RECEIVED
SEP 24 6 49 PM '01
POSTAL RATE COMMISSION
OFFICE OF THE SECRETARY

LIBRARY REFERENCE USPS-LR-J-128
ESTIMATION OF LONG-RUN INCOME ELASTICITIES

Category Two Library Reference
(Thress, USPS-T-8)

Table of Contents

LIBRARY REFERENCE USPS-LR-J-128: ESTIMATION OF LONG-RUN INCOME ELASTICITIES

I. Introduction

The objective of this library reference is to calculate long-run income elasticities using data from the 1994 *Household Diary Study* as used in my testimony, USPS-T-7.

Estimating long-run income elasticities requires two stages of estimation. First, mail volume is analyzed and observed income is used to estimate income elasticities. Then a method is developed to relate observed income elasticities to long-run income elasticities. This is necessary because long-run income is not an observed variable. Having identified a relationship between income elasticity and long-run income elasticity, we calculate long-run income elasticities for every category of mail. The nature of the mail volume data, number of pieces of mail received in a period of time, leads to the use of count data models for the analysis.

This library reference is organized as follows. Section II provides an overview of the permanent income hypothesis and the process of calculating long-run income elasticities from measured income elasticities. Section III presents the calculation of current income elasticities from the Household Diary Study. Section IV describes the necessary steps to calculate P_y, the proportion of variance in observed income explained by permanent income, which is then used to transform the current income elasticities calculated in section III into long-run income elasticities.

II. Permanent Income Hypothesis: Theory and Other Issues

The ultimate objective of this study is to infer how changes in income affect the demand for mail services. According to the permanent income hypothesis, the answer depends on how the household sees the change in income. If this change affects the evaluation of the long-term income prospects, the response will be different than if the change in income is perceived as short-run, or transitory.

Friedman (1957) postulated the permanent income hypothesis, which states that consumption decisions are based on permanent income. Variations in income that do not alter permanent income do not affect consumption decisions. Friedman defined permanent income as an index of wealth, where wealth is to be regarded as including not only the nonhuman assets but also human capacities.

While the permanent income hypothesis was originated by Friedman when he was at the University of Chicago, the permanent income hypothesis has gained a lasting place in the economic literature with widespread, indeed almost universal, acceptance. The permanent income hypothesis is used widely in studies using both macroeconomic and cross-sectional data. This library reference focuses on the use of cross-sectional data. Previous studies considering the implications of permanent income in cross-sectional analysis include Hall and Mishkin (1982), Hayashi (1985), and Altonji and Siow (1987).

The basic idea in using the permanent income hypothesis in cross-sectional analysis is that observed income of households as reported in surveys may not truly capture the income that determines the demand for goods and services. The reported income, for instance, does not usually recognize assets that households have and that constitute part of their permanent income. Nor does reported income take into account characteristics of households, such as level of education, that could affect their long-term income.

The present study will first estimate income elasticities using the reported income in the *Household Diary Study*. Afterward, these elasticities will be transformed into long-run (or permanent) income elasticities using a procedure described in section IV below.

III. Mail Volume Analysis

A. Econometric Methodology

In cross-sectional studies of demand equations, it is usually assumed that all households face identical prices. Thus, explanations of behavioral differences are sought in differences in income. Equations that relate demand for goods and services to income for different households are called Engel curves.

Two issues should be noted concerning the estimation of the relation between income and mail volume using cross-sectional data. One is the choice of a functional form, the other is the choice of econometric methodology to perform the mail volume analysis.

The first issue concerns the choice of the functional form, specifically, how income will enter in the Engel curve. The Engel curve can be written as:

$$mv_i = G(y_i) \qquad (128.1)$$

where mv_i is mail volume and y_i is income for household i. A wide selection of functional forms for Engel curves has been explored in the literature. Most studies experiment with double logarithmic,

$$\ln(mv_i) = \alpha + \beta \ln(y_i) \qquad (128.2)$$

The income elasticity in equation (128.2) is given by β, so that this functional form imposes a constant income elasticity. In order to bypass this assumption, some studies have expanded equation (128.2) to include a quadratic term in log income,

$$\ln(mv_i)=\alpha+\beta\ln(y_i)+\gamma(\ln(y_i))^2 \qquad (128.3)$$

In this equation, the income elasticity is equal to $\beta + 2\gamma(\ln(y_i))$, so that equation (128.3) allows income elasticity to vary with the level of income. However, it also allows for the possibility of a constant income elasticity if γ is equal to zero. The estimation of the income elasticity will be carried out using this functional form.

The last issue is the econometric methodology to apply to the mail volume analysis. This question arises because the dependent variable takes only non-negative integer values, corresponding to the number of pieces of mail received in a given interval of time. Therefore regular econometric techniques cannot be applied to mail volume data. Cameron and Trivedi (1986), Mullahy (1986) and Greene (1994) are some of the studies dealing with the specification and estimation of count data models.

The idea is to discover a data generating process that fits the distribution of the data. As a starting point, a Poisson model will be fitted to the data. The expected value of mail volume is made a function of the explanatory variables as follows,

$$E(mv_i)=\exp\left[\alpha+\beta\ln(y_i)+\gamma(\ln(y_i))^2\right] \qquad (128.4)$$

However, this model is very restrictive because it implies that the mean of the process and the variance are the same. The Negative Binomial model is a modification of the

Poisson model that allows for over-dispersion of the data, meaning that the variance can be larger than the mean. In this model, the mean of the mail volume is specified as in the Poisson model. See appendix A for a complete description of these models and their estimation. In this analysis 3,698 observations will be usable.

B. Calculation of Income Elasticities

This section discusses how to calculate income elasticities from the different models considered in the previous section. These models are Poisson and Negative Binomial. The definition of expected mail volume for each model will be exploited in order to calculate income elasticities.

Income elasticity, η, is the percentage change in mail volume when income increases by one percent, i.e.,

$$\eta = \frac{\partial mv_i / mv_i}{\partial y_i / y_i} = \frac{\partial \ln(mv_i)}{\partial \ln(y_i)} \qquad (128.5)$$

The expected value of mail volume in the Poisson and Negative Binomial models is defined as in equation (128.4). Then it is possible to write the following,

$$\hat{mv}_i = \exp\left[\hat{\alpha} + \hat{\beta}\ln(y_i) + \hat{\gamma}(\ln(y_i))^2\right] \qquad (128.6)$$

Therefore the income elasticity will be,

$$\hat{\eta} = \hat{\beta} + 2\hat{\gamma}\ln(y_i) \qquad (128.7)$$

and the variance of the income elasticity will be given by,

$$Var(\hat{\eta}) = Var(\hat{\beta}) + (\ln(y_i))^2 Var(\hat{\gamma}) + 2\ln(y_i)Cov(\hat{\beta},\hat{\gamma}) \qquad (128.8)$$

Given the fact that income elasticity is a function of income, we need to find a unique value that represents the distribution of income elasticity. This value should be the integral over the distribution of income. This is,

$$\mathrm{H} = \int_{-\infty}^{+\infty} \hat{\eta}\ f(\ln(Y))\ d(\ln(Y)) \qquad (128.9)$$

where *f(.)* is the density function of the logarithm of income. This value turns out to be equal to,

$$\mathrm{H} = \hat{\beta} + 2\hat{\gamma}\int_{-\infty}^{+\infty} \ln(y)\ f(\ln(Y))\ d(\ln(Y)) = \hat{\beta} + 2\hat{\gamma}E[\ln(y)] \qquad (128.10)$$

The values reported as income elasticities are the values calculated using equation (128.10). Their variances are calculated using equation (128.8).

C. Results

In order to be able to estimate income elasticities from the Household Diary Study, the income reported in the Household Diary Study needs to be transformed. Income is reported by brackets. We assume that the natural logarithm of income is a normally distributed variable. With this assumption it is possible to estimate the mean and standard deviation of the process. With this information, the income ranges are substituted by the expected value of income inside that particular range.

The mean of the natural logarithm of income is estimated to be equal to 10.31 and the standard deviation is equal to 0.7331. With this mean and standard deviation the expected value of income for incomes less than $7,000 is equal to $5,319 and so on.

Appendix B includes the results for the Poisson and Negative Binomial models for all categories of mail for which long-run income elasticities are estimated from the Household Diary Study (nonpresort First-Class letters, First-Class cards, and Periodical mail). The Negative Binomial model fits the data better than the Poisson one for all categories of mail considered here. Consequently, the income elasticities are calculated with the coefficients estimated in the Negative Binomial models. The values of income elasticities for different values of income are included in appendix B.

Table 128-1 exhibits the income elasticities and their standard errors calculated at the mean of income.

