Abstract

This paper presents a personal income tax revenue forecasting model which incorporates a functional specification of the entire size distribution of income estimated using aggregate data. The model is shown to yield substantially more accurate forecasts for Minnesota than other revenue forecasting models which do not incorporate a functional specification of the income distribution. Although the choice of a revenue forecasting model is somewhat subjective, our results indicate that the general approach of modeling the entire income distribution by a functional form is highly desirable in terms of forecast accuracy.

THE SIZE DISTRIBUTION OF INCOME IN AN INCOME TAX REVENUE FORECASTING MODEL

WILLIAM C. WEILER
University of Minnesota

An income tax revenue forecasting model should account for two factors related to the distribution of income which can affect tax revenues. First, incomes of individual taxpayers generally grow over time, which pushes them into higher tax brackets in a progressive tax structure. As cogently described by Neil Singer (1970), this "rate effect" has a systematic impact on tax revenues over a wide range of individual incomes. Second, the number of taxpayers varies over time. To the extent that new taxpayers have an income distribution different from that of individuals who are already taxpayers, a change in gross income resulting from additional taxpayers is unlikely to produce the same revenues as the identical change resulting from increasing incomes of previous taxpayers.

In addition, a revenue forecasting model should have the capacity to forecast revenues under alternative effective tax

AUTHOR'S NOTE: I wish to thank Colleen T. Davidson, Stephen A. Hoenack, William S. Reece, David L. Sjoquist, and two referees for exceptionally helpful comments on earlier drafts of this paper.

PUBLIC FINANCE QUARTERLY, Vol. 7 No. 2, April 1979, 182-198

rates. The ability to simulate revenues under alternative rate structures is a useful feature, especially at the state level.[1]

We will develop a simple income tax revenue forecasting model for Minnesota which includes specification of the size distribution of income and has the desired simulation properties. In the model, total individual income tax liability is calculated as the sum of the products of the effective tax rates at each level of income and the total dollar income taxed in each rate class.[2] The latter is the product of the proportion of taxpayers in the class, the average individual income level in the class, and the total number of taxpayers. The proportion of taxpayers in each income class is determined through a statistical model of the complete size distribution of income. The model is conceptually similar to Reaume's (1978) model for North Carolina, but is simpler in construction and use.[3] We show that our model provides substantially more accurate forecasts for Minnesota than forecasting models suggested elsewhere which do not incorporate a functional specification of the income distribution. Further, comparison of our forecast accuracy with Reaume's results for North Carolina suggests that his model can be simplified with a relatively slight loss in forecast accuracy. These comparisons confirm the desirability of the general approach of including a functional specification of the income distribution in an income tax revenue forecasting model.

The paper is organized as follows. The next section is a description of the model. Econometric estimates of the equations in the model are presented in the second section. The third section includes a comparison of the forecast accuracy of the model relative to other revenue projection models and a discussion of the sensitivity of forecasts to the estimated parameters of the income distribution. The final section is a brief conclusion.

THE REVENUE FORECASTING MODEL

The identity relating total revenue to the size distribution of income, the total number of taxpayers, and the tax rates is

$$R \equiv \sum_{i=1}^{n} f_i N y_i^* r_i \qquad [1]$$

where R is total tax revenue, N is the total number of taxpayers, and for each income class i f_i, y_i^* and r_i are the proportion of taxpayers, the average income, and the effective tax rate, respectively. To forecast revenue using this identity, we must forecast the values of the variables in the summation. The average income level and the effective tax rate for each income class are considered exogenous in this model. The tax rates are policy variables to be set at the levels desired for the forecast. There has been little variation in the average individual incomes in the relatively narrow Minnesota income brackets over time, and the averages have been near the midpoints of the brackets. Thus, for forecasting purposes it seems reasonable to assume that each average is the midpoint of the respective bracket. Of course, alternative values, either estimated or assumed, could be used, but a procedure for generating them is not part of the model.

