# InterGen:

# Technical Documentation[1]

by

Mark Fossett
Department of Sociology and
Racial and Ethnic Studies Institute
Texas A&M University
College Station, Texas

November 2001

## Summary

This document provides an overview of the methods and techniques the InterGen program uses to implement a quantitative model of the transmission of inter-group inequality across generations. It begins with a brief overview of the InterGen model. Then it presents discussions of how key elements of the model are implemented, including specifications of attainment equations used in the model and algorithms and techniques used to "calibrate" these equations in ways that implement the model's basic parameters while insuring that key assumptions about status distributions are satisfied.

[1] The development of the InterGen program, the VLAB-RESI website, and this document was supported by a grant from the National Science Foundation (Division of Undergraduate Education, Educational Materials Division). Additional support also was provided by the Department of Sociology and the Racial and Ethnic Studies Institute at Texas A&M University.

## Overview of the Simulation Model

The intergenerational model involves a relatively simple system of two attainment equations. Complexity arises from the fact that the equations "feed" into each other and indirectly back into themselves. Four variables are involved in the equations.

P-ED – Parent's educational status,

P-SES – Parent's socio-economic status,

R-ED – Respondent's educational status, and

R-SES – Respondent's socio-economic status.

In the InterGen model, education and socio-economic status are both measured as percentile scores. By definition, percentile scores range from 0 to 100 and follow a "rectangular" probability distribution. The mean and median of a percentile distribution always equal 50; the standard deviation is 28.9.[2] Percentile scores index rank position, a key focus of the model, and they provide a simple, intuitive metric for measuring outcomes over time and across groups and generations.

The InterGen model has a temporal dimension. Each of the four variables is a dependent variable in an equation at some point in time. In this situation, they can be described as "outcomes" and "attainments". Each variable also serves as an independent variable at some point in time. In this situation they can be described as "resources" that contribute to attainments.

There are two basic equations in the model, one for education and one for socio-economic status. They are

R-ED = P-ED + P-SES, and

R-SES = R-ED + P-SES.

The temporal aspect of the model provides for "cohorts" to progress through a generational life cycle with four distinct stages. Each cohort is "born" in Stage 1 and during it completes a process of educational attainment (R-ED) represented by the first equation. In Stage 2 the cohort enters the labor force and begins a process of socio-economic status attainment (R-SES) represented by the second equation. In this

[2] The standard deviation is obtained from the formula $(100^2/12)^{1/2}$ given in Freund (1962:146).

same stage the cohort has children and they complete a process of educational attainment. In Stage 3 the cohort completes its participation in the labor force, continuing to attain socio-economic status. In Stage 4 the cohort exits the labor force.

When cohorts are in Stages 2 and 3 of the life cycle, their education and socio-economic status attainments (i.e., R-ED and R-SES), take on relevance for the education and socio-economic status attainments of their children, the next generation in the model. Because of this, status outcomes (dependent variables) for one generation always become status resources (independent variables) for the next generation. That is, R-ED for Generation t becomes P-ED for Generation t+1.

Race differences in attainment outcomes are a major focus of the InterGen model. Consequently, the model allows for the possibility that the process just described will differ by race. This means that each attainment equation may differ by race. Significantly, this group-specific variation in attainment is structured in such a way that the "overall" attainment distributions always conform to the assumption of percentile distributions. That is, each attainment outcome (e.g., education) for every generation always follows a percentile distribution. Group-specific variations in the attainment process serve to "arrange" groups within this distribution. This has implications for how the equations in the model are implemented and this will be made clearer in the next sections that review the formal specification of the attainment process for education.