TABLE 128-1: INCOME ELASTICITY AT MEANS

	Elasticity	Standard Error	T-Ratio
First Class			
Envelopes: Non-presort	0.4598	0.0216	21.30
Cards: Total	0.6375	0.0401	15.89
Second Class	0.4808	0.0341	14.09

IV. Transformation of Income Elasticities into Long-Run Income Elasticities

The current income elasticities calculated in section III above are transformed into long-run income elasticities by multiplying them by a value, P_y, which is equal to the percentage of variance of income contributed by long-term factors. The theoretical value of P_y is derived in section A. below. In section B. below, the value of P_y used in this case is derived.

A. Theoretical Derivation of P_y

Let the relationship of mail volume and long-run income be[1]:

$$\ln(mv_i) = \eta^p \ln(y_i^p) + u_i \qquad (128.11)$$

where mv_i is mail volume for household i, η^p is long-run income elasticity, y^p_i is long-run income for household i and u_i is a random error term which is not correlated with any other variable. In this case, the estimate of long-run income elasticity will be given by:

$$\hat{\eta}^p = \frac{\sum_{i=1}^{N} \ln(mv_i)\ln(y_i^p)}{\sum_{i=1}^{N} (\ln(y_i^p))^2} \qquad (128.12)$$

where N is the number of households in the sample. Equation (128.12) is obtained by applying Ordinary Least Squares (OLS) techniques to equation (128.11).

Now consider the estimate of income elasticity obtained by regressing the logarithm

[1] All the equations in this section are expressed in deviations with respect to the mean for simplicity of notation. For this reason there are no constant terms in these equations.

of mail volume on the logarithm of observed income:

$$\hat{\eta} = \frac{\sum_{i=1}^{N} \ln(mv_i)\ln(y_i)}{\sum_{i=1}^{N} (\ln(y_i))^2} \qquad (128.13)$$

where OLS is used to estimate income elasticity.

The logarithm of observed income is decomposed into long-run (permanent) and short-run (transitory) components as,

$$\ln(y_i) = \ln(y_i^{p}) + \ln(y_i^{t}) \qquad (128.14)$$

where the short-run component is uncorrelated with the long-run one. Therefore, this short-run component is not correlated with the logarithm of mail volume under the assumptions of equation (128.11). If we substitute equation (128.14) into equation (128.13), we observe that

$$\hat{\eta} = \frac{\sum_{i=1}^{N} \ln(mv_i)(\ln(y_i^{p}) + \ln(y_i^{t}))}{\sum_{i=1}^{N} (\ln(y_i))^2} \qquad (128.15)$$

Using the fact that the logarithm of short-run income is uncorrelated with the logarithm of mail volume, so that the sum of the logarithm of mail volume times the logarithm of short-run income, $\sum_{i=1}^{N} \ln(mv_i)\ln(y_i^{t})$, will be equal to zero, we obtain the following,

$$\hat{\eta} = \frac{\sum_{i=1}^{N} \ln(mv_i)\ln(y_i^p)}{\sum_{i=1}^{N} (\ln(y_i))^2} \tag{128.16}$$

Multiplying the numerator and denominator by the sum of the logarithm of long-run income squared, and rearranging equation (128.16) we find,

$$\hat{\eta} = \left(\frac{\sum_{i=1}^{N} \ln(mv_i)\ln(y_i^p)}{\sum_{i=1}^{N} (\ln(y_i^p))^2} \right) \left(\frac{\sum_{i=1}^{N} (\ln(y_i^p))^2}{\sum_{i=1}^{N} (\ln(y_i))^2} \right) \tag{128.17}$$

The first term of the multiplication is the estimate of the long-run income elasticity. Thus, equation (128.17) can be written as,

$$\hat{\eta} = \hat{\eta}^p \frac{\sum_{i=1}^{N} (\ln(y_i^p))^2}{\sum_{i=1}^{N} (\ln(y_i))^2} \tag{128.18}$$

or,

$$\hat{\eta} = \hat{\eta}^p \ P_y \tag{129.19}$$

where P_y is the proportion of the variance of observed income contributed by long-run income, i.e.:

$$P_y = \frac{\sum_{i=1}^{N} (\ln(y_i^p))^2}{\sum_{i=1}^{N} (\ln(y_i))^2} \qquad (128.20)$$

Notice that P_y is a ratio of variances because the variables are expressed in deviations with respect to their mean.

B. Derivation of P_y used in this Case

Estimating the value of P_y, which is needed to correct the elasticity coefficient from regressing mail volume on observed income, involves estimating an equation relating reported household income to various demographic characteristics of the households, in the following way,

$$\ln(y_i) = \alpha + \sum_{k=1}^{K} \beta_k X_{k,i} + \varepsilon_i \qquad (128.21)$$

where $X_{k,i}$ is a variable that represents a demographic characteristic, k, of household i, K is the total number of explanatory variables included in the equation and ε_i is the error term. There are 4,107 usable observations for the estimation of this equation.

The explanatory variables are: a constant, the number of children whose age is less than 18; dummy variables indicating the race of the head of the household; a dummy variable indicating whether the head of the household is married or not; dummy variables showing the age of the head of the household; dummy variables indicating the level of education attained by the head of the household; number of males and females in the household (older than 18) that are employed full- or part-time, and retired; and finally, the total number of males and females in the household over 18 years of age. The fitted values from this regression, $\ln(\hat{y}_i)$,will measure income that includes the long-run as well as the short-run components of income.

Data from the Household Diary Study consists of income ranges rather than actual values of income. This feature of the data constrains the econometric techniques that can be applied to equation (128.21). For example, it will be infeasible to estimate equation (128.21) by Ordinary Least Squares. Therefore maximum likelihood estimation techniques are implemented to estimate the parameters. Appendix C contains a detailed description of the methodology used in this regression.

The next step is to remove the transitory component of income, creating a variable, $\ln(\hat{y}_i^P)$. This is done by estimating expected values of the work status variables for each household. It is believed that using the expected values of the work status variables, instead of the actual ones, in predicting income will give a measure of income that contains only the long-run component.

The work status variables are the number of males and females that are full-time employed, part-time employed, retired, and the total number of males and females in the household. The total number of males and females in equation (128.21) picks up the effect of people that are not employed. The explanatory variables used in predicting work status are the remaining explanatory variables used in the equation (128.21). Count data models are applied to these regressions because of the nature of the data.

The final step is to estimate the extent to which variation in observed income is associated with variation in long-run income. The measure of the extent of this variation is an estimate of P_y. To carry out this step, a linear regression was run using

$\ln(\hat{y}_i)$ as the dependent variable and $\ln(\hat{y}_i^{\,p})$ as the independent one. This regression gives an R^2 of 0.8969. The value of R^2 is equal to the proportion of total variance of the dependent variable ($\ln(\hat{y}_i)$ in this case) that is explained by the fitted values of the regression ($\ln(\hat{y}_i^{\,p})$ in this case). Hence, R^2 is used as a proxy for P_y. All the results concerning these regressions are contained in appendix D.

C. Final Long-Run Income Elasticities

The value of P_y calculated above (0.8969) is used as shown in equation (128.19) to adjust the current income elasticities presented in Table 128-1 above. The final long-run income elasticities are displayed in Table 128-2.

Table 128-2: LONG-RUN INCOME ELASTICITY AT MEANS

	Elasticity	Standard Error	T-Ratio
First Class			
Envelopes: Non-presort	0.5127	0.0241	21.30
Cards: Total	0.7108	0.0447	15.89
Second Class	0.5361	0.0380	14.09

REFERENCES

ALTONJI, J. and A. SIOW (1987), 'Testing the Response of Consumption to Income Changes with (noisy) Panel Data,' *Quarterly Journal of Economics*, 102, 293-328.

CAMERON, A.C. and P.K. TRIVEDI (1986), 'Econometric Model Based on Count Data: Comparisons and Applications of Some Estimators and Tests,' *Journal of Applied Econometrics*, vol. 1, 29-53.

DEATON, A., (1992), *Understanding Consumption*, Oxford University Press.

GREENE, W.H., (1994), 'Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models,' Working Paper, Department of Economics, Stern School of Business, New York University, EC-94-10.

HALL, R.E. and F.S. MISHKIN (1982), 'The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households,' *Econometrica*, 50, 461-481.

HAYASHI, F., (1985), The Permanent Income Hypothesis and Consumption Durability: Analysis Based on Japanese Panel Data,' *Quarterly Journal of Economics*, 100, 1083-1113.

FRIEDMAN, M., (1957), *A Theory of the Consumption Function*, New York: National bureau of Economic Research, (distr. Princeton University Press).