THE INCOME DISTRIBUTION

The endogenous variables are the proportion of taxpayers in each income class and the total number of taxpayers. With regard to the former, income distributions in many countries are positively skewed so that most analyses have used variants of the lognormal or Pareto distribution functions to model them.[4] A few studies have also utilized hybrid functional forms, most commonly a combination of the Pareto and normal or the Pareto and lognormal distribution functions, for different parts of the income distribution.[5] We utilize the displaced lognormal distribution as the best functional form for this analysis.[6] This distribution, which has been used in other studies of the size distribution of income (especially Metcalf, 1972), is a simple extension of the lognormal distribution where the range of the variate y is $-c < y < \infty$ where c is a constant greater than zero. It is a three-parameter distribution that is completely specified by its mean (mu) and standard deviation (sigma) plus the constant of dis-

placement, c.[7] The proportion of taxpayers in each income class can be easily calculated from these parameters, since knowledge of them uniquely identifies the density function of the distribution.

We treat the parameters of the distribution as endogenous variables to be estimated as functions of macroeconomic variables. The distribution function can be specified by mu, sigma, and c or by three quantile income levels or the first three moments of the absolute data.[8] We use the quantile income levels as the endogenous variables for several reasons. First, when the data are grouped, the direct estimation of moments is less efficient than using quantile levels to determine the parameters (Metcalf, 1972, esp. appx. to ch. 2). Second, the quantile levels are directly observable from the absolute data while mu, sigma, and c are derived under the assumption of a displaced lognormal distribution. Third, the quantiles are money income levels in the distribution so that the independent variables are those which influence the relative position of the distribution, a well-defined if not simple problem. However, it is difficult to think of economic variables that would be similarly related to mu, sigma, and c.

The second quantile income level can be chosen as the median; however, we may select any symmetric quantile levels about the median as the first and third. Since the resultant values of mu, sigma, and c depend on the particular quantile levels that are chosen, we must specify criteria for choosing them. Because our primary consideration is revenue forecasting accuracy, our criterion is to minimize the deviations between actual and estimated tax revenues over our data sample period. For each year we used several quantile values to estimate the parameters of the distribution function, calculated the resultant proportion of taxpayers in each income class, and inserted these proportions and the actual values of the other variables into equation 1 to compute tax revenues. Using the median income and the third and ninety-seventh percentile income levels as the three quantile income levels minimized the average yearly error (2%) between estimated and actual tax revenues over the sample period.

The independent variables in the estimating equations for the quantile income levels include the components of Minnesota personal income that are taxable: wages and salaries, proprietors' income, and dividends, interest, and rents. We expect that the coefficients of these variables will be positive. Because the income distribution of new taxpayers is unlikely to coincide with that of current taxpayers, the number of taxpayers is also an independent variable. The effect of this variable is indeterminate because the negative effect of new entrants in the labor force could be balanced by the potentially possitive effect of new immigrants. The unemployment rate could also affect the income distribution through its effect on individual incomes. The effect is also indeterminate because the negative effect of reducing taxable income of individuals incurring spells of unemployment could be balanced by the potentially positive effect of a reduction in the number of low income taxpayers if their taxable incomes fall below the minimum filing requirement. We also include dummy variables for tax law changes likely to alter the income distribution.

THE NUMBER OF TAXPAYERS

The other endogenous variable in equation 1 is the total number of taxpayers. Soltow (1955) found that aggregate income and state population were the causal factors influencing this variable in Wisconsin. Over the data period in his analysis, however, the tax laws which could influence the number of taxpayers did not change. Further, the strong relationship he found between the number of taxpayers and aggregate income was due to an unchanged high minimum income filing requirement.

The independent variables in our equations for the number of taxpayers are population over age 15, the unemployment rate, and dummy variables to account for changes in tax laws. Income variables were excluded, since the minimum filing requirement in Minnesota was very small over the sample period. The effect of population is expected to be positive. The effect of the unemployment rate is expected to be negative, since an increase in the unemployment rate is likely to decrease the number of low income filers.

