} = \frac{\left( \alpha_i - \dfrac{\$4{,}000\, \bar{\alpha}_t}{\bar{x}_t} \right)}{\bar{x}_t} \tag{17}$$

If the numerator on the RHS of (17) is nearly constant, it can be factored out of the numerator and denominator of (16) leaving:

$$\frac{\displaystyle\sum_{\tau=0}^{\tau=(t-1)} \frac{\Delta \bar{x}_{(\tau+1)}}{\bar{x}_\tau}}{\displaystyle\sum_{\tau=0}^{\tau=(T-1)} \frac{\Delta \bar{x}_{(\tau+1)}}{\bar{x}_\tau}} \tag{18}$$

which is the same expression for all education groups.

(18) and the approximate equivalence of $y_{ilt}$ after transformation (1) and $y_{irt}$ after transformation (2) follow from the empirical asumptions that: a) $y_{ilt}$ is monotonic decreasing and $y_{irt}$ is monotonic increasing, b) the model, $f_{it}(x)$, defined by (3),(4), and (5), fits the data, c) $0 < a_{irt} << 1.0$, $0 < |a_{ilt}| << 1.0$, and d) the numerator of the RHS of (17) is approximately constant over time.

The approximate equivalence of the left tail relative frequency $y_{lt}$ of the unconditional distribution in figure 1 after transformation (1) and the unconditional right tail relative frequency $y_{rt}$ of figure 2 after transformation (2) requires additional empirical assumptions. $y_{lt}$ in terms of the partial distribution of the conditional distributions, $f_{it}(x)$ in the left tail at $x_l$ is:

$$y_{lt} \equiv f_t(x_l) = \sum_{i=1}^{i=5} w_{it} f_{it}(x_l)$$

and $y_{rt}$ is in the right tail at $x_r$:

$$y_{rt} \equiv f_t(x_r) = \sum_{i=1}^{i=5} w_{it} f_{it}(x_r)$$

The unconditional distribution, $f_t(x)$, is a gamma mixture distribution and not itself gamma. If e) the distribution of the $w_{it}$'s by level of education i is symmetric and f) $\alpha_i$ is monotonic increasing or decreasing with level of education i, it follows where:

$$x_{rt} = \sum_i w_{it} x_{irt}$$

and where:

$$\sum_i w_{it} |x_{irt} - \bar{x}_{it}| \approx \sum_i w_{it} |x_{ilt} - \bar{x}_{it}|$$

$$\text{for} \quad x_{ilt} = \$4{,}000$$

$$x_{irt} = \bar{x}_{it} + \left( \bar{x}_{it} - \$4{,}000 \right)$$

that $y_{lt}$ after transformation (1) approximates $y_{rt}$ after transformation (2). Demonstrating that these conditions are met by the data requires fitting the model and obtaining the parameter estimates of the fit.

**Estimation of Parameters of Model, $\{f_{it}(x)$, [(3),(4),(5)]$\}$, and Fit of Model to Data**

Validating this paper's model, $f_{it}(x_0)$, requires that it fits the data almost as well as a comparable, unconstrained parametric functional form and accounts for the puzzling pattern.. The data fitted are partial distributions[4] of the conditional distribution of nonmetro annual wage and salary income conditioned on education in the years, estimated from March Current Population Surveys. There are 165 such partial distributions: 165 = 33 years x 5 levels of education. The estimate of each partial distribution consists of 18 estimated relative frequencies, one in each of the 18 bins, each \$4,000 wide, from \$1-\$4,000 to \$68,001-\$72,000 in constant 1995 dollars. There are 165 x 18 = 2,970 relative frequencies estimated altogether. Each of these 2,970 estimates have been bootstrapped 50 times. The model requires the estimation of the unconditional mean in each year but only the unconditional mean of incomes from \$1 to \$72,000 in each year is known. Incomes greater than \$72,000 have been "topcoded", that is, altered by the U.S. Bureau of the Census to safeguard the identity of the respondent. The unconditional mean of nonmetro annual wage and salary income is estimated in a year as the product of a constant, k, by that year's unconditional mean of incomes from \$1 to \$72,000. There are six parameters in the vector of parameters searched over to minimize the sum of squared differences between the 2,970 observed relative frequencies and those expected under the model. The algorithm used to search the parameter vector to minimize squared error is a stochastic search algorithm known for not being trapped by local minima, simulated annealing ( Kirkpatrick, Gelatt, and Vecchi, 1983).