MULLAHY, J., (1986), 'Specification and Testing of Some Modified Count Data Models,' *Journal of Econometrics*, vol. 33, 341-365.

RCF, (1994), 'Estimation of Permanent Income Elasticities for Mail Categories from the 1987 Household Diary Study Used in USPS-T-2,' Work Paper 1, R-94.

APPENDIX A: COUNT DATA MODELS

This appendix describes the models and the estimation procedure of the count data models applied to the mail volume analysis. For all model specifications it is assumed that there is a sample of *N* households (i=1,2,..., N). mv_i denotes the quantity of mail received by household *i*. The K x 1 vector of explanatory variables is denoted by X_i and *L* is the likelihood function.

A.1 - Poisson Model

The observed mail volume received, mv_i, is assumed to be an independent draw from a Poisson distribution with parameter λ_i (λ_i >0). The probability density of the Poisson random variable MV_i is then denoted by:

$$Pr(MV_i{=}mv_i)=\frac{e^{-\lambda_i}\lambda_i^{mv_i}}{mv_i!},mv_i{=}0,1,2,... \qquad (A.1)$$

The Poisson parameter is made a function of the explanatory variables as follows:

$$\lambda_i{=}\exp(X_i^{\prime}\beta), \qquad (A.2)$$

where β is the K x 1 vector of parameters. The exponential function prevents negative values for λ_i. The conditional mean and variance of MV_i are given by the Poisson parameter λ_i. This is a major drawback of the Poisson specification as actual data could be characterized by over- or under-dispersion (meaning that the variance could be greater or smaller than the mean).

The parameters in the model can be estimated by maximum likelihood procedures. The likelihood function of this model is:

$$L = \Pi_{i=1}^{N} Pr(MV_i = mv_i) = \Pi_{i=1}^{N} \frac{e^{-\lambda_i} \lambda_i^{mv_i}}{mv_i!} \qquad \text{(A.3)}$$

A.2 - Negative Binomial Model

The Negative Binomial model is an extension of the Poisson regression model which allows the variance of the process to differ from the mean. The probability density of the random variable is expressed by:

$$Pr(MV_i = mv_i) = \frac{\Gamma(v_i + mv_i)}{\Gamma(v_i)\Gamma(mv_i + 1)} \left[\frac{v_i}{v_i + \lambda_i}\right]^{v_i} \left[\frac{\lambda_i}{v_i + \lambda_i}\right]^{mv_i}, mv_i = 0,1,2,... \qquad \text{(A.4)}$$

where λ_i is defined as in the Poisson regression model and v_i is the index or precision parameter. v_i can be a function of the explanatory variables or simply a constant. It is assumed to be a constant and equal to $(1/\alpha)$ in this study. Now λ_i is the expected value of MV_i but the variance of MV_i is equal to:

$$V(MV_i) = \lambda_i + \alpha\lambda_i^2 \qquad \text{(A.5)}$$

When α is equal to zero the Negative Binomial model is reduced to the Poisson regression model. Therefore it is possible to test which of these two models fit the data better testing the significance of the α parameter. α will be always greater than zero, therefore this formulation allows for over-dispersion of the data.

Again this model can be estimated by maximum likelihood methods. The likelihood function will be:

$$= \Pi_{i-1}^{N} Pr(MV_i = mv_i) = \Pi_{i-1}^{N} \frac{\Gamma(1/\alpha + mv_i)}{\Gamma(1/\alpha)\Gamma(mv_i + 1)} \left[\frac{1}{1 + \alpha\lambda_i}\right]^{1/\alpha} \left[\frac{\alpha\lambda_i}{1 + \alpha\lambda_i}\right]^{m} \qquad \text{(A.6)}$$

APPENDIX B: EMPIRICAL RESULTS FOR THE MAIL VOLUME ANALYSIS

This appendix shows the empirical results for every category of mail included in the *Household Diary Study*. *Results for a Poisson model and a Negative Binomial model* are included in every table. Income elasticities for different values of income are calculated and shown at the end of the tables.

All these equations were estimated with the statistical package LIMDEP version 7.0.

Table 128-3: RESULTS FOR FIRST CLASS ENVELOPES _ NON-PRESORT

Poisson Regression	
Maximum Likelihood Estimates	
Dependent variable	ENVNPS1
Number of observations	3698
Iterations completed	6
Log likelihood function	-10655.65
Restricted log likelihood	-11368.72
Chi-squared	1426.147
Degrees of freedom	2
Significance level	0.0000000
Chi- squared =	14287.
G - squared =	11618.

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	MEANof X
LINC	-0.56206E-02	0.26337	-0.021	0.98297	10.31
LINC2	0.22492E-01	0.12663E-01	1.776	0.07571	106.9
Constant	-1.0803	1.3664	-0.791	0.42914	

Negative Binomial Regression	
Maximum Likelihood Estimates	
Dependent variable	ENVNPS1
Number of observations	3698
Iterations completed	8
Log likelihood function	-8757.982
Restricted log likelihood	-10655.65
Chi-squared	3795.327
Degrees of freedom	1
Significance level	0.0000000

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	MEANof X
LINC	-0.25933	0.47009	-0.552	0.58118	10.31
LINC2	0.34874E-01	0.23043E-01	1.513	0.13018	106.9
Constant	0.21276	2.3914	0.089	0.92911	
α	0.58396	0.18274E-01	31.955	0.00000	

Income	Elasticity	Standard Error	z=b/s.e.
$10,000	0.3831	0.4973E-01	7.702
$20,000	0.4314	0.2469E-01	17.47
MEAN	0.4598	0.2158E-01	21.30
$40,000	0.4798	0.2805E-01	17.11
$50,000	0.4953	0.3575E-01	13.86
$60,000	0.5080	0.4285E-01	11.86
$70,000	0.5188	0.4918E-01	10.55
$80,000	0.5281	0.5481E-01	9.634
$90,000	0.5363	0.5987E-01	8.958

Table 128-4: RESULTS FOR FIRST CLASS CARDS _ TOTAL

Poisson Regression	
Maximum Likelihood Estimates	
Dependent variable	CARDT1
Number of observations	3698
Iterations completed	7
Log likelihood function	-3369.496
Restricted log likelihood	-3547.241
Chi-squared	355.4891
Degrees of freedom	2
Significance level	0.0000000
Chi- squared = 5057.3	
G - squared = 4113.6	

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	MEANof X
LINC	-0.74347	0.75220	-0.988	0.32296	10.31
LINC2	0.66971E-01	0.35881E-01	1.866	0.06197	106.9
Constant	-0.34656	3.9340	-0.088	0.92980	

Negative Binomial Regression	
Maximum Likelihood Estimates	
Dependent variable	CARDT1
Number of observations	3698
Iterations completed	2
Log likelihood function	-3287.140
Restricted log likelihood	-3369.496
Chi-squared	164.7116
Degrees of freedom	1
Significance level	0.0000000

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	MEANof X
LINC	-0.74347	0.84536	-0.879	0.37914	10.31
LINC2	0.66974E-01	0.40956E-01	1.635	0.10199	106.9
Constant	-0.34656	4.3513	-0.080	0.93652	
α	0.66758	0.80984E-01	8.243	0.00000	

Income	Elasticity	Standard Error	z=b/s.e.
$10,000	0.4902	0.9850E-01	4.977
$20,000	0.5831	0.5206E-01	11.20
MEAN	0.6375	0.4011E-01	15.89
$40,000	0.6759	0.4653E-01	14.53
$50,000	0.7058	0.5798E-01	12.17
$60,000	0.7302	0.6953E-01	10.50
$70,000	0.7509	0.8018E-01	9.365
$80,000	0.7688	0.8981E-01	8.560
$90,000	0.7845	0.9854E-01	7.962

Table 128-5: RESULTS FOR SECOND CLASS

Poisson Regression	
Maximum Likelihood Estimates	
Dependent variable	TOTAL2
Number of observations	3698
Iterations completed	7
Log likelihood function	-7028.702
Restricted log likelihood	-7313.732
Chi-squared	570.0601
Degrees of freedom	2
Significance level	0.0000000
Chi- squared =	11223.
G - squared =	8920.6

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|>z]	MEANof X
LINC	0.34023	0.44006	0.773	0.43944	10.31
LINC2	0.67991E-02	0.21135E-01	0.322	0.74767	106.9
Constant	-3.9706	2.2858	-1.737	0.08238	

Negative Binomial Regression	
Maximum Likelihood Estimates	
Dependent variable	TOTAL2
Number of observations	3698
Iterations completed	4
Log likelihood function	-5847.402
Restricted log likelihood	-7028.702
Chi-squared	2362.599
Degrees of freedom	1
Significance level	0.0000000

