A Tobit Analysis of the Demand for Life Insurance in Canada

by

Hongyan Fu

(2729104)

Major Paper presented to the

Department of Economics of the University of Ottawa

in partial fulfillment of the requirements of the M.A. Degree

Supervisor: Professor Kathleen Day

ECO 7997

Ottawa, Ontario

August 2004

Abstract

This paper investigates the demographic and economic factors determining the demand for life insurance in Canada using a Tobit model. Both the quantity of life insurance demanded and the probability of purchasing are examined in the empirical study. Some new factors such as country of birth and expenditure on tobacco products and games of chances are included in the estimating equation. Tests for heteroskedasticity are also carried out in the paper. The results show that income, marriage, age, educational level, number of earners, family size and home ownership have a significant positive relationship with the demand for life insurance. The marginal effects on the quantity demanded of life insurance, as well as the change in the probability of purchasing are also studied. Among all of the explanatory variables, the respondent's after-tax income has the highest elasticity.

1. Introduction

Life insurance has existed for a long time. It is used by people to share an unknown risk through paying a small certain premium. Since individuals may die during their productive years, or live a longer life than expected, a loss can result from premature or overdue death. Life insurance is used by people to eliminate these risks and maintain their lives as expected over their entire life span. As the standard of living rises, the purchase of life insurance becomes an increasingly important decision for each family. Furthermore, life insurance is becoming a key source of profit for financial services companies. According to data from the Canadian Life and Health Insurance Association (CLHIA), by the end of 2002, the total amount of life insurance owned by Canadians reached $2369.3 billion. Almost 17.5 million Canadians owned life insurance, and the average amount owned by insured individuals was $136,400. Although 2002 was a challenging year, during this year, Canadians purchased $246.6 billion of life insurance and about 871,100 individual policies were purchased. The average size of new individual policies was $188,100. Hence, research on the purchasing decision and the demand for life insurance will continue to be a hot topic.

Theoretically speaking, individuals demand life insurance in order to compensate for the financial loss caused by uncertain risks. Based on theoretical models of this demand, many empirical studies have been done, yielding many useful results regarding the demographic and economic factors that influence the demand for life insurance. However, these studies mainly focused on some specific populations, such as baby boomers and young married families, or a limited number of explanatory variables. Moreover, most of the studies only examined the relationship between the amount of life insurance demanded and the explanatory variables, but neglected their effect on the probability of purchasing insurance.

In this paper, the Tobit regression model is employed to investigate the demand for life insurance using cross-section data for Canada. Besides the general explanatory factors examined by prior studies, such as income, age and occupation, some new explanatory factors are included in this empirical model. Because Canada is a country of immigrants, many people come from different countries and regions. These differences in culture will probably influence life insurance demand, so country of birth is used as an explanatory factor in the model to indicate whether individuals are immigrants or native-born Canadians.

As we know, people need life insurance to avoid the financial loss associated with uncertain risk, but some of them do not have a good understanding of risk and do not take action to purchase insurance. Therefore attitude towards risk in real life is an important factor to consider. However, it is a little complex to measure this variable since it belongs to the field of psychology. This study tries to find the impact of this element on the demand for life insurance by using expenditure on games of chance as an explanatory variable.

Section 2 of this paper briefly reviews the literature relating to theoretical and empirical studies on life insurance demand and purchasing behavior, while section 3 describes the Tobit model and the calculation of marginal effects, which are used in the empirical analysis. The data set and explanatory variables included in the empirical model are discussed in section 4. The regression results are analyzed in section 5. Finally, the findings and suggestions for future research are summarized in a brief conclusion.

2. Literature Review

Because this study constitutes an empirical study of the demand for life insurance in Canada, a number of relevant empirical models and methods are reviewed in this section. Since the theoretical research is the basis of the empirical studies, first some theoretical studies are briefly examined. For a more complete review of theoretical and empirical models, see Zietz (2003).

Life insurance has existed since Roman times. It is based on the idea of a large number of people contributing a small amount of money (a premium) into a fund or pool, to cover the needs of those people who suffer a loss. Because nobody knows when he might be the one in need, by paying a small certain cost which can be budgeted, the individual can shift to the insurer a potentially large uncertain loss that cannot be budgeted.