THE COMPLETE MODEL

The complete revenue forecasting model consists of the revenue determination identity (equation 1), estimated equations for the three quantile income levels and for the number of taxpayers (equations 2a-3), and the formulae relating the proportion of taxpayers in each income class to the parameters of the distribution function (equations 4-8). Symbolically, the model is

$$R \equiv \sum_{i=1}^{n} f_i \, N \, y_i^* \, r_i \qquad [1]$$

$$Y3 = g(WAGE, PROPY, DIR, UN, TIME, D1, D2) \qquad [2a]$$

$$Y50 = g(WAGE, PROPY, DIR, UN, TIME, D1, D2) \qquad [2b]$$

$$Y97 = g(WAGE, PROPY, DIR, UN, TIME, D1, D2) \qquad [2c]$$

$$N = g(POP, UN, D1, D2) \qquad [3]$$

$$c = Y50 \, [(L \cdot H - 1)/(2 - L - H)] \qquad [4]$$

$$mu = 1n \, (Y50 + c) \qquad [5]$$

$$sigma = 1n \, [(L \cdot Y50 + c)/(Y50 + c)] \, / X \qquad [6]$$

$$z_i = (1n \, (y_{i+1} + c) - mu)/sigma \qquad [7]$$

$$f_i = (2\pi)^{-1/2} \int_{z_i}^{z_{i+1}} e^{-t^2/2} \, dt \qquad [8]$$

The variables R, f_i, N, y^*, and r_i were defined above. Y3, Y50, and Y97 are the third, fiftieth, and ninety-seventh percentile income levels from the absolute data. WAGE, PROPY, and DIR are wage and salary income, proprietors' income, and dividends, interest, and rents, respectively. UN is the Minnesota unemployment rate. TIME is a time trend variable to control for omitted variables. POP is Minnesota population aged 15 and over. D1 and D2 are dummy variables for a minimum tax on gross income and for a tax law change which expedited the filing of separate returns. L and H are equal to Y3/Y50 and Y97/Y50, respectively. In the general form X is the inverse of the standard normal cumulative distribution evaluated at the percentile of the third quantile income level. In our case X = 1.881, the value at the ninety-seventh

percentile. The y_i's are the income levels defining the boundaries of the income classes ($y_o = 0$).

A revenue forecast is then made as follows. Values of the exogenous variables in the equations are obtained and the effective tax rates are determined through assumed policy decisions.[9] The total number of taxpayers is computed from equation 3. The quantile income levels are calculated from equations 2a, 2b, and 2c and inserted in equations 4-6 to calculate the parameters of the displaced lognormal distribution. Equation 7 then is used to calculate the appropriate integration limits for the standard normal cumulative distribution function ($z_o = -\infty$ and $z_{n+1} = \infty$ by definition) for computing the f_i's from equation 8. Tax revenue is then calculated using equation 1.

ESTIMATION OF THE MODEL

Regression estimates of equations 2a, 2b, and 2c are presented in Table 1. These equations include most of the variables discussed in the previous section. Five of the eighteen coefficients in the equations are not significant at the 5% level, but only one has a t-statistic below unity. When these variables were included in their respective equations, the significance of other coefficients was enhanced or autocorrelation problems were eliminated.

For each equation the signs of the income variables are positive. The median income equation was the only one significantly improved by using disaggregated income variables. The coefficients for the number of taxpayers are all negative. These results support the hypothesis that new taxpayers are probably new entrants in the labor force whose incomes are below those of previous taxpayers. The dummy variables for tax law changes have negative coefficients in four of five cases. Since both dummies account for tax law changes likely to increase the relative number of low income filers, negative coefficients would be expected. The insignificant negative and significant positive coefficients for the separate return dummy for Y50 and Y97, respectively, suggest that the effect of this tax law change was

TABLE 1[a]

Regression Results for Equations 2a, 2b, and 2c

Independent Variable	Dependent Variable Y3	Y50	Y97
N	-.000683 (2.84)[b]	-.00356 (20.63)	-.00193 (1.74)
UN	-14.718 (0.82)	-27.998 (1.35)	
TIME		139.942 (6.34)	
D1[c]	-499.602 (5.33)		-1226.132 (2.79)
D2[c]	-506.488 (8.13)	-135.337 (1.57)	971.254 (2.92)
(WAGE + PROPY)		.213 (8.43)	
DIR		.696 (3.59)	
(WAGE + PROPY + DIR)	.0288 (1.63)		1.120 (12.38)
Constant	1553.794 (9.63)	5359.493 (30.49)	8432.909 (18.87)
R^2[d]	.984	.996	.997
Durbin-Watson	2.46	2.39	3.00
Standard Error	49.099	70.152	320.840