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	MEANof X
LINC	0.33988	0.70557	0.482	0.63002	10.31
LINC2	0.68332E-02	0.34485E-01	0.198	0.84293	106.9
Constant	-3.9705	3.6007	-1.103	0.27016	
α	1.3426	0.57214E-01	23.466	0.00000	

Income	Elasticity	Standard Error	z=b/s.e.
$10,000	0.4657	0.7721E-01	6.032
$20,000	0.4752	0.3997E-01	11.89
MEAN	0.4808	0.3413E-01	14.09
$40,000	0.4847	0.4249E-01	11.41
$50,000	0.4877	0.5333E-01	9.145
$60,000	0.4902	0.6360E-01	7.708
$70,000	0.4923	0.7285E-01	6.758
$80,000	0.4942	0.8114E-01	6.090
$90,000	0.4958	0.8860E-01	5.596

APPENDIX C: INCOME ANALYSIS

As stated in section IV, the estimation of the parameters in equation (128.21) is constrained by the way income is reported by the *Household Diary Study*. This appendix describes how income is reported and how the parameters in equation (128.21) are estimated. For convenience, equation (128.21) is written here again:

$$\ln(y_i) = \alpha + \sum_{k=1}^{K} \beta_k X_{k,i} + \varepsilon_i \qquad (2.21)$$

Households report their income as being in one of twelve different income ranges. Thus we do not observe y_i but a variable called k_i. The variable, k_i, is described as follows:

$k_i = 1$ if $y_i <$ \$7,000
$k_i = 2$ if \$7,000 $\leq y_i <$ \$10,000
$k_i = 3$ if \$10,000 $\leq y_i <$ \$15,000
$k_i = 4$ if \$15,000 $\leq y_i <$ \$20,000
$k_i = 5$ if \$20,000 $\leq y_i <$ \$25,000
$k_i = 6$ if \$25,000 $\leq y_i <$ \$30,000
$k_i = 7$ if \$30,000 $\leq y_i <$ \$35,000
$k_i = 8$ if \$35,000 $\leq y_i <$ \$50,000
$k_i = 9$ if \$50,000 $\leq y_i <$ \$65,000
$k_i = 10$ if \$65,000 $\leq y_i <$ \$80,000
$k_i = 11$ if \$80,000 $\leq y_i <$ \$100,000
$k_i = 12$ if \$100,000 $\leq y_i$

In general, $k_i=j$ if $\alpha_j \leq y_i < \alpha_{j+1}$. Taking logarithms, $k_i=j$ if $\ln(\alpha_j) \leq \ln(y_i) < \ln(\alpha_{j+1})$.

Substituting equation (128.21),

$$k_i = j \text{ if } \ln(\alpha_j) \leq \alpha + \sum_{k=1}^{K} \beta_k X_{k,i} + \varepsilon_i < \ln(\alpha_{j+1}).$$

Subtracting $\alpha+\sum_{k=1}^{K}\beta_k X_{k,i}$ from every term,

$$k_i = j \text{ if } \ln(\alpha_j) - \alpha-\sum_{k=1}^{K}\beta_k X_{k,i} \le \varepsilon_i < \ln(\alpha_{j+1}) - \alpha-\sum_{k=1}^{K}\beta_k X_{k,i} \ .$$

Let *F(.)* be the distribution function of ε. Then the probability of k_i being equal to *j* can be written as:

$$Pr(k_i=j)=F\left(\ln(\alpha_{j+1})-\alpha-\sum_{k=1}^{K}\beta_k X_{k,i}\right)-F\left(\ln(\alpha_j)-\alpha-\sum_{k=1}^{K}\beta_k X_{k,i}\right) \quad \text{(C.1)}$$

where $ln(\alpha_1)$ is equal to $-\infty$ and $ln(\alpha_{13})$ is equal to $+\infty$. The likelihood function will be:

$$L=\Pi_{i=1}^{N}\Pi_{j=1}^{12}[Pr(k_i=j)]^{z_{ij}} \qquad \text{(C.2)}$$

where *N* is the number of households in the sample, Z_{ij} is equal to 1 if y_i falls in the j^{th} category and Z_{ij} equal to 0 otherwise.

The parameters can be estimated by maximizing the log-likelihood function:

$$\begin{aligned}\ln L=\Sigma_{i=1}^{N}\Sigma_{j=1}^{12}Z_{ij}\ln\Bigg[&F\left(\ln(\alpha_{j+1})-\alpha-\sum_{k=1}^{K}\beta_k X_{k,i}\right)\\ &-F\left(\ln(\alpha_j)-\alpha-\sum_{k=1}^{K}\beta_k X_{k,i}\right)\Bigg]\end{aligned} \qquad \text{(C.3)}$$

Assuming that ε_i follows a $N(0,\sigma^2)$ distribution, the log-likelihood function can be written as:

$$\ln L = \Sigma_{i=1}^{N} \Sigma_{j-1}^{12} Z_{ij} \ln \left[\Phi\left(\frac{\ln(\alpha_{j+1}) - \alpha - \sum_{k=1}^{K} \beta_k X_{k,i}}{\sigma} \right) - \Phi\left(\frac{\ln(\alpha_j) - \alpha - \sum_{k-1}^{K} \beta_k X_{k,i}}{\sigma} \right) \right] \tag{C.4}$$

where $\Phi(.)$ is the cumulative function of a normal distribution. Maximizing (C.4) with respect to α, β and σ requires the use of non linear optimization techniques. The estimation was performed using the statistical package LIMDEP version 7.0.

APPENDIX D: EMPIRICAL RESULTS FOR ANALYSIS OF INCOME

This appendix contains the empirical results that lead to the calculation of P_y, the proportion of the income variance due to variations in the permanent component of income. Table 128-6 shows the estimates of equation (128.21). Table 128-7 contains all the regressions of the work status variables. Notice that work status variables are the number of people employed full-time and part-time, retired people and the total number of people in the household. The last variable is picking up the effect of unemployed people in the household. The estimation of the number of retired people proved to be problematic. Therefore, an estimate of unemployed people was modeled instead, and the expected value of retired people was then backed out from this estimate. The estimation of all these variables was carried out with Poisson models because there is no evidence of over-dispersion of these data. Finally, Table 128-8 shows the OLS regression of income on its permanent component.

Here is a list and a description of the variables used in these regressions:

OLDINC, household income which is classified in 12 categories (see appendix C)

KIDS18, number of children whose age is less than 18

WHITE, dummy variable which is equal to 1 if the head of the household is white

BLACK, dummy variable which is equal to 1 if the head of the household is black

HISP, dummy variable which is equal to 1 if the head of the household is hispanic

ORIENT, dummy variable which is equal to 1 if the head of the household is oriental (the omitted variable reflects people that belong to some other race)

MARRIED, dummy variable which is equal to 1 if the head of the household is married

AGEDUMn, dummy variable which is equal to 1 if the age of the head of the household belongs to group n. The groups are as follows:
n=1, age between 18-21 (this group is the omitted one in the regressions)

n=2, age between 22-24
n=3, age between 25-34
n=4, age between 35-44
n=5, age between 45-54
n=6, age between 55-64
n=7, age between 65-69
n=8, age 70 and above

EDHDUMn, dummy variable equal to 1 if the level of education of the head of the household belongs to group n. The groups are as follows:
n=1, 8th grade or less (this group is the omitted one in the regressions)
n=2, high school incomplete
n=3, high school completed
n=4, some college
n=5, professional technical school
n=6, college graduate
n=7, postgraduate work

EFTM, number of males employed full-time in the household
EFTF, number of females employed full-time in the household
EPTM, number of males employed part-time in the household
EPTF, number of females employed part-time in the household

RETM, number of retired males in the household
RETF, number of retired females in the household

M, number of males in the household
F, number of females in the household

NEM, number of males that are not employed (M-EFTM-EPTM-RETM)
NEF, number of females that are not employed (F-EFTF-EPTF-RETF)

LINC, it corresponds to $\ln(\hat{y}_i)$ which is explained in the text

LPINC, it corresponds to $\ln(\hat{y}_i^p)$ which is explained in the text

Table 128-6: EMPIRICAL RESULTS FOR EQUATION (128.21)

```
Limited Dependent Variable Model - CENSORED
Maximum Likelihood Estimates
Dependent variable                 OLDINC
Number of observations             4107
Iterations completed                  3
Log likelihood function        -8145.713
Censoring Thresholds for the 12 cells:
Cell   Lower      Upper  Cell    Lower     Upper
 1  ********       8.9     2       8.9       9.2
 3       9.2       9.6     4       9.6       9.9
 5       9.9      10.1     6      10.1      10.3
 7      10.3      10.5     8      10.5      10.8
 9      10.8      11.1    10      11.1      11.3
11      11.3      11.5    12      11.5  ********
```