If a major source of income for a family is eliminated due to death, disability, or other impairment of one or more family members, it will be necessary for the family to make economic and social adjustments. The human life value concept was developed to measure the economic value of life. With respect to human life value, life insurance has two functions: to contribute toward its conservation and to protect against financial loses resulting from its destruction. Individuals may die during their productive years, or live far beyond them. Loss can result if death is premature or overdue. In first case, the income earner cannot realize his total value and ceases to bring in cash for his family. However in the latter case, his retirement plan or legacy may need to be changed, also resulting in losses, because of his unexpectedly long life. If a business interest or a large estate is involved, the owner's death can result in loss regardless of when death occurs. Therefore, life risk is uncertainty concerning the time of death. If people knew exactly when death would occur, no risk would exist even though death could cause losses. The purpose of life insurance is to eliminate these risks for individuals and to reduce them in the economy.

The principal results in the theoretical literature on the life insurance demand are derived by Yaari (1965), Fischer (1973), Campbell (1980) and Lewis (1989). Yaari (1965) provides the first theoretical analysis of the topic and derived a mathematical model to explain the consumer's behavior when faced with an uncertain lifetime. However, he does not examine demand functions for consumption or insurance in detail. Fischer (1973) uses a discrete-time model and compared statics and dynamics of the insurance demand functions. Campbell (1980) analyzes the household's optimal reactions to income uncertainty because of the wage earner's premature death and finds that a person's aggregate future earnings have a moderate positive effect on the total insurance held on that person, while aggregated social security survivor benefits have a moderate negative effect. He calculates social security benefits based on earnings, age, and family size.

All of these models assume that the possible premature death of a household's primary wage earner will cause uncertainty with respect to the households' income, and view life insurance as the way to reduce this uncertainty. Lewis (1989) extends Yaari's model by explicitly modeling the demand for insurance of the primary wage earner's dependents. In other words, in his model the optimal amount of insurance is determined by maximizing the utility of the beneficiaries (i.e., the children or spouse). He argues that this approach makes sense even though young children are unlikely to actually purchase insurance, because the life insurance premium is paid by the parents on behalf of their offspring.

Based on these theoretical models, many empirical studies have made an effort to identify all of the factors that affect people's demand and purchase behavior. Over the past 40 years, many researchers have tried different methods to find the significant elements that influence people's life insurance purchasing behavior, including demographic and economic factors. Most of them use income level, age, family size, education and occupation as major explanatory variables. As for the dependent variable, previous studies mostly use premium expenditures, partly because it is available in most datasets. The amount of purchased life insurance and the type of life insurance (with cash value or without cash value) are also studied as dependent variables, for example by Anderson and Nevin (1975). However, because they focus on different groups and used different methods, their empirical results are very different.

Hammond, Houston, and Melander (1967) were the first to study the factors associated with life insurance consumption through a multiple linear regression model. Cross-sectional data from surveys in 1952 and 1961 are used in the research. In their empirical model, the life insurance premium is a function of income, net worth, race, age, education and the occupation of the family head. The results show that income, net worth, household life cycle stage, education and occupation significantly influence life insurance purchases. Among these variables, income and net worth are positively related to life insurance purchases. In addition, the coefficient of the occupation variable implies that households in professional, self-employed and managerial occupations spend more on premiums than those in other occupational categories. Among the other variables, life stage and unmarried status are negatively related to insurance purchases. The coefficient of age is significant only in the low and middle-income groups, but it has a different sign for each group. The coefficient of race is not significant for any of the income groups.

Hammond, Houston, and Melander (1967) also estimate the income elasticity of demand for life insurance. The result shows that income changes have an inelastic effect on expenditure on life insurance in the low- and high-income groups, but in the middle-income group, the elasticity is high, which means expenditure on life insurance is highly responsive to the income changes in this group. The reason is that with a middle-income level, households need proportionately less income to buy necessities, while the low-income households must spend most of their increased income on current consumption. Therefore, middle-income households have a greater capability of increased expenditures on life insurance. However, for individuals with a high income, substitutes for individual life insurance such as stock options and pensions are available, which reduces the need for life insurance.

Berekson (1972) extends the scope of empirical analyses of insurance purchasing behavior by investigating the determinants of both the amount of insurance and premium expenditures. The amount of insurance is the face death benefit for the issued person. However, the premium expenditure is the cost of purchasing the life insurance, i.e., the regular payment to the insurance companies. Therefore, the amount of life insurance is a much more accurate measure of the quantity of life insurance demanded than the premium expenditure. The explanatory variables include demographic and socio-economic variables. Using two data sets from surveys of college students, he finds that the coefficient of age is positively significant; however birth order is both positively and negatively related to life insurance demand depending on the data set. The number of children and income are also associated with life insurance purchasing behavior.