## Education Attainment

The following linear regression equations represent the individual-level attainment process for education:[3]

$$\text{R-ED}_T = b_{0_T} + b_{1_T}(\text{P-ED}_T) + b_{2_T}(\text{P-SES}_T), \quad [1]$$

$$\text{R-ED}_W = b_{0_W} + b_{1_W}(\text{P-ED}_W) + b_{2_W}(\text{P-SES}_W), \text{ and} \quad [2]$$

$$\text{R-ED}_B = b_{0_B} + b_{1_B}(\text{P-ED}_B) + b_{2_B}(\text{P-SES}_B). \quad [3]$$

The subscripts T, W, and B signify the total population, the white population and the black population, respectively. Thus, $\text{P-ED}_i$ stands for parent's education for members of group i, $\text{P-SES}_i$ stands for parent's socio-economic status for members of group i, $b_{0_i}$ is the intercept of the education attainment equation for group i, $b_{1_i}$ is the slope for parent's education for group i, and $b_{2_i}$ is the slope for parent's socio-economic status for group i.

One of the properties of linear regression is that regression equations yield the mean for the dependent variable when the means for the independent variables are specified in the right-hand side of the equation. Thus, the following relations hold:

$$\textbf{R-ED}_T = b_{0_T} + b_{1_T}\textbf{P-ED}_T + b_{2_T}\textbf{P-SES}_T, \quad [4]$$

$$\textbf{R-ED}_W = b_{0_W} + b_{1_W}\textbf{P-ED}_W + b_{2_W}\textbf{P-SES}_W, \text{ and} \quad [5]$$

$$\textbf{R-ED}_B = b_{0_B} + b_{1_B}\textbf{P-ED}_B + b_{2_B}\textbf{P-SES}_B, \quad [6]$$

where $\textbf{R-ED}_i$ stands for the *mean* for respondent's education for group i, $\textbf{P-ED}_i$ is the *mean* for parent's education for members of group i, $\textbf{P-SES}_i$ is the *mean* for parent's socio-economic status for members of group i, and where $b_{0_i}$, $b_{1_i}$, and $b_{2_i}$ are as defined earlier.

By definition, the average percentile score on education for the total population is 50. This can also be expressed as the weighted sum of the average percentile scores of the white and black populations as follows:

$$50 = \textbf{R-ED}_T = p\cdot\textbf{R-ED}_W + q\cdot\textbf{R-ED}_B . \quad [7]$$

Based on Equations 5 & 6 above, this can be re-stated as

$$50 = p\cdot(b_{0_W} + b_{1_W}\textbf{P-ED}_W + b_{2_W}\textbf{P-SES}_W) + q\cdot(b_{0_B} + b_{1_B}\textbf{P-ED}_B + b_{2_B}\textbf{P-SES}_B) \quad [8]$$

where p is the proportion white in the total population and q is the proportion black in the total population, and $\textbf{R-ED}_T$, $\textbf{R-ED}_W$, and $\textbf{R-ED}_B$ are as defined above.

As will be seen later, this equation is used in the process of solving to obtain effect parameters for Equations 5 & 6 that satisfy the requirements of regression analysis and the attainment system.

## Calculating Values for Model Parameters: Education

The effect parameters $b_{0_W}$, $b_{1_W}$, and $b_{2_W}$ for Equation 5, the education attainment equation for whites, and the effect parameters $b_{0_B}$, $b_{1_B}$,

[3] These and other linear regression equations introduced here share all of the standard characteristics of regression equations save one – the form of the distribution of the disturbance (error) term. This point is addressed in more detail in a later section of the paper.

and $b_{2_B}$ for Equation 6, the education attainment equation for blacks are obtained by an iterative, multi-step, numerical search.

**Start Values**

In the first step the *initial* (but not necessarily final) estimates of effect parameters for whites are obtained by taking the values that would obtain in a population of *whites only* where the "impact coefficient" of the variables in the education attainment equation is given by $IC_{ED}$ – a parameter of the InterGen model.