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|>z]	Mean of X
Constant	8.6834	0.12149	71.472	0.00000	
KIDS18	-0.52386E-02	0.78930E-02	-0.664	0.50688	0.8692
WHITE	0.11651	0.91550E-01	1.273	0.20316	0.8593
BLACK	-0.11228	0.95961E-01	-1.170	0.24196	0.6793E-01
HISP	-0.10713	0.97382E-01	-1.100	0.27130	0.5186E-01
ORIENT	-0.12826	0.11209	-1.144	0.25252	0.1388E-01
MARRIED	0.28855	0.20406E-01	14.140	0.00000	0.6713
AGEDUM2	0.16140	0.77971E-01	2.070	0.03846	0.3433E-01
AGEDUM3	0.38679	0.68607E-01	5.638	0.00000	0.2048
AGEDUM4	0.48276	0.68408E-01	7.057	0.00000	0.2520
AGEDUM5	0.56083	0.68763E-01	8.156	0.00000	0.1724
AGEDUM6	0.54978	0.69896E-01	7.866	0.00000	0.1278
AGEDUM7	0.44728	0.74277E-01	6.022	0.00000	0.7280E-01
AGEDUM8	0.36671	0.72928E-01	5.028	0.00000	0.1217
EDHDUM2	0.35071E-01	0.44140E-01	0.795	0.42688	0.1001
EDHDUM3	0.32431	0.39915E-01	8.125	0.00000	0.3414
EDHDUM4	0.53935	0.42162E-01	12.792	0.00000	0.1921
EDHDUM5	0.50026	0.50493E-01	9.908	0.00000	0.5381E-01
EDHDUM6	0.81112	0.43269E-01	18.746	0.00000	0.1600
EDHDUM7	0.95345	0.45419E-01	20.993	0.00000	0.1059
EFTM	0.49558	0.27637E-01	17.932	0.00000	0.6518
EFTF	0.29325	0.18338E-01	15.991	0.00000	0.4531
EPTM	0.15116	0.39041E-01	3.872	0.00011	0.6233E-01
EPTF	0.97122E-01	0.23707E-01	4.097	0.00004	0.1497
RETM	0.26021	0.36746E-01	7.081	0.00000	0.1483
RETF	0.14559E-01	0.27957E-01	0.521	0.60253	0.1704
M	-0.87027E-01	0.27300E-01	-3.188	0.00143	0.9535
F	-0.37336E-02	0.20107E-01	-0.186	0.85269	1.079
σ	0.48121	0.57284E-02	84.005	0.00000	

Table 128-7: EMPIRICAL RESULTS FOR WORK STATUS VARIABLE REGRESSIONS

(a) Number of males employed full-time

Poisson Regression	
Maximum Likelihood Estimates	
Dependent variable	EFTM
Number of observations	4107
Iterations completed	8
Log likelihood function	-3520.878
Restricted log likelihood	-4005.999
Chi-squared	970.2421
Degrees of freedom	19
Significance level	0.0000000
Chi- squared =	2455.9
G - squared =	2032.9

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	Mean of X
Constant	-1.4335	0.33490	-4.280	0.00002	
KIDS18	-0.31982E-01	0.18189E-01	-1.758	0.07869	0.8692
WHITE	0.19717	0.24399	0.808	0.41901	0.8593
BLACK	0.31806E-01	0.25814	0.123	0.90194	0.6793E-01
HISP	0.33811	0.25623	1.320	0.18699	0.5186E-01
ORIENT	0.51682	0.27667	1.868	0.06176	0.1388E-01
MARRIED	0.73766	0.53348E-01	13.827	0.00000	0.6713
AGEDUM2	0.28284	0.20392	1.387	0.16544	0.3433E-01
AGEDUM3	0.24068	0.18508	1.300	0.19347	0.2048
AGEDUM4	0.22513	0.18495	1.217	0.22351	0.2520
AGEDUM5	0.28661	0.18519	1.548	0.12169	0.1724
AGEDUM6	-0.11810	0.18969	-0.623	0.53354	0.1278
AGEDUM7	-1.4339	0.23667	-6.059	0.00000	0.7280E-01
AGEDUM8	-1.9601	0.24173	-8.109	0.00000	0.1217
EDHDUM2	0.15216	0.15267	0.997	0.31893	0.1001
EDHDUM3	0.27503	0.14135	1.946	0.05169	0.3414
EDHDUM4	0.25073	0.14456	1.734	0.08284	0.1921
EDHDUM5	0.31931	0.15998	1.996	0.04594	0.5381E-01
EDHDUM6	0.28831	0.14578	1.978	0.04796	0.1600
EDHDUM7	0.27329	0.15011	1.821	0.06866	0.1059

(b) Number of females employed full-time

Poisson Regression	
Maximum Likelihood Estimates	
Dependent variable	EFTF
Number of observations	4107
Iterations completed	8
Log likelihood function	-3205.904
Restricted log likelihood	-3435.224
Chi-squared	458.6381
Degrees of freedom	19
Significance level	0.0000000
Chi- squared =	3002.5
G - squared =	2886.2

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|>z]	Mean of X
Constant	-1.0814	0.38219	-2.830	0.00466	
KIDS18	-0.76415E-01	0.22886E-01	-3.339	0.00084	0.8692
WHITE	0.14962	0.29042	0.515	0.60642	0.8593
BLACK	0.31008	0.30192	1.027	0.30441	0.6793E-01
HISP	0.20244	0.30754	0.658	0.51038	0.5186E-01
ORIENT	0.47640	0.33057	1.441	0.14954	0.1388E-01
MARRIED	0.73207E-01	0.53626E-01	1.365	0.17221	0.6713
AGEDUM2	-0.61828E-01	0.22707	-0.272	0.78540	0.3433E-01
AGEDUM3	0.49437E-01	0.19696	0.251	0.80181	0.2048
AGEDUM4	0.18267	0.19608	0.932	0.35155	0.2520
AGEDUM5	0.25009	0.19639	1.273	0.20288	0.1724
AGEDUM6	-0.10204	0.20126	-0.507	0.61216	0.1278
AGEDUM7	-1.1029	0.23838	-4.627	0.00000	0.7280E-01
AGEDUM8	-2.1531	0.26886	-8.008	0.00000	0.1217
EDHDUM2	-0.10841E-01	0.17838	-0.061	0.95154	0.1001
EDHDUM3	0.21233	0.16280	1.304	0.19216	0.3414
EDHDUM4	0.23646	0.16660	1.419	0.15580	0.1921
EDHDUM5	0.24021	0.18616	1.290	0.19694	0.5381E-01
EDHDUM6	0.26170	0.16837	1.554	0.12012	0.1600
EDHDUM7	0.23749	0.17344	1.369	0.17089	0.1059

(c) Number of males employed part-time

Poisson Regression	
Maximum Likelihood Estimates	
Dependent variable	EPTM
Number of observations	4107
Iterations completed	9
Log likelihood function	-946.7605
Restricted log likelihood	-982.1240
Chi-squared	70.72699
Degrees of freedom	19
Significance level	0.0000000
Chi- squared = 4398.8	
G - squared = 1412.3	

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|>z]	Mean of X
Constant	0.25134E-01	0.59481	0.042	0.96629	
KIDS18	-0.23542	0.72793E-01	-3.234	0.00122	0.8692
WHITE	-1.3638	0.41860	-3.258	0.00112	0.8593
BLACK	-0.94622	0.46464	-2.036	0.04170	0.6793E-01
HISP	-1.4621	0.51562	-2.836	0.00457	0.5186E-01
ORIENT	-0.62727	0.55887	-1.122	0.26170	0.1388E-01
MARRIED	0.36991	0.14648	2.525	0.01156	0.6713
AGEDUM2	-0.54887	0.43595	-1.259	0.20803	0.3433E-01
AGEDUM3	-1.3300	0.38351	-3.468	0.00052	0.2048
AGEDUM4	-0.96856	0.36994	-2.618	0.00884	0.2520
AGEDUM5	-0.56931	0.36062	-1.579	0.11440	0.1724
AGEDUM6	-1.0098	0.37810	-2.671	0.00757	0.1278
AGEDUM7	-1.1760	0.41266	-2.850	0.00437	0.7280E-01
AGEDUM8	-1.6686	0.41462	-4.024	0.00006	0.1217
EDHDUM2	-0.48681	0.31523	-1.544	0.12251	0.1001
EDHDUM3	-0.65169	0.28238	-2.308	0.02101	0.3414
EDHDUM4	-0.66620	0.30160	-2.209	0.02718	0.1921
EDHDUM5	-0.51749	0.36951	-1.400	0.16137	0.5381E-01
EDHDUM6	-0.85620	0.31989	-2.677	0.00744	0.1600
EDHDUM7	-0.40047	0.31519	-1.271	0.20388	0.1059