In subsequent research, Anderson and Nevin (1975) extend the dependent variable again by examining two variables: the amount of life insurance and the type of life insurance. However, just a special group, the young married family, is examined. The data come from the Panel on Consumer Decision Processes and are composed of 230 couples. They use a technique known as Multiple Classification Analysis (MCA) to explain the relationship between the dependent and independent variables. The results show that the coefficients of the level of education of the husband, the husband's current income, expected household income, and the net worth of the household are statistically significant in explaining the *amount* of life insurance purchased. The education of the husband is negatively related to the amount of life insurance purchased. Insurance on the wife before marriage is positively related to life insurance purchased; however, insurance on the husband before marriage is negatively related to it. The coefficient of net worth, the wife's insurance portfolio before marriage and the influence of the insurance agent were found to be statistically significant in explaining the *type* of life insurance purchased. Income is an obvious explanatory variable that is positively related to the amount of insurance. However, an unexpected finding is that middle-income couples purchased considerably less life insurance than did the lower- or upper-income couples, which means a non-linear model is better for explaining the relationship between income and insurance. The explanation proposed by Anderson and Nevin (1975) is that insurance companies neglect the market for middle-income households, since salespeople always exert more effort on the higher income households in order to get a higher commission. Besides, advertisements and other sales efforts work better for the lower-income households than for the middle-income households.

Browne and Kim (1993) extend the literature by examining international differences in life insurance demand. Their sample includes 45 underdeveloped and developed countries spread throughout the world. Both the life insurance premiums and life insurance in force are used as the dependent variable to measure insurance consumption. In the demand model, they take the mean national values of these dependent variables as representative of a typical household in that country and use ordinary least squares to estimate a log linear equation. Their empirical results show that the dependency ratio (the ratio of the total number of children under 15 to the total number of persons aged 15 to 64), national income, and government spending on social security are significant determinants and have a positive effect on life insurance demand. On the other hand, inflation, the price of insurance and religion are significantly negatively related to life insurance demand. The analysis is based on macro data and cannot accurately estimate the effect of household characteristics. Since the model used national income to represent household income, it may not estimate the effect of individual characteristics on life insurance demand very well.

All of the above studies use cross-section data. In contrast, Truett and Truett (1990) estimate a linear regression model with time series data covering the US from 1960 to 1982 and Mexico from 1964 to 1979. The quantity of life insurance is used as the dependent variable. The results show that age, education and the level of income are significant determinants of insurance demand. One of their main contributions is that they find that the income elasticity of demand is much higher in Mexico than in the U.S. This finding provides empirical evidence for the hypothesis that the income elasticity of demand for life insurance is higher in less developed countries than in developed countries.

Chen, Wong, and Lee (2001) is a recent study which uses time series data to examine trends in the life insurance purchase rate (the ratio of the number of purchasers to the reference group) in the U.S. from 1940 to 1996. The data are obtained from two main sources. One is the *Life Insurance Fact Book* published by the American Council of Life Insurance, and the other is *The Buyer Study: United States.* They question the causal modeling approach for they think that the standard model does not separate the age, period and cohort effects. Using a cohort regression model proposed by Mason et al. (1973) to examine age, period and cohort effects on life insurance purchasing in the U.S., they find that people born in the baby boom period purchase less life insurance than consumers of earlier generations, which has resulted in a decrease in life insurance purchases in the U.S.

Bernstein and Geehan (1988) analyze the household demand for insurance services in Canada using both cross-section data for 1982 and time series data from 1952 to 1982 provided by Statistics Canada. Property, auto and life insurance are analyzed as the dependent variables. Besides income, household characteristics such as family size, age, region, urban or rural, sex and tenure are used to estimate Engel functions. Their results show that insurance services are luxury goods because the income elasticity is greater than one. In addition, female household heads spend less on life insurance than males and tenant households spend less on life insurance than homeowners. The results also show that the demand for life insurance is significantly higher in Quebec than in other provinces.

Another empirical study for Canada was done by Matteo and Emery (2002). They studied the relationship between wealth and the demand for life insurance using a dataset for Ontario in 1892. The dependent variable is the amount of insurance payments. Wealth, age, number of children, marriage, urban, residency and race are studied in the Probit and OLS models. Their primary result is that there is a negative correlation between wealth and the demand for life insurance. That is because with wealth accumulation, self-insurance is available and becomes a substitute for life insurance.