SOURCES: The sources of Y50, Y3, Y97, N, D1 and D2 are the Minnesota Commissioner of Revenue (1977) and unpublished data provided by the Minnesota Department of Revenue. Income and unemployment for Minnesota were collected from various issues of the **Survey of Current Business.**

a. The equations were estimated using annual data for 1956-1961, 1963-1966, and 1968-1975, a total of 18 observations.

b. The absolute values of the t-statistics are shown beneath each estimated coefficient. The respective 5% significance levels (one-tail test) for these equations are 1.78, 1.80, and 1.77.

c. D1 is 1 for 1961-1966, 0 otherwise. D2 is 1 for 1971-1975, 0 otherwise.

d. R^2 is adjusted for degrees of freedom.

greatest on filers in the lower part of the income distribution. The unemployment rate coefficients are negative, but not statistically significant. These results suggest that the previously discussed opposing effects of the unemployment rate offset each other.

The regression estimates of equation 3 are shown in Table 2. The equations were estimated by entering the dummy variables in their usual forms and also by entering them multiplicatively with population.[10] The coefficients on population are positive and statistically significant in both equations. The coefficients of the dummy variables are positive and statistically significant whether entered in intercept form or interactively with population. The coefficients of the unemployment rate are negative and statistically significant. This result suggests that the probable effect of unemployment is to reduce the number of low income filers, which emphasizes the potentially offsetting effects of the unemployment rate in the quantile income equations. We chose the second equation in Table 2 for the model because it has a smaller standard error of forecast (evaluated at the 1975 values of the independent variables).

FORECASTS OF TAX REVENUE

We can classify income tax revenue forecasting models as follows:

(1) The earliest models were single equation models with tax revenue regressed on an income aggregate (Groves and Kahn, 1952). Singer (1968) extended this model by adding dummy variables to the regressors to account for changes in tax rates or the taxable income base.
(2) Singer (1970) developed a model with revenue-income elasticity as the dependent variable and included an independent variable to measure the effect of the tax rate structure on revenues.
(3) Multiple equation models with gross or taxable income in each income class regressed on aggregate income (Norman and Russell, 1970; McLaren, 1973) were designed to estimate income in each class directly rather than through the specification of a

TABLE 2[a]
Regression Estimates for Equation 3

Independent Variable	Equation Number 3-1	3-2
POP	1.448	1.445
	(21.85)	(22.72)
UN	-.0428	-.0417
	(4.71)	(4.83)
D1	.0331	
	(16.66)	
D2	.0572	
	(1.37)	
POP X D1		.139
		(17.82)
POP X D2		.0201
		(1.42)
Constant	-2.274	-2.272
	(13.06)	(13.56)
R^2	.991	.992
D-W	1.77	1.83
Standard Error	.0345	.0324

SOURCES: The sources for UN, D1, and D2 are given in Table 1. POP is collected from Minnesota Department of Health (1972) and unpublished data provided by the Section of Vital Statistics of the Department of Health.
a. As in Table 1, the positive values of the t-statistics are shown beneath the estimated coefficients, and R^2 is adjusted for degrees of freedom. The 5% significance level of the t-statistics is 1.77.

statistical distribution function. Revenues were then calculated as the sum of the products of the endogenous variables and tax rates.

(4) The final category is multiple equation models which include specification of a functional form to describe the distribution of income (Reaume, 1978). Equations are estimated for the

parameters of the distribution function and for the number of taxpayers as functions of macroeconomic variables.[11] Revenue is calculated using an identity like equation 1. Our model clearly fits in this category.

Our model is more complex and requires more data to generate forecasts than the models in the first three categories. On the other hand, our model is not as intricate as the Reaume model. In this section we compare the forecast accuracy of our model with the simpler models to see whether our model is sufficiently more accurate to justify its additional complexity. We also compare our forecast accuracy with Reaume's results. In addition, we analyze the sensitivity of our model's forecast to estimated values of Y3, Y50, and Y97.