The estimated five shape parameters with standard errors estimated from 50 bootstrap samples in parentheses are: 1.126 (.0099) for an elementary school education, 1.263 (.01011) for some high school, 1.494 (.01098) for completion of four years of high school, 1.631 (.01753) for some college, and 2.331 (.02385) for completion of at least four years of college or university. The estimate of k is 1.1974 (.00265).

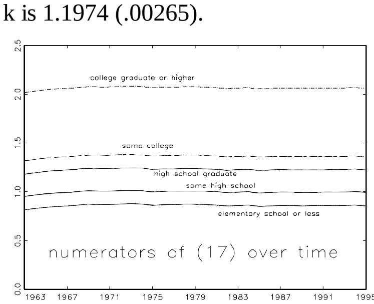


Figure 4

The fit is measured by the sum of squared errors between the 2,970 observed relative frequency distributions and the expected relative frequency distributions under the model, $f_{it}(x)$, [(3),(4), and (5)]. The sum of squared errors between the two sets of 2,970 numbers is .991. Another way of evaluating the fit of the model, one that emphasizes larger departures less than does squared error, is the sum and mean of absolute differences between the observed and expected relative frequencies. The sum of these is 39.98 with a mean absolute difference of .0135. The correlation coefficient between the observed and expected relative frequencies is .933. This six parameter fit of a highly constrained model to 2,970 relative frequencies is close.

The perception of closeness of fit in absolute terms is subjective. An objective standard of closeness of fit of a model is the cost of parsimony in terms of excess error over that of the comparable unconstrained model. The comparable unconstrained model is an unconstrained two parameter gamma pdf fit to each of the 165 partial distributions.[4] Where this paper's model, $f_{it}(x)$, has six parameters to be estimated, the comparable unconstrained parametric model has 330 parameters. This alternative is also fitted by simulated annealing a partial distribution[4] at a time. Each search is over a parameter vector with two cells, one for the gamma pdf's shape parameter and the other for its scale parameter.

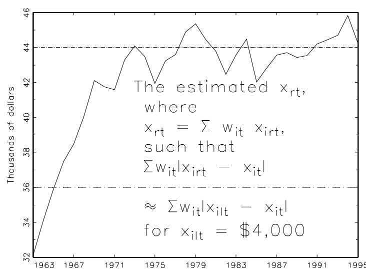


Figure 5

The sum of squared errors between the observed relative frequencies and those expected under this flexible parametric alternative is .845 where this paper's model has a sum of squared errors of .991, i.e., 17.2% more squared error than the 330 parameter unconstrained alternative model. The sum of absolute deviations of the 2,970 relative frequencies under the comparable unconstrained model is 37.66 whereas this paper's model, $f_{it}(x)$, has a sum of absolute deviations of 39.98. So by choosing this paper's model over the comparable unconstrained model, only 6.1% more absolute error is incurred. This paper's model, $f_{it}(x)$, [(3),(4) and (5)], fits almost as well as the comparable unconstrained alternative with 98.2% fewer parameters.