Substantively, this model parameter controls the "strength" of the connection between resources (parent's education and parent's socio-economic status) and outcomes (respondent's education) in the education attainment process. Formally, $IC_{ED}$ is the increase in educational attainment that results when *all* resource variables (i.e., independent variables) increase by 1 point. Since the variables are measured on a percentile scale, the impact coefficient registers the extent to which rank position on resources (i.e., predictors) translates into rank position on education.[4] For example, an impact coefficient of 30% indicates that improvements of 1 point (an increase in rank position of one percentile) on all predictors of education translates into an improvement of 0.3 points on education (an increase in rank position of three-tenths of one percentile).

The InterGen model specifies that there are two independent variables (P-$ED_W$ and P-$SES_W$) in the equation for education (R-$ED_W$). It also specifies that these two predictors have equal effects on the dependent variable.[5] Since the effects of the predictors are additive, the overall impact coefficient is the sum of the impact coefficients for the separate predictors. Thus, in the example at hand, the overall impact coefficient of 30% (or 0.30) represents the sum of separate impact coefficients of 0.15 for both predictors. Restated from the other direction, the impact coefficients for both predictors are equal to one-half the value of the overall impact coefficient ($IC_{ED}$).

Those who are familiar with the terminology of regression equations will recognize that the impact coefficients for the separate predictors are the unstandardized regression coefficients $b_{1_W}$ and $b_{2_W}$ in Equation 5. Once these are established, the regression intercept $b_{0_W}$ for the equation is also established and thus $b_{0_W}$, $b_{1_W}$, and $b_{2_W}$ are given by

$$b_{1_W} = IC_{ED} / 2 \quad [9]$$

$$b_{2_W} = IC_{ED} / 2 \quad [10]$$

$$b_{0_W} = 50.0 - b_{1_W} \cdot 50.0 - b_{2_W} \cdot 50.0. \quad [11]$$

Equation 11 is a restatement of the standard equation for the constant or "y-intercept" of a regression equation

$$b_0 = \mathbf{Y} - b_1 \cdot \mathbf{X}_1 - b_2 \cdot \mathbf{X}_2$$

where $\mathbf{Y}$, $\mathbf{X}_1$, and $\mathbf{X}_2$ represent *means* on the dependent and independent variables. The values of 50.0 in Equation 11 are dictated by the assumption (adopted only for this initial step) that the population consists only of whites, in which case, the means on all variables are necessarily equal to 50.0. Application of Equation 11 is straightforward. For example,

if $IC_{ED}$ is 0.40, $b_{0_W} = 30$, $b_{1_W} = 0.20$, and $b_{2_W} = 0.20$;

if $IC_{ED}$ is 0.50, $b_{0_W} = 25$, $b_{1_W} = 0.25$, and $b_{2_W} = 0.25$;

if $IC_{ED}$ is 0.60, $b_{0_W} = 20$, $b_{1_W} = 0.30$, and $b_{2_W} = 0.30$;

and so on. The application of these relations yields initial estimates of the effect parameters of the education equation for whites.

Next, *initial* (not necessarily final) estimates of the effect parameters $b_{0_B}$, $b_{1_B}$ and $b_{2_B}$ of the education attainment equation for blacks are obtained by taking the corresponding effect parameters for whites and multiplying them by 1.0 - $Disc_{ED}$ where $Disc_{ED}$ represents the impact of discrimination on black educational attainment, a parameter of the InterGen model. If $Disc_{ED}$ is 0, then 1 - $Disc_{ED}$ is 1.0 and the parameters of the education attainment equation are the same for whites and blacks. If $Disc_{ED}$ is > 0, then blacks convert status resources into education attainments at a *lesser* rate than whites and the relative disparity will be the same at any combination of status resources. Thus, the following expressions are used

$$b_{0_B} = (1.0\text{-}Disc_{ED})\, b_{0_W} \quad [12]$$

[4] The impact coefficient ($IC_{ED}$) has have a second, more technical interpretation. If the population consisted of *only* whites, it is equal to the $R^2$ statistic for the regression equation for education for whites.