(d) Number of females employed part-time

Poisson Regression	
Maximum Likelihood Estimates	
Dependent variable	EPTF
Number of observations	4107
Iterations completed	8
Log likelihood function	-1763.448
Restricted log likelihood	-1811.315
Chi-squared	95.73321
Degrees of freedom	19
Significance level	0.0000000
Chi- squared =	4052.3
G - squared =	2352.8

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	Mean of X
Constant	-1.9189	0.58876	-3.259	0.00112	
KIDS18	0.11816	0.35451E-01	3.333	0.00086	0.8692
WHITE	-0.86707E-01	0.45060	-0.192	0.84741	0.8593
BLACK	-0.29933E-01	0.47602	-0.063	0.94986	0.6793E-01
HISP	-0.29534E-01	0.48184	-0.061	0.95112	0.5186E-01
ORIENT	-0.55280E-01	0.54840	-0.101	0.91971	0.1388E-01
MARRIED	0.28805	0.98511E-01	2.924	0.00346	0.6713
AGEDUM2	-0.29852	0.35470	-0.842	0.40000	0.3433E-01
AGEDUM3	-0.53854	0.30600	-1.760	0.07842	0.2048
AGEDUM4	-0.40485	0.30381	-1.333	0.18267	0.2520
AGEDUM5	-0.30357	0.30541	-0.994	0.32024	0.1724
AGEDUM6	-0.51352	0.31562	-1.627	0.10374	0.1278
AGEDUM7	-0.49233	0.33515	-1.469	0.14184	0.7280E-01
AGEDUM8	-1.2193	0.34820	-3.502	0.00046	0.1217
EDHDUM2	-0.36206E-02	0.27838	-0.013	0.98962	0.1001
EDHDUM3	0.13573	0.25090	0.541	0.58854	0.3414
EDHDUM4	0.44739	0.25582	1.749	0.08031	0.1921
EDHDUM5	-0.27964E-01	0.31471	-0.089	0.92920	0.5381E-01
EDHDUM6	0.38298	0.26055	1.470	0.14158	0.1600
EDHDUM7	0.44502	0.26813	1.660	0.09697	0.1059

(e) Number of males

Poisson Regression	
Maximum Likelihood Estimates	
Dependent variable	M
Number of observations	4107
Iterations completed	6
Log likelihood function	-4297.256
Restricted log likelihood	-4515.567
Chi-squared	436.6232
Degrees of freedom	19
Significance level	0.0000000
Chi- squared =	1448.1
G - squared =	1535.8

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	Mean of X
Constant	-0.21789	0.23967	-0.909	0.36327	
KIDS18	-0.53086E-01	0.16257E-01	-3.265	0.00109	0.8692
WHITE	-0.24703	0.17293	-1.429	0.15314	0.8593
BLACK	-0.27040	0.18481	-1.463	0.14344	0.6793E-01
HISP	-0.86225E-01	0.18478	-0.467	0.64076	0.5186E-01
ORIENT	0.75891E-01	0.20657	0.367	0.71333	0.1388E-01
MARRIED	0.73171	0.42123E-01	17.371	0.00000	0.6713
AGEDUM2	0.14472	0.16668	0.868	0.38526	0.3433E-01
AGEDUM3	-0.48651E-01	0.15003	-0.324	0.74573	0.2048
AGEDUM4	0.89963E-02	0.14953	0.060	0.95203	0.2520
AGEDUM5	0.12779	0.14947	0.855	0.39257	0.1724
AGEDUM6	-0.80298E-02	0.15166	-0.053	0.95777	0.1278
AGEDUM7	-0.13231	0.15848	-0.835	0.40382	0.7280E-01
AGEDUM8	-0.25107	0.15510	-1.619	0.10549	0.1217
EDHDUM2	-0.52131E-01	0.92994E-01	-0.561	0.57508	0.1001
EDHDUM3	-0.88921E-01	0.84039E-01	-1.058	0.29001	0.3414
EDHDUM4	-0.92568E-01	0.88306E-01	-1.048	0.29452	0.1921
EDHDUM5	-0.38302E-01	0.10523	-0.364	0.71586	0.5381E-01
EDHDUM6	-0.10076	0.90149E-01	-1.118	0.26369	0.1600
EDHDUM7	-0.85732E-01	0.94174E-01	-0.910	0.36264	0.1059

(f) Number of females

Poisson Regression	
Maximum Likelihood Estimates	
Dependent variable	F
Number of observations	4107
Iterations completed	5
Log likelihood function	-4526.598
Restricted log likelihood	-4567.546
Chi-squared	81.89572
Degrees of freedom	19
Significance level	0.0000000
Chi- squared =	966.97
G - squared =	1082.1

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	Mean of X
Constant	0.17220	0.22835	0.754	0.45078	
KIDS18	0.44755E-01	0.14503E-01	3.086	0.00203	0.8692
WHITE	-0.61676E-01	0.18082	-0.341	0.73303	0.8593
BLACK	0.68490E-01	0.18844	0.363	0.71626	0.6793E-01
HISP	0.99112E-01	0.19040	0.521	0.60269	0.5186E-01
ORIENT	0.22548	0.21258	1.061	0.28883	0.1388E-01
MARRIED	0.14181	0.34532E-01	4.106	0.00004	0.6713
AGEDUM2	-0.22996	0.14777	-1.556	0.11965	0.3433E-01
AGEDUM3	-0.24329	0.12655	-1.922	0.05454	0.2048
AGEDUM4	-0.15447	0.12589	-1.227	0.21979	0.2520
AGEDUM5	-0.36320E-01	0.12626	-0.288	0.77360	0.1724
AGEDUM6	-0.27497E-01	0.12786	-0.215	0.82972	0.1278
AGEDUM7	-0.11502	0.13389	-0.859	0.39032	0.7280E-01
AGEDUM8	-0.16573	0.12977	-1.277	0.20157	0.1217
EDHDUM2	-0.20503E-01	0.82240E-01	-0.249	0.80312	0.1001
EDHDUM3	-0.50364E-01	0.74333E-01	-0.678	0.49806	0.3414
EDHDUM4	-0.47375E-01	0.78713E-01	-0.602	0.54726	0.1921
EDHDUM5	-0.12352	0.97624E-01	-1.265	0.20577	0.5381E-01
EDHDUM6	-0.95000E-01	0.81146E-01	-1.171	0.24171	0.1600
EDHDUM7	-0.13730	0.85998E-01	-1.597	0.11036	0.1059

(g) Number of males not employed

Poisson Regression	
Maximum Likelihood Estimates	
Dependent variable	NEM
Number of observations	4107
Iterations completed	8
Log likelihood function	-1227.647
Restricted log likelihood	-1284.444
Chi-squared	113.5944
Degrees of freedom	19
Significance level	0.0000000
Chi- squared = 4099.7	
G - squared = 1735.3	

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	Mean of X
Constant	-0.58500E-01	0.53818	-0.109	0.91344	
KIDS18	-0.12596E-02	0.48490E-01	-0.026	0.97928	0.8692
WHITE	-1.3376	0.36159	-3.699	0.00022	0.8593
BLACK	-1.0660	0.39636	-2.690	0.00716	0.6793E-01
HISP	-0.61958	0.38898	-1.593	0.11120	0.5186E-01
ORIENT	-0.64676	0.50186	-1.289	0.19750	0.1388E-01
MARRIED	0.74169E-01	0.11606	0.639	0.52278	0.6713
AGEDUM2	0.14108	0.40982	0.344	0.73065	0.3433E-01
AGEDUM3	-0.86609	0.38683	-2.239	0.02516	0.2048
AGEDUM4	-0.40046	0.37540	-1.067	0.28609	0.2520
AGEDUM5	-0.40335E-01	0.37311	-0.108	0.91391	0.1724
AGEDUM6	-0.22686	0.38125	-0.595	0.55182	0.1278
AGEDUM7	-0.47863	0.40624	-1.178	0.23873	0.7280E-01
AGEDUM8	-0.91842	0.40472	-2.269	0.02325	0.1217
EDHDUM2	-0.33300	0.21179	-1.572	0.11587	0.1001
EDHDUM3	-0.76101	0.19869	-3.830	0.00013	0.3414
EDHDUM4	-0.96551	0.22365	-4.317	0.00002	0.1921
EDHDUM5	-0.71888	0.28951	-2.483	0.01302	0.5381E-01
EDHDUM6	-0.97710	0.23552	-4.149	0.00003	0.1600
EDHDUM7	-1.2210	0.27282	-4.475	0.00001	0.1059