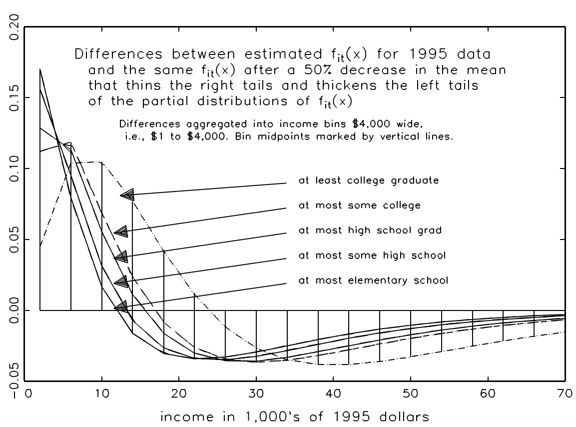


Figure 6

There are six empirical conditions that have to be met to explain the puzzling quasi-symmetry between figures 1 and 2. Condition a) that $y_{lt}$ is monotonic decreasing and $y_{rt}$ is monotonic increasing is approximately demonstrated by figures 1 and 2 themselves. Condition b) is that the model fits. It does. Condition c) is that $a_{irt}$ and $|a_{ilt}|$ are close to zero. The unweighted mean of $|a_{ilt}|$ over all times and education levels is .056; .048 is the comparable figure for $a_{irt}$. Condition c) is satisfied. Condition d) is demonstrated by figure 4. Condition e) holds because high school graduate, the middle category of education, is the modal category from 1964 through 1996, with the two adjacent categories greater than the extremal categories. Condition f) holds because it has been shown that $\alpha_i$ scales with level of education. Figure 5 shows the result of the approximate satisfaction of these conditions. For an $x_{lt}$ of $4,000, the estimated $x_{rt}$ which allows $y_{rt}$ after transformation (2) to approximate $y_{lt}$ after transformation (1) is in the vicinity of of $40,000, the mid-point of the bin $36,001-$44,000 from 1963 through 1995,
explaining the overlapping of the graphs in figure 3.

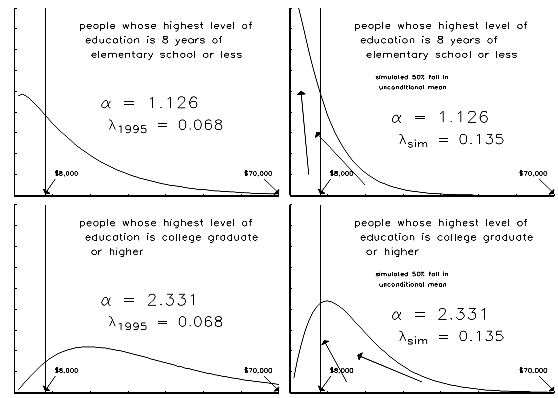


Figure 7

## Conclusions

A large percentage fall in the unconditional mean of nonmetro annual wage and salary income causes, in this paper's model, a whiplash increase in relative frequencies in the left tail of the distributions of all education levels in the nonmetro labor force, but especially among the least educated nonmetro workers. 18.5% of nonmetro workers with only an elementary school education (eight years of education or less) earned $4,000 or less in wage and salary income in 1995. $4,000 or less is clearly too little money to sustain that worker during a year, let alone a worker's family. This level of wage and salary income clearly requires assistance, whether from family, charity or the state in the form of welfare benefits, for example, food stamps. The comparable percentages for more educated nonmetro workers are: 14.1% of those with some high school education, 10.6% for high school graduates, 9.4% for those with some college, and 6.4% for nonmetro workers with four years of college or more.

The simulated scenario in this paper's model of a 50% decrease in the unconditional mean of nonmetro annual wage and salary income greatly increases the percentage in the low income bracket of $1 to $4,000 for all education groups. However, the least well educated are the most severely impacted. The percentage of nonmetro workers with an elementary school education earning $1-$4,000 in annual wage and salary income goes from 18.5% to 35.5% under this scenario. The comparable forecasted relative frequencies for more educated nonmetro workers are: 14.1% to 29.8% for those with some high school, 10.6% to 23.5% for high school graduates, 9.4% to 20.6% for those with some college, and only 6.4% to 10.9% for those with four years of college or more. In this paper's model the least well educated are the most at risk of ending up with a wage and salary income too small to sustain them in the scenario of a 50% drop in the unconditional mean of nonmetro wage and salary income. Figure 15 displays the forecasted differences in relative frequency in bins $4,000 wide by level of education in the scenario of a 50% decrease in the unconditional mean of nonmetro wage and salary income. Clearly, education provides protection against the hardest and farthest falls in wage and salary income in this scenario. Because the nonmetro labor force is less well educated than the metro labor force, it is, under this paper's model, more vulnerable to being thrust into the lowest annual wage and salary income brackets, for example, at or below $8,000, the income marked on figure 16, for a given percentage lowering of its unconditional mean of annual wage and salary income.

## References


Council of Economic Advisers. 2002. **Economic Report to the President**. Washington, DC: U.S. Government Printing Office.

Duncan, Cynthia. 1992. **Rural Poverty in America**. New York: Auburn House.