[5] This is the only possibility in the Java applet version of the InterGen program. The standalone version of the InterGen program for windows allows the user to vary the impact of parent's education and parent's socio-economic status such that they contribute *unequally* to the determination of education, but still generate the same overall impact coefficient.

$$b_{1_B} = (1.0\text{-}Disc_{ED})\, b_{1_W} \quad [13]$$

$$b_{2_B} = (1.0\text{-}Disc_{ED})\, b_{2_W} \quad [14]$$

When these values (i.e., $b_{0_B}$, $b_{1_B}$, and $b_{2_B}$) are applied in Equation 3, the equation generates predictions for respondent's education for blacks ($PV\text{-}R\text{-}ED_B$) that equal $(1.0\text{-}Disc_{ED}) \cdot PV\text{-}R\text{-}ED_W$, where $PV\text{-}R\text{-}ED_W$ is the predicted value of respondent's education for whites based on Equation 2. Significantly, this relationship will hold for any combination of values of the independent variables in the education attainment equation (i.e., any values for the social background characteristics P-ED and P-SES).

If two conditions are met, the right-hand side of Equation 8 will yield exactly 50 using these initial values of the effect parameters based on Equations 9-14. The first condition is that the levels of parental education and parental socio-economic status are the same for whites and blacks.[6] The second is that there is no discrimination in educational attainment (i.e., $Disc_{ED}$ is 0.0). When these conditions hold, the process for calculating coefficients for the group-specific attainment equations for education can stop at this point. Usually, however, one or both conditions are not met and further steps must be taken to calculate the final values of the coefficients.

### Calculating Final Values

If the initial estimates of the coefficients in Equation 8 are correct, the right-hand side of the equation will yield a result ($Y_T$*) of exactly 50. This is necessary for the system of equations in the inter-generational attainment model to satisfy the model's assumptions regarding the distribution of education in the total population. To clarify, education is measured by percentile scores and thus by definition must have means of 50 for the total population in every generation. Consequently, the group-specific attainment equations that generate new education distributions for each generation must produce means of 50 for the total population.

When the model parameters include discrimination in education, or when the white and black means on parent's education and parent's socio-economic status are not equal, the right-hand side of Equation 8 will *not* yield a result ($Y_T$*) of exactly 50 using the initial coefficient estimates. This signals that the coefficients must be "re-calibrated". This is accomplished by performing an iterative numerical search using the following steps.

First, compute $Y_T$* using the right-hand side of Equation 8 with the initial coefficient estimates.

Next adjust the term $b_{0_W}$, increasing it if $Y_T^* < 50$, decreasing it if $Y_T^* > 50$.

Then reset $b_{0_B}$ to $(1.0\text{-}Disc_{ED})\, b_{0_W}$.

Then recalculate $Y_T$* and continue the process until $Y_T$* equals 50.[7]

The recalibrated set of coefficients has two important properties. First, it insures that Equation 8 yields a result of exactly 50 which in turn guarantees that the attainment process produces a status distribution that satisfies the model's distributional assumptions. Second, it maintains the desired discrimination relationship wherein the predicted value of education for blacks ($PV\text{-}R\text{-}ED_B$) equals $(1.0\text{-}Disc_{ED}) \cdot PV\text{-}R\text{-}ED_W$ under any combination of values for the independent variables in the education attainment equation (i.e., any values for the resource characteristics P-ED and P-SES).

## Socio-Economic Status Attainment

The following linear regression equations represent the individual-level attainment process for socio-economic status:

$$R\text{-}SES_T = z_{0_T} + z_{1_T}(R\text{-}ED_T) + z_{2_T}(P\text{-}SES_T), \quad [15]$$

$$R\text{-}SES_W = z_{0_W} + z_{1_W}(R\text{-}ED_W) + z_{2_W}(P\text{-}SES_W), \text{ and} \quad [16]$$

$$R\text{-}SES_B = z_{0_B} + z_{1_B}(R\text{-}ED_B) + z_{2_B}(P\text{-}SES_B). \quad [17]$$

As before, the subscripts T, W, and B signify the total population, the white population and the black population, respectively. Thus, $R\text{-}ED_i$ stands for respondent's education for members of group i, $P\text{-}SES_i$ stands for parent's socio-economic status for members of group i, $z_{0_i}$ is the intercept of the socio-economic status attainment equation for group i, $z_{1_i}$ is the slope for respondent's education for group i, and $z_{2_i}$ is the slope for parent's socio-economic status for group i.