(h) Number of females not employed

Poisson Regression	
Maximum Likelihood Estimates	
Dependent variable	NEF
Number of observations	4107
Iterations completed	7
Log likelihood function	-2667.088
Restricted log likelihood	-2803.156
Chi-squared	272.1361
Degrees of freedom	19
Significance level	0.0000000
Chi- squared = 3287.4	
G - squared = 2944.5	

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	Mean of X
Constant	-0.24783	0.36587	-0.677	0.49818	
KIDS18	0.18868	0.22837E-01	8.262	0.00000	0.8692
WHITE	-0.40980	0.29140	-1.406	0.15963	0.8593
BLACK	-0.13859	0.30404	-0.456	0.64852	0.6793E-01
HISP	-0.54661E-01	0.30434	-0.180	0.85746	0.5186E-01
ORIENT	0.51165E-01	0.35685	0.143	0.88599	0.1388E-01
MARRIED	0.27138	0.67365E-01	4.029	0.00006	0.6713
AGEDUM2	-0.51644	0.23632	-2.185	0.02886	0.3433E-01
AGEDUM3	-0.58459	0.19807	-2.951	0.00316	0.2048
AGEDUM4	-0.61795	0.19753	-3.128	0.00176	0.2520
AGEDUM5	-0.46018	0.19943	-2.307	0.02103	0.1724
AGEDUM6	-0.39426	0.20265	-1.946	0.05171	0.1278
AGEDUM7	-0.68794	0.22264	-3.090	0.00200	0.7280E-01
AGEDUM8	-0.85764	0.21432	-4.002	0.00006	0.1217
EDHDUM2	-0.49315E-01	0.13142	-0.375	0.70748	0.1001
EDHDUM3	-0.31331	0.12234	-2.561	0.01044	0.3414
EDHDUM4	-0.52048	0.13409	-3.881	0.00010	0.1921
EDHDUM5	-0.57312	0.17720	-3.234	0.00122	0.5381E-01
EDHDUM6	-0.67461	0.14188	-4.755	0.00000	0.1600
EDHDUM7	-0.88037	0.16158	-5.449	0.00000	0.1059

Table 128-8: OLS RESULTS FOR INCOME

```
Ordinary    least squares regression    Weighting variable = ONE
Dependent variable is LINC      Mean =   10.31780, S.D. =      0.7224
Model size: Observations =     4107, Parameters =   2, Deg.Fr. =  4105
Residuals:  Sum of squares=    220.832     Std.Dev. =          0.23194
Fit:        R-squared = 0.89695, Adjusted R-squared =          0.89693
Model test: F[  1,    4105] = *******,    Prob value =         0.00000
Autocorrel: Durbin-Watson Statistic =   1.90171,   Rho =       0.04915
```

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]	Mean of X
Constant	-0.15255E-01	0.54784E-01	-0.278	0.78065	
LPINC	1.0015	0.52981E-02	189.028	0.00000	10.32

APPENDIX E: EMPIRICAL RESULTS FOR ANALYSIS OF INCOME

I.- EXPLANATION OF CONTENTS

This appendix describes the programs used to estimate the long-run income elasticities and standard errors. These programs have been filed with this library reference. These programs are written to be used with the statistical software LIMDEP version 7.0. These programs and data are the same as the ones used in my last testimony.

This appendix is divided into two sections. The first section describes the estimation and calculation of income elasticities and their standard errors. Section III of this library reference deals with this issue. The second section of this appendix will be devoted to explain the estimation of P_y, the percentage of variance of income contributed by long-run income, and the necessary previous steps to achieve this result. Section IV of this library reference explains the theoretical derivation of this ratio. With these two results and equation (128.19) above long-run income elasticities can be calculated.

Accompanying this document there is a diskette containing input (EST_Py.PRG and NBMAIL.PRG) and output (EST_Py.OUT and NBMAIL.OUT) files and the data (HDS94.DAT) used to estimate income elasticities and standard errors. The data file had to be compressed (DATA.EXE) and in order to decompress it all one has to do is type DATA at the dos prompt and the HDS94.DAT file will be created.

II.- ESTIMATION OF INCOME ELASTICITIES

A. NBMAIL.PRG FILE

```
? ***************************************************************
? *                                                             *
? * THIS PROGRAM ESTIMATES NEGATIVE BINOMIAL MODELS FOR *
? *    MAIL VOLUME AND CALCULATES THE ELASTICITIES.     *
? *                                                             *
? ***************************************************************
fast $
?
? LOAD THE DATA CONTAIN IN THE HDS94.DAT FILE AND
? CREATE AND OUTPUT FILE
?
LOAD; file = HDS94.DAT $
OPEN; Output = NBMAIL.OUT $
?
? CREATE LIST WITH THE EXPLANATORY VARIABLES
? FOR THE INCOME REGRESSION
?
NAMELIST; XV = one $
?
? DELETE OBSERVATIONS WITH MISSING VALUES FOR THE INCOME ANALYSIS
?
reject; OLDINC=-999$
reject; white=-999$
reject; agedum1=-999 $
reject; bank=-999$
reject; funds=-999$
reject; ownhome=-999$
reject; ira=-999$
? ***********************************************
?         INCOME REGRESSION
?  Notice that it is specified in log's
?  ln(7000)=8.8536654 and so on.
?  "linc" is the predicted value of log of income
? ***********************************************
GROUPEDDATA; lhs=OLDINC; rhs=xv
   ; LIMITS=8.8536654,9.2103404,9.6158055,9.9034876,10.126631,
10.308953,10.463103,10.819778,11.082143,11.289782,11.512925;keep=linc
$
?
? DELETE OBSERVATIONS WITH MISSING VALUES FOR THE MAIL VOLUME ANALYSIS
?
?
reject; center=-999$
reject; bank=-999$
reject; ira=-999$
reject; funds=-999$
```

```
reject; ccst=-999$
reject; ccgas=-999$
reject; ccbk=-999$
reject; ccot=-999$
reject; ownhome=-999$
reject; pc=-999$

? CREATE DIFFERENT TRANSFORMATIONS OF INCOME
?
create; linc2=linc^2 $
?
NAMELIST; delete xv $
NAMELIST; XV = linc, linc2, one  $
NAMELIST; MAILDATA=TOTAL1, ENVPKG1, ENVTOT1, ENVNPS1, ENVPS1, CARDT1,
                   CARDNPS1, CARDSPS1,TOTAL2, SINGLE3, TOTAL4, PP4,
BPM4,                         SR4, DOMESTIC $
?
? THESE VALUES ARE USED IN THE CALCULATIONS OF INCOME ELASTICITY
?
calculate; a1= 2*log(10000)$
calculate; a2= 2*log(20000)$
calculate; a3= 2*10.31$
calculate; a4= 2*log(40000)$
calculate; a5= 2*log(50000)$
calculate; a6= 2*log(60000)$
calculate; a7= 2*log(70000)$
calculate; a8= 2*log(80000)$
calculate; a9= 2*log(90000)$

PROC
  POISSON; lhs=\MAIL; rhs=xv; Model=N; Maxit=200 $
  matrix; vn=part(varb,1,2,1,2)
        ; bn=part(b,1,2)
        ; vinc= [1,a1/1,a2/1,a3/1,a4/1,a5/1,a6/1,a7/1,a8/1,a9]
        ; vae=vinc*vn*vinc'
        ; elast=vinc*bn
        ; dv=vecd(vae)
        ; sd=esqr(dv)
        ; isd=Diri(sd); ts=Dirp(elast,isd)
        ; list; r=[elast,sd,ts] $
ENDPROC
EXEC; \MAIL=MAILDATA $
close $
stop $
```

B. EXPLANATIONS ON NBMAIL.PRG

All the lines in the programs that begin with a question mark are comments. Therefore they do not execute any command, they are only explanations. We first load the data, which is accomplished in line 12. The command in line 13 creates an output file, NBMAIL.OUT, where all the results will be saved. Missing values in LIMDEP version 7.0 are represented by the value -999. Lines 22 through 28 remove observations for specific households with missing values in certain variables.

The main part of this program is the procedure described in lines 73 through 84 which estimates the Negative Binomial models for the different categories of mail volume and calculates the income elasticities and their standard errors. The independent variables in these regressions are the logarithm of income, the logarithm of income squared and a constant.