COMPARISON WITH OTHER MODELS

Data for 1976 that became available after the model was estimated are used to test the forecast accuracy of the model. The values of the exogenous variables for these forecasts were as follows. For the reasons discussed above, the average income level in each of the income classes was set to the midpoint of the class. Because there were no tax changes between 1975 and 1976, effective tax rates for each income class were also set at their actual values for 1975. The exogenous variables for the estimated equations in Tables 1 and 2 were set at their actual 1976 values. The results of this ex post forecast for 1976 are reported in Table 3. The predicted values of the endogenous variables are all within two standard errors of forecast of the actual values. The tax revenue forecast for 1976 is low by $15.6 million or 1.6%.[12] In comparison, the absolute forecast error in North Carolina tax collections three quarters beyond the sample period using Reaume's model is 1.1% (Reaume, 1978: 87).

To compare our model with the simpler models, we estimated three forecasting models corresponding to the model categories discussed above.[13] Each was estimated with the same data used to

TABLE 3
Actual and Predicted Values for 1976

	Number of Taxpayers (Millions)	3rd Percentile Income	Median Income	97th Percentile Income	Income Tax Revenues (Millions)
Actual	1.803	$330	$7,996	$29,976	$974.2
Predicted	1.823	$346	$7,927	$31,044	$958.6

estimate our model for Minnesota. The "category one" model is a single equation of the form

$$R = a_o + a_1 Y + a_2 RC1 + a_3 RC2$$

where R is tax revenue, Y is total personal income, and RC1 and RC2 are dummy variables for tax rate changes over the sample period (Singer, 1968). The "category two" model is a single equation of the form

$$\epsilon = a_o + a_1 \left(\frac{R}{Y}\right)_{-1} + a_2 RC1 + a_3 RC2 + a_4 \left(\frac{GNP - GNP_{-1}}{GNP}\right)$$

where ϵ is the revenue income elasticity between years t-1 and t and GNP is Gross National Product for the United States (Singer, 1970). The "category three" model is a set of equations of the form

$$GY_i = a_o + a_1 Y$$

where GY_1 is the adjusted gross income in each income class (McLaren, 1973). The revenue forecasts for 1976 for these three models were $890.4 million (a forecast error of 8.6%), $1,047.7 million (7.5%), and $900.8 million (7.5%), respectively.

The results indicate that our model is substantially more accurate than the simpler models but less accurate than the

Reaume model. The dollar errors for the simpler models ranged from $73.4 to $83.8 million compared with the error of $15.6 million for our model. If we assume the 1.1% error shown for North Carolina to be applicable for Reaume's model if it were estimated for Minnesota, the resulting dollar error for 1976 would be $10.7 million. Thus, the relative decline in forecast accuracy between our model and Reaume's is much smaller than the relative decline from our model to the simpler models. These results indicate that the increased complexity and more extensive data requirements for models that incorporate the specification of a functional form for the size distribution may be worth the cost.

SENSITIVITY ANALYSIS

The usefulness of our model also depends on the sensitivity of its forecasts to estimates of the parameters of the income distribution. If forecasts are altered markedly by small changes in the estimates of the parameters of the distributions, then we must provide estimating equations for these parameters which have small standard errors of forecast.

Table 4 shows the revenues for 1976 predicted by the model when the quantile income levels are individually changed one standard deviation from the estimated values for 1976. For each row the values of the other income levels remain at their estimated values. The number of taxpayers was set to the predicted value for 1976.

As can be seen, the impact of changes in Y3 on revenues is nil whereas the effects of changes in Y50 and Y97 are approximately equal in percentage terms. The ratio of the percentage changes in revenues to the percentage changes in Y50 and Y97 are 0.7 and 0.8, respectively, which translate into predicted revenue changes of about $7 and $8 million per 1% change in Y50 and Y97. That the effects of changes in Y50 and Y97 are substantially larger than those for Y3 is not surprising because Minnesota has a highly progressive tax structure. Changes in Y3 materially affect the predicted number of taxpayers only in the lower part of the income distribution where effective tax rates are low and revenues col-

TABLE 4
Revenue Forecasts When Each Quantile Income Level is Changed by One Standard Deviation

Quantile Income Level	Predicted Value for 1976 (μ)	Standard Deviation (σ)	Revenue Forecast[a] for 1976 at μ + σ	Revenue Forecast[a] for 1976 at μ - σ
Y3	$346	$49	$958.6	$958.6
Y50	7,927	70	964.3	952.9
Y97	31,044	321	966.2	951.0

a. Revenues are in millions of dollars. Forecast revenue for 1976 using the predicted values of Y3, Y50, and Y97 was $958.6 million (see Table 3).

lected are small. If our model is to be used in a state with a highly progressive tax structure or one that is moderately progressive over a wide range of individual incomes, these results suggest that a substantial part of the effort in the implementation should be devoted to obtaining estimating equations for the parameters of the income distribution which have small standard errors of forecast.