[6] If this is true, the means for both groups will be 50.0.

[7] InterGen stops the process when $Y_T$* is within 0.00001 of 50.0.

Again drawing on the properties of regression analysis, the means for the group-specific means on respondent's socio-economic status can be expressed as follows:

$$\mathbf{R\text{-}SES}_T = z_{0_T} + z_{1_T}\mathbf{R\text{-}ED}_T + z_{2_T}\mathbf{P\text{-}SES}_T, \quad [18]$$

$$\mathbf{R\text{-}SES}_W = z_{0_W} + z_{1_W}\mathbf{R\text{-}ED}_W + z_{2_W}\mathbf{P\text{-}SES}_W, \text{ and} \quad [19]$$

$$\mathbf{R\text{-}SES}_B = z_{0_B} + z_{1_B}\mathbf{R\text{-}ED}_B + z_{2_B}\mathbf{P\text{-}SES}_B, \quad [20]$$

where $\mathbf{R\text{-}SES_i}$ stands for the *mean* for respondent's socio-economic status for group i, $\mathbf{R\text{-}ED}_i$ is the *mean* for respondent's education for members of group i, $\mathbf{P\text{-}SES}_i$ is the *mean* for parent's socio-economic status for members of group i, and where $z_{0_i}$, $z_{1_i}$, and $z_{2_i}$ are as defined earlier.

By definition, the average percentile score on respondent's socio-economic status for the total population is 50. This can also be expressed as the weighted sum of the average percentile scores of the white and black populations as follows:

$$50 = \mathbf{R\text{-}SES}_T = p\cdot\mathbf{R\text{-}SES}_W + q\cdot\mathbf{R\text{-}SES}_B\,. \quad [21]$$

Based on Equations 19 & 20 above, this can be re-stated as

$$50 = p\cdot(z_{0_W} + z_{1_W}\mathbf{R\text{-}ED}_W + z_{2_W}\mathbf{P\text{-}SES}_W) + q\cdot(z_{0_B} + z_{1_B}\mathbf{R\text{-}ED}_B + z_{2_B}\mathbf{P\text{-}SES}_B)\,. \quad [22]$$

As will be seen later, this equation is used in the process of solving to obtain effect parameters for Equations 19 & 20 that satisfy the requirements of regression analysis and the attainment system.

## Calculating Values for Model Parameters: Socio-Economic Status

The effect parameters $z_{0_W}$, $z_{1_W}$, and $z_{2_W}$ for Equation 19, the socio-economic status attainment equation for whites, and the effect parameters $z_{0_B}$, $z_{1_B}$, and $z_{2_B}$ for Equation 20, the socio-economic status attainment equation for blacks are obtained by an iterative, multi-step, numerical search.

### Start Values

In the first step the *initial* (but not necessarily final) estimates of effect parameters for whites are obtained by taking the values that would obtain in a population of *whites only* where the "impact coefficient" of the variables in the socio-economic attainment equation is given by $IC_{SES}$ – a parameter of the InterGen model.

Substantively, this model parameter controls the "strength" of the connection between "resources" and socio-economic status attainments. Formally, $IC_{SES}$ is the increase in socio-economic status that results when *all* resource variables (i.e., independent variables) increase by 1 point. Since the variables are measured on a percentile scale, the impact coefficient registers the extent to which rank position on predictors translates into rank position on socio-economic status.[8] For example, an impact coefficient of 30% indicates that improvements of 1 point (an increase in rank position of one percentile) on all predictors of socio-economic status translates into an improvement of 0.3 points (an increase in rank position of three-tenths of one percentile) in socio-economic status.