Fuguitt, Glenn, David Brown and Calvin Beale. 1989. **Rural and Small Town America**. New York: Russell Sage.

Kirkpatrick, S., C.D. Gelatt,and M.P. Vecchi. 1983. "Optimization by simulated annealing." **Science** 220: 671-680.

Unicon, inc. 1997. **Manual for March Current Population Surveys.** Santa Monica, CA: Unicon.


## Appendix A: The Partial Derivative of $f_{it}(x)$ With Respect to $\bar{x}_t$

The partial derivative of $f_{it}(x)$ with respect to $\bar{x}_t$ gives an expression for how $f_{it}(x)$ changes as a function of $\bar{x}_t$.

$$
\begin{aligned}
f_{it}(x) &= \frac{\lambda_t^{\alpha_i}}{\Gamma(\alpha_i)} x^{\alpha_i - 1} e^{-\lambda_t x} \\
&= \exp\Big[\alpha_i \ln(\bar{\alpha}_t) - \alpha_i \ln(\bar{x}_t) \\
&\quad - \ln(\Gamma(\alpha_i)) + (\alpha_i - 1)\ln(x) \\
&\quad - \bar{\alpha}_t \left(\frac{x}{\bar{x}_t}\right) \Big]
\end{aligned}
$$

where,

$$\lambda_t = \frac{\bar{\alpha}_t}{\bar{x}_t}$$

and,

$$\bar{\alpha}_t = w_{1t}\alpha_1 + w_{2t}\alpha_2 + \ldots + w_{It}\alpha_I$$

So,

$$
\begin{aligned}
\frac{\partial f_{it}(x)}{\partial \bar{x}_t} &= \exp\Big[\alpha_i \ln(\bar{\alpha}_t) - \alpha_i \ln(\bar{x}_t) \\
&\quad - \ln(\Gamma(\alpha_i)) + (\alpha_i - 1)\ln(x) \\
&\quad - \bar{\alpha}_t \left(\frac{x}{\bar{x}_t}\right) \Big] \cdot \\
&\quad \Big[\alpha_i \ln(\bar{\alpha}_t) - \alpha_i \ln(\bar{x}_t) \\
&\quad - \ln(\Gamma(\alpha_i)) + (\alpha_i - 1)\ln(x) \\
&\quad - \bar{\alpha}_t \left(\frac{x}{\bar{x}_t}\right) \Big]'
\end{aligned}
$$

and,

$$
\begin{aligned}
&\Big(\alpha_i \ln(\bar{\alpha}_t) - \alpha_1 \ln(\bar{x}_t) \\
&- \ln(\Gamma(\alpha_i)) + (\alpha_i - 1)\ln(x) \\
&- \bar{\alpha}_t \left(\frac{x}{\bar{x}_t}\right) \Big)' = \\
&= \left(\frac{-\alpha_i}{\bar{x}_t} + \bar{\alpha}_t \left(\frac{x}{\bar{x}_t^2}\right)\right) \\
&= \frac{\bar{\alpha}_t x - \alpha_i \bar{x}_t}{\bar{x}_t^2}
\end{aligned}
$$

Thus,

$$\frac{\partial f_{it}(x)}{\partial \bar{x}_t} = f_{it}(x) \cdot \left(\frac{\bar{\alpha}_t x - \alpha_i \bar{x}_t}{\bar{x}_t^2}\right)$$

Since, in a gamma pdf:

$$\frac{\alpha_i}{\lambda_t} = \bar{x}_{it}$$

and:

$$\lambda_t = \frac{\bar{\alpha}_t}{\bar{x}_t}$$

where $\bar{x}_{it}$ is the mean of $f_{it}(x)$, the conditional mean, as distinguished from $\bar{x}_t$, the unconditional mean. It follows that:

$$\frac{\partial f_{it}(x)}{\partial \bar{x}_t} = f_{it}(x) \cdot \left(\frac{\bar{\alpha}_t (x - \bar{x}_{it})}{\bar{x}_t^2}\right)$$

or:

$$\frac{\partial f_{it}(x)}{\partial \bar{x}_t} = f_{it}(x)\ a_{it}$$

where:

$$a_{it} = \left[\frac{\bar{\alpha}_t (x - \bar{x}_{it})}{\bar{x}_t^2}\right]$$