CONCLUSION

In this paper we showed that an income tax revenue forecasting model could be designed to include a functional specification of the size distribution of income. The equations used to estimate the parameters of the distribution function were estimated with readily available data, and the estimated coefficients were mostly statistically significant with the expected signs. We also found that revenue forecasts were fairly sensitive to the estimated parameters of the income distribution function so that it is necessary to devote some effort to the specification of the estimated equations for these parameters in implementing this type of model.

The importance of including a functional specification of the size distribution was underscored by comparing forecast errors of

different models. We found that our model produced significantly lower forecast errors than models which do not incorporate the size distribution and that our forecast errors were only slightly larger than those obtained by Reaume in a much more complex model that also includes a functional specification of the size distribution. Our results, together with Reaume's, strongly suggest that including a specification of the size distribution in an income tax revenue forecasting model, while making specification and estimation more complex and requiring additional data, may well be worth the cost.

NOTES

1. Wilford (1975) found 39 legislated changes in income tax rates in 24 states over the 1959-1970 period. Minnesota, the state to which the revenue forecasting model developed in this paper is applied, legislated two rate changes between 1956 and 1976 and has other changes scheduled for 1978 and 1979.

2. We use the terms *revenue* and *liability* interchangeably because we do not separate collected revenue into its components (withholding, final payments, declarations, and refunds).

3. The differences in the two models are that our model is estimated using annual rather than quarterly data; final tax collections are estimated directly without separation into their components; and state tax rates, exemptions, and deductions, rather than entering directly, are used outside the model to compute effective tax rates on gross income.

4. We may estimate the proportion of taxpayers in each income class directly, or we may characterize the distribution of income using a particular functional form. For example, Norman and Russell (1970) and McLaren (1973) estimate linear equations of the form $Y_i = a_i + b_i Y$ for total income in each of the the income classes. However, these models do not account for the impact of new taxpayers on revenues. Additional problems with direct estimation are discussed by Metcalf (1972: 11-12).

5. References to studies of the income distribution and discussions of the underlying theoretical generating processes which lead to the use of lognormal and Pareto distribution functions appear in Lydall (1968) and Mincer (1970, 1974). In two other studies, Salem and Mount (1974) and Thurow (1970) use the Gamma and Beta distributions, respectively, as models of the income distribution.

6. Clearly, the selection of the "best" form must be guided by criteria which describe the desirable properties of the chosen distribution. We selected three criteria: (1) the chosen form had to closely approximate the data over time; (2) the number of parameters that were necessary to describe the distribution function had to be small, and (3) the parameters had to be easily predicted for tax revenue forecasting purposes. See Aitchison

and Brown (1957: 108-119) for a more extensive discussion of possible criteria for choosing the appropriate functional form of an income distribution in the context of the purposes of various analyses.

7. Several alternative forms of the income distribution function were also tested. The normal distribution function yielded poor results even in the lower percentiles of the distribution. The Pareto distribution provided an accurate fit above the sixtieth percentile but a poor fit below that point. A hybrid distribution, much like the "Standard Distribution" developed by Lydall (1968) in which the Pareto distribution was used to approximate the upper percentiles and the lognormal distribution the lower percentiles of size distribution, was equivalent to the displaced lognormal distribution in terms of the first criterion in note 6. However, this distribution has five parameters as opposed to three for the displaced lognormal distribution.

8. Aitchison and Brown (1957: 57-58) and Metcalf (1972: 19-23) derive the formulae relating the parameters of the displaced lognormal distribution to the quantile income levels or moments of the absolute data. To avoid iterative solutions for mu, sigma, and c, the quantile levels are chosen with the second quantile midway between the first and third (in standard deviation units). While the median is usually chosen as the second quantile level, it is not necessary to use the median as the point around which the symmetric quantiles are picked nor is it absolutely necessary that the first and third quantile levels be symmetric about the second. However, the goodness of fit is not improved by nonsymmetric choices (Aitchison and Brown, 1957: 53-54).