The InterGen model specifies that there are two independent variables ($R\text{-}ED_W$ and $P\text{-}SES_W$) in the equation for socio-economic status ($R\text{-}SES_W$). It also specifies that these two predictors have equal effects on the dependent variable.[9] Since the effects of the predictors are additive, the overall impact coefficient is the sum of the impact coefficients for the separate predictors. Thus, in the example at hand, the overall impact coefficient of 30% (or 0.30) represents the sum of separate impact coefficients of 0.15 for both predictors.

As before, this implies that the impact coefficients for the separate predictors are the unstandardized regression coefficients $z_{1_W}$ and $z_{2_W}$ in Equation 19. Once these are established, the regression intercept $z_{0_W}$ for the equation can be computed. Following the earlier example, $z_{0_W}$, $z_{1_W}$, and $z_{2_W}$ are given by

$$z_{1_W} = IC_{SES} / 2 \quad [23]$$

$$z_{2_W} = IC_{SES} / 2 \quad [24]$$

[8] As before, the impact coefficient ($IC_{SES}$) has have a second, more technical interpretation. If the population consisted of *only* whites, it is equal to the $R^2$ statistic for the regression equation for socio-economic status for whites.

[9] This is the only possibility in the Java applet version of the InterGen program. The standalone version of the InterGen program for windows allows the user to vary the impact of parent's education and parent's socio-economic status such that they contribute *unequally* to the determination of education, but still generate the same overall impact coefficient.

$$z_{0_W} = 50.0 - z_{1_W} \cdot 50.0 - z_{2_W} \cdot 50.0 \quad [25]$$

The application of these relations yields *initial* estimates of the effect parameters of the socio-economic status equation for whites.

Next, InterGen obtains *initial* (not necessarily final) estimates of the effect parameters $z_{0_B}$, $z_{1_B}$ and $z_{2_B}$ of the socio-economic status attainment equation for blacks by taking the corresponding effect parameters for whites and multiplying them by $1.0 - Disc_{SES}$ where $Disc_{SES}$ represents the impact of discrimination on black socio-economic status attainment, a parameter of the InterGen model. If $Disc_{SES}$ is 0, then 1 - $Disc_{SES}$ is 1.0 and the parameters of the socio-economic status attainment equation are the same for whites and blacks. If $Disc_{SES}$ is > 0, then blacks convert attainment resources into socio-economic status at a *lesser* rate than whites. Thus, the following expressions are used

$$z_{0_B} = (1.0\text{-}Disc_{SES})\, z_{0_W} \quad [26]$$

$$z_{1_B} = (1.0\text{-}Disc_{SES})\, z_{1_W} \quad [27]$$

$$z_{2_B} = (1.0\text{-}Disc_{SES})\, z_{2_W} \quad [28]$$

When these values (i.e., $z_{0_B}$, $z_{1_B}$, and $z_{2_B}$) are applied in Equation 17, the equation generates predictions for respondent's socio-economic status for blacks ($PV\text{-}R\text{-}SES_B$) that equal $(1.0\text{-}Disc_{SES}) \cdot PV\text{-}R\text{-}SES_W$, where $PV\text{-}R\text{-}SES_W$ is the predicted value of respondent's socio-economic status for whites based on Equation 16. Significantly, this relationship will be exact for any combination of values of the independent variables in the socio-economic status attainment equation (i.e., any values for the characteristics R-ED and P-SES).

If two conditions are met, the right-hand side of Equation 22 will yield a value of exactly 50 using the initial estimates of the effect parameters based on Equations 23-28. The first condition is that the levels of respondent's education and parent's socio-economic status are the same for whites and blacks.[10] The second is that there is no discrimination in socio-economic status attainment (i.e., $Disc_{SES}$ is 0.0). When these conditions hold, the process for calculating coefficients for the group-specific attainment equations for socio-economic status can stop at this point. Usually, however, one or both conditions are not met and further steps must be taken to calculate the at final values of the coefficients.