Income is reported by brackets, so we do not observed the actual value of income. OLDINC is a variable that only takes values 1,2,..,12 indicating to which bracket of income the household belongs. If OLDINC is equal to 1, the income of the household is between 0 and $7,000. If OLDINC is equal to 2, the income of the household is between $7,000 and $10,000 and so on. We need to express income in dollar terms to be able to use it in the mail volume regressions. This is accomplished by assuming that the logarithm of income is normally distributed. With the command GROUPEDDATA in lines 35 through 37, we are able to estimate the mean and standard error of the distribution of income. We are also able to calculate the expected value of income for every bracket of income. We substitute the income bracket defined by OLDINC by the expected value of income in that bracket which is the variable named *linc*. The command GROUPEDDATA estimates models where the dependent variable is reported only by category. This command requires the values of the thresholds and these are provided in LIMITS. Since the estimation is performed in logarithms, the limits are the logarithms of $7,000, $10,000 and so on. The only explanatory variable in this regression is a constant and the estimate of this coefficient is the mean of the distribution of the logarithm of income.

We create the square of the logarithm of income in line 54. And we defined the list of explanatory variables for the mail volume regression in line 57. When the Negative Binomial model is estimated for a mail volume category we save the coefficients, *bn* (line 76), and the covariance, *vn* (line75) of the two coefficients on income. Then we construct a matrix, *vinc* (line 77), with two columns. The first one is just a column with ones. The second column contains the logarithm for different values of income times two. This matrix, *vinc*, times the vector of coefficients, *bn*, will give us the value of the income elasticities, *elast* (line 79) for the different values of income defined in *vinc*. For example the third row of *elast* will be the elasticity of income evaluated at the mean of income. The last row of *elast* will be the elasticity of income for an income of $90,000. The formula we applied in line 79 is the equation (128.7) defined above.

The matrix *vinc* times *vn* time the transpose of *vinc* will give us a matrix, *vae*, which diagonal, *dv*, contains the variances of the income elasticities. Line 78 is equation (128.8) above. Line 83 prints the elasticities, *elast*, their standard errors, *sd*, and the t-ratios, *ts*.

Line 85 executes the procedure described in lines 73 through 84 for the variables in the MAILDATA list. This list is defined in lines 58 through 60. The definition of these mail volume varibles are as follows:

TOTAL1: First Class Total

ENVPKG1: First Class Envelopes and Packages

ENVTOT1: First Class Envelopes Total

ENVNPS1: First Class Envelopes Non-Presort

ENVPS1: First Class Envelopes Presort

CARDT1: First Class Cards Total

CARDNPS1: First Class Cards Non-Presort

CARDSPS1: First Class Cards Presort

TOTAL2: Second Class

SINGLE3: Third Class Single Piece

TOTAL4: Fourth Class Total

PP4: Fourth Class Total

BPM4: Fourth Class Bound Printed Matter

SR4: Fourth Class Special Rate

DOMESTIC: Domestic

III. ESTIMATION OF P_y

A. EST_Py.PRG FILE

```
? ****************************************
? *                                      *
? * THIS PROGRAM ESTIMATES Py            *
? *                                      *
? ****************************************
fast$
?
? LOAD THE DATA CONTAIN IN THE HDS94.DAT FILE AND
? CREATE AND OUTPUT FILE
?
LOAD; file = HDS94.DAT $
OPEN; Output = EST_Py.OUT $
?
? CREATE LIST WITH THE EXPLANATORY VARIABLES
?        FOR THE INCOME REGRESSION
?
NAMELIST; XV = one,kids18,white,black,hisp,orient,married,
                 agedum2,agedum3,agedum4,agedum5,agedum6,agedum7,agedum8,
                  edhdum2,edhdum3,edhdum4,edhdum5,edhdum6,edhdum7,
                  eftm,eftf,eptm,eptf,retm,retf,m,f $
?
? DELETE OBSERVATIONS WITH MISSING VALUES FOR THE INCOME ANALYSIS
?
reject; OLDINC=-999$
reject; kids18=-999$
reject; white=-999$
reject; married=-999 $
reject; agedum1=-999 $
reject; edhdum1=-999 $
? **********************************************
?         INCOME REGRESSION
?  Notice that it is specified in log's
?  ln(7000)=8.8536654 and so on.
?  "linc" is the predicted value of log of income
? **********************************************
GROUPEDDATA; lhs=OLDINC; rhs=xv
   ; LIMITS=8.8536654,9.2103404,9.6158055,9.9034876,10.126631,
     10.308953,10.463103,10.819778,11.082143,11.289782,11.512925
   ; keep=linc $
MATRIX; bethat1=[b(21)];bethat2=[b(22)];bethat3=[b(23)]
      ; bethat4=[b(24)]; bethat5=[b(25)]; bethat6=[b(26)]
      ; bethat7=[b(27)]; bethat8=[b(28)] $
NAMELIST; delete xv $
NAMELIST; XV = one,kids18,white,black,hisp,orient,married,
                 agedum2,agedum3,agedum4,agedum5,agedum6,agedum7,agedum8,
                  edhdum2,edhdum3,edhdum4,edhdum5,edhdum6,edhdum7 $
```

```
POISSON; lhs=eftm; rhs=xv;  Maxit=300; res=eeftm $
POISSON; lhs=eftf; rhs=xv;  Maxit=300; res=eeftf $
POISSON; lhs=eptm; rhs=xv;  Maxit=300; res=eeptm$
POISSON; lhs=eptf; rhs=xv;  Maxit=300; res=eeptf$
POISSON; lhs=m; rhs=xv;  Maxit=300; res=em$
POISSON; lhs=f; rhs=xv;  Maxit=300; res=ef$

create; nem=m-eftm-eptm-retm; nef=f-eftf-eptf-retf $
POISSON; lhs=nem; rhs=xv;  Maxit=300; res=enem$
POISSON; lhs=nef; rhs=xv;  Maxit=300; res=enef$
create; lpinc=linc-eeftm*bethat1-eeftf*bethat2
            -eeptm*bethat3-eeptf*bethat4-(em-eeftm-eeptm-enem)*bethat5
            -(ef-eeftf-eeptf-enef)*bethat6-em*bethat7-ef*bethat8 $
REGRESS; lhs=linc; rhs=one,lpinc $
close $
stop $
```

B. EXPLANATIONS ON EST_Py.PRG

We first load the data, which is accomplished in line 11. The command in line 12 creates an output file, EST_Py.OUT, where all the results will be saved.

The first step is to estimate equation (128.21) above. This equation regresses income on several characteristics of the households. The set of explanatory variables was explained in detail in Appendix B. Lines 17 through 20 create a list with all these explanatory variables. Lines 36 through 39 perform the estimation of equation (128.21). The dependent variable is OLDINC and it only takes values 1,2,..,12 indicating to which bracket of income the household belongs. The command GROUPEDDATA estimates models where the dependent variable is reported only by category. This command requires the values of the thresholds and these are provided in LIMITS. Since the estimation is performed in logarithms, the limits are the logarithms of $7,000, $10,000 and so on. We keep the predicted values of income under the variable named *linc*.

The next step is to get predictions for the work-status variables which are number of males (EFTM) and females (EFTF) employed full-time, number of males (EPTM) and females (EPTF) employed part-time, number of males (RETM) and females (RETF) that are retired and the total number of males (M) and females (F) in the household. The explanatory variables are all the ones we used in the income regression except for the work-status variables. The list of these variables is defined in lines 44 through 46. The work-status variables are count data and, since there is no evidence of over dispersion of these data, we applied Poisson regressions. Lines 47 through 52 are the estimation by Poisson regressions for number of males and females employed full-time and part-time and total number of males and females. As discussed above, we encountered some difficulties when estimating a regression for number of retired members of the household. Therefore we construct in line 53 the number of not employed people as the total number of males and females that are not working full-time, part-time or retired. And we run Poisson regressions for these two variables (lines 54 and 55). We keep the residuals for these eight Poisson regressions.

We construct next long-run income, which is defined as the component of the predicted income, *linc*, that contains only permanent components. Therefore long-run income, *lpinc*, is created as the predicted income minus the residuals for the work-status variables times the effects of each of those in the income regression. The residuals of the work-status variables reflect the short-run, or transitory, components of these variables. The long-run income is created in lines 56 through 58.

Last we run an OLS regression (line 59) of the predicted income, *linc*, on a constant and the long-run income, *lpinc*. The R-squared of this regression is the value for P_y, which is used to transform the income elasticities into long-run income elasticities.