9. Estimates of personal income components and the unemployment rate are available from the Minnesota Department of Employment Services. They could also be obtained from an appropriately designed model of the Minnesota economy.

10. Income variables and a dummy variable for changes in the minimum filing requirement were included in a preliminary analysis to test the possibility that there were significant numbers of very low income individuals pushed into the minimum income level for filing a return in the data period. The coefficients of both the income variables and the dummy variable had the expected positive signs but were not statistically significant.

11. Two other analyses which incorporate a functional form of income distribution do not include forecasting equations for the distribution function parameters. Wellink (1974) uses the Pareto distribution to model the upper 80% of the Dutch income distribution and calculates an elasticity based on time trends in the parameters of this distribution. Spahn (1975) rejects the use of a statistical approximation to the income distribution and instead fits a spline function to the income distribution for West Germany. For forecasting, he assumes that all individual incomes increase at the rate of increase of total income, an assumption of dubious validity because it implies that the shape of the income distribution does not change over time.

12. Absolute forecast errors for 1974 and 1975 calculated using an earlier version of the model estimated from data through 1973 were 0.7% and 3.6%, respectively.

13. The estimated coefficients and other relevant statistics for these models are available from the author on request.

REFERENCES

AITCHISON, J. and J.A.C. BROWN (1957) The Lognormal Distribution. London: Cambridge Univ. Press.

GROVES, H. M. and C. H. KAHN (1952) "The stability of state and local tax yields." Amer. Economic Rev. 42 (March): 87-102.

LYDALL, H. (1968) The Structure of Earnings. London: Oxford Univ. Press.

McLAREN, J. A. (1973) "An income tax simulation model for the state of Minnesota." National Tax J. 26: 71-77.

METCALF, C. E. (1972) An Econometric Model of the Income Distribution. Chicago: Markam.

MINCER, J. (1970) "The distribution of labor incomes: a survey with special reference to the human capital approach." J. of Economic Literature 8 (March): 1-26.

——— (1974) Schooling, Experience, and Earnings. New York: Columbia Univ. Press for the National Bureau of Economic Research.

Minnesota Commissioner of Revenue (1977) The 1975 Minnesota Individual Income Tax. Income Tax Bulletin No. 43. St. Paul: Minnesota Commissioner of Revenue.

Minnesota Department of Health (1972) Minnesota Population: Trends, Estimates, Projections. Minneapolis: Minnesota Department of Health.

NORMAN, M. and R. R. RUSSELL (1970) "A personal income tax simulation model with an application to the state of Hawaii." National Tax J. 23: 429-434.

REAUME, D. M. (1978) "Forecasting state income tax collections: a new approach." Public Finance Q. 6: 67-91.

SALEM, A.B.Z. and T. D. MOUNT (1974) "A convenient descriptive model of the income distribution: the gamma density." Econometrica 42: 1115-1127.

SPAHN, P. B. (1975) "Simulating long term changes of income distribution within an income tax model for West Germany." Public Finance 30: 231-250.

SINGER, N. M. (1968) "The use of dummy variables in estimating the income elasticity of state income-tax revenues." National Tax J. 21: 200-204.

——— (1970) "Estimating state income-tax revenues: a new approach." Rev. of Economics and Statistics 52: 427-433.

SOLTOW, L. (1955) "The historic rise in the number of taxpayers in a state with constant tax law." National Tax J. 8: 365-381.

THUROW, L. C. (1970) "Analyzing the American income distribution." Amer. Economic Rev. 60: 261-269.

WELLINK, N. (1974) "Sensitivity of personal income tax revenue." National Tax J. 27: 357-360.

WILFORD, W. T. (1975) "A comment on 'The stability, growth, and stabilizing influence of state taxes.'" National Tax J. 29: 452-458.

William C. Weiler is Associate Director of the Management Information Division at the University of Minnesota. Mr. Weiler has also published work on enrollment demand and other policy issues related to higher education.