### Calculating Final Values

When the coefficients in Equation 22 are calibrated correctly, the right-hand side of the equation will yield a result ($Y_T$*) of exactly 50. This is necessary for the system of equations in the inter-generational attainment model to satisfy the model's assumptions regarding the distribution of socio-economic status in the total population. As is the case for education, socio-economic status is measured by percentile scores and by definition must have a mean of 50 for the total population in every generation. Consequently, the group-specific attainment equations that generate new socio-economic status distributions for each generation must produce means of 50 for the total population.

When the model parameters include discrimination in socio-economic status, or when the white and black means on respondent's education and parent's socio-economic status are not equal, the right-hand side of Equation 22 will not yield a result ($Y_T$*) of exactly 50 using the initial coefficient estimates. This signals that the coefficients need to be "re-calibrated". This is accomplished by performing an iterative numerical search using the following steps.

First, compute $Y_T$* using the right-hand side of Equation 22 with the initial coefficient estimates.

Next, adjust the term $z_{0_W}$, increasing it if $Y_T^* < 50$, decreasing it if $Y_T^* > 50$.

Then reset $z_{0_B}$ to $(1.0\text{-}Disc_{SES})\, z_{0_W}$.

Then recalculate $Y_T$* and repeat the process until $Y_T$* equals 50.[11]

The recalibrated set of coefficients has two important properties. First, it insures that Equation 22 yields a result of exactly 50. Second, it maintains the desired discrimination relationship wherein the predicted value of socio-economic status for blacks ($PV\text{-}R\text{-}SES_B$) equals $(1.0\text{-}Disc_{SES}) \cdot PV\text{-}R\text{-}SES_W$ under any combination of values for the independent variables in the socio-economic status attainment equation (i.e., any values for the resource characteristics R-ED and P-SES).

[10] If this is true, the means for both groups will be 50.0.

[11] InterGen repeats the process until $Y^*_T$ is within 0.00001 of 50.0.

## Departure from Standard Regression Assumptions

The regression equations introduced in the sections above share all of the standard characteristics of regression equations with one exception. The distribution of the disturbance term of the equation (i.e., the error distribution) is not normal. The reason for this is simple. The dependent variables are all measured as percentiles which, by definition, are bounded, stopping at 0 on the low end and 100 and the high end. Normal distributions have infinite tails so the standard regression assumption of normally distributed errors would imply scores that are outside the logical bounds of percentile scores.

As a practical matter, this problem only comes into play when predictions from the regression equations begin to approach the upper and lower boundaries of the percentile distributions. Such extreme predictions only occur when the model's impact coefficient settings ($IC_{ED}$ and $IC_{SES}$) are very high (e.g., 85% or above). To insure that this does not occur, the available settings for the impact coefficients are restricted to the range of 0-80% - a range that is more than satisfactory for exploring status inheritance dynamics and their impact on the inter-generational transmission of inequality.

A more technically correct model would specify an error distribution that is compatible with the bounded range of the percentile scores. One possibility would be to specify the equations using logit-style transformations of the percentile scores.[12] This non-linear transformation would allow for linear regression equations with normally distributed error terms such that all logically possible outcomes would, when converted back to percentile scores (via a reverse logit transformation) within the logical bounds of 0-100.

Unfortunately, a technical solution along these lines would compromise one of the major goals of the InterGen model, that of representing attainment processes in a way that is readily intelligible to students and laypersons who have limited backgrounds in statistical modeling. In view of this, I have chosen to use the less technically correct model specification (i.e., linear equations with bounded dependent variables in the form of percentile scores). I acknowledge that the behavior of the model will violate some technical regression assumptions under extreme conditions and I minimize the practical relevance of this violation of technical assumptions by imposing constraints on the model (i.e., restrictions on the impact coefficients) to insure that the extreme conditions do not come into play.

## Substantive Representation of Attainment and Ethnic Stratification

The InterGen model presents status attainment dynamics from a point of view that focuses on whites and considers minority attainment in terms of departure from whites. The choice is intentional and has clear substantive implications. Status attainment dynamics in the total population are seen as reflecting a stratification system that is shaped by the dominant ethnic group (whites) and includes attainment rules that generate stratification within that subpopulation and also (if specified) additional rules that impact minority outcomes.

Given this point of view, the four basic model parameters (i.e., $Disc_{ED}$, $Disc_{SES}$, $IC_{ED}$, and $IC_{SES}$) have the following interpretations:

- $IC_{ED}$ – The impact coefficient governing the education attainment equation controls the important of status resources for educational attainment for whites. Since the resources for educational attainment are parent's education and parent's socio-economic status, this setting regulates the strength of status inheritance dynamics for whites.
- $IC_{SES}$ – The impact coefficient governing the status attainment equation controls the important of status resources for socio-economic status attainment for whites. Since the resources for socio-economic status attainment are respondent's education and parent's socio-economic status, this setting regulates both the strength of "merit-based" stratification (the effect of respondent's education) and the strength of status inheritance dynamics (the effect of parent's socio-economic status.
- $Disc_{ED}$ – The impact of discrimination in minority education attainment controls the relative efficacy (compared to whites) of status resources for education attainment by blacks.
- $Disc_{SES}$ – The impact of discrimination in minority socio-economic status attainment controls the relative efficacy (compared

[12] The logit-style transformation would take the form ln(Y/(100-Y) where Y is the percentile score (and where Y approaches but never actually reaches the lower and upper boundaries of 0 and 100).

to whites) of status resources for socio-economic status attainment by blacks.

## Some Observations

The InterGen model is one in which attainment outcomes for one generation become resources that shape attainment outcomes for later generations. Those who are new to the study of social stratification may be interested to note the following points that are implied by such a model.

- Discrimination that impacts minorities and creates inequality at one point in time is transmitted indirectly to future generations through status inheritance dynamics that serve to perpetuate the inequality.
- Assuming discrimination is ongoing, the resulting disadvantages are compounded over successive generations and inequality will increase over time.
- If the basic parameters of the model are left unchanged, the system will move toward stable equilibrium values for inequality and group status levels.
- Furthermore, every unique set of values for the model's basic parameters will generate a particular equilibrium outcome for ethnic inequality and group status levels.
- And, this equilibrium condition will emerge regardless of the prior state of the system. Thus, if model parameters changed, the system will move toward the new set of equilibrium values associated with those model parameters regardless of what the prior parameter settings for the model were.

This leads to the following conclusions about the InterGen stratification system: The long-term equilibrium levels of education and socio-economic status for whites and blacks, and hence the average education and socio-economic status differences between the groups depends on (a) the important of status resources for education, (b) the importance of status resources for socio-economic status, (c) the relative efficacy of status resources for blacks in educational attainment, and (d) the relative efficacy of status resources for blacks in status attainment. *Significantly, interventions at any of these four points controls the long-term white-black difference in mean education and the long-term white-black difference in mean status.*

Changing values in these four controlling parameters allows one to assess the impact of "virtual" interventions in the areas of (a) varying the strength of status inheritance dynamics, (b) varying the degree to which status attainments are merit based, (c) varying discrimination in education, and (d) varying discrimination in status attainment. Several important conclusions about the behavior of this system can be offered here.

- In the long-term, interventions that reduce discrimination drive the system toward equality.
- In the long-term, interventions that reduce the strength of status inheritance dynamics quickens the pace with which the effects of the equality-promoting intervention just noted (i.e., reductions in discrimination) are realized.
- In the long-term, if discrimination is present, interventions that reduce the strength of status inheritance dynamics will reduce the magnitude of group inequality.

Many other observations could be offered. But the ones just listed are among the more basic and important that can be noted. The reader is invited to use the InterGen model to explore these issues and to see what other observations are appropriate.

## References

Freund, John E., 1962. *Mathematical Statistics*. Prentice-Hall.