Prior research has typically used the linear regression model to try to find the factors that affect the quantity of life insurance demanded. However, because in reality many people do not purchase life insurance, in samples of microdata corner solutions of zero insurance purchases will be observed. In this case using ordinary least squares estimation will result in biased and inconsistent estimates. The Tobit model proposed by Tobin (1958) is more appropriate in this case because it explicitly takes into account the zero observations. Showers and Shotick (1994) analyze the effects of household characteristics on the demand for total insurance using the Tobit regression model. The data for this study came from the interview portion of the U.S. Consumer Expenditure Survey (CES) in 1987. The Tobit regression model is employed to estimate both the marginal change in the demand for insurance and the change in the probability of purchasing insurance. The regression results show that significant determinants of the demand for total insurance include the number of household earners, income, age, and family size, all of which are positively related to demand. In addition, they examine the marginal effects of each determinant. They find that the family will buy more insurance as income increases because they need more protection for their high level of quality of life. Unfortunately, their study is focused on total insurance, an aggregate of health, life, auto and homeowners insurance, and it is not clear if the results apply to life insurance alone. Furthermore, just a few explanatory variables are examined.

Hau (2000) also uses the Tobit model to analyze life insurance demand. However, his study is focused on a special group, retired people. The paper found that demographic and personal variables appear to be less significant than financial factors in determining life insurance for the retired people.

Until now, to the author's best knowledge, there is no empirical study using the Tobit method that analyzes all the demographic and economic factors determining insurance demand and the probability of purchasing insurance. In addition, there are few empirical studies for Canada. Thus this paper will analyze cross-section data in

Canada and contribute to the empirical literature on the demand for life insurance.

3. Econometric model

The censored regression model proposed by Tobin (1958) can be expressed by the following equations (1) and (2):

$$Y_t^* = X_t\beta + \mu_t, \tag{1}$$

$$Y_t = \begin{cases} Y_t^*, \text{ if } Y_t^* > 0 \\ 0, \text{ if } Y_t^* \leq 0, \end{cases} \tag{2}$$

where t is from 1 to N, N is the number of observations, Y_t is the observed dependent variable, X_t is a row vector of independent variables, β is a vector of unknown coefficients, and μ_t is an independently normally distributed error term. In equation (1), Y_t^* is also called a latent variable, since it is observed only when it is positive and hence is an underlying index for Y_t.

The parameters of (1) can be estimated using maximum likelihood estimation. The log-likelihood function for the Tobit model can be expressed as:

$$L = \sum_{Y_t>0} -\tfrac{1}{2}[\ln(2\pi) + \ln(\sigma^2) + (\tfrac{Y_t - X_t\beta}{\sigma})^2] + \sum_{Y_t=0} \ln[1 - F(\tfrac{X_t\beta}{\sigma})], \tag{3}$$

where σ is the standard error of μ_t, and $F(z)$ is the cumulative normal distribution function.

In the Tobit model, the expected value of Y_t is given by equations (4) and (5):

$$z_t = \frac{X_t\beta}{\sigma}, \tag{4}$$

$$E(Y_t) = X_t\beta F(z_t) + \sigma f(z_t). \tag{5}$$

where σ is the standard error of μ_t, $f(z_t)$ is the normal probability density and $F(z_t)$ is the cumulative normal distribution function.

The Tobit model coefficients can be used to examine changes in both the probability of being above the limit (zero) and changes in the value of the dependent variable when it is already above the limit. The decomposition can be achieved using the method proposed by McDonald and Moffitt (1980). The change in the probability that Y_t will be above the limit and the change in the level of Y_t when the i^{th} variable in X_t changes is given by (7) and (8), respectively:[1]

$$\lambda = \frac{f(z_t)}{F(z_t)}, \tag{6}$$

$$\frac{\partial F(z_t)}{\partial X_{it}} = \frac{f(z)\beta_i}{\sigma}, \tag{7}$$

$$\frac{\partial E(Y_t \mid Y_t > 0)}{\partial X_{it}} = \beta_i(1 - \lambda_0 - \lambda_0^{\,2}). \tag{8}$$

McDonald and Moffitt (1980) show that the total marginal effect of the change in X_{it} on Y_i can be written as the following function of equations (7) and (8):

$$\frac{\partial E(Y_t)}{\partial X_{it}} = F(z_t)\frac{\partial E(Y_t \mid Y_t > 0)}{\partial X_{it}} + E(Y_t \mid Y_t > 0)\frac{\partial F(z_t)}{\partial X_{it}} \tag{9}$$

Sometimes, the elasticity gives a more comparable marginal change when examining the effect on the dependent variable due to a change in the independent variables. The total elasticity with respect to a change in X_{it} can be expressed as the following function:

$$\eta_i = \frac{\partial E(Y)}{\partial X_{it}} * \frac{X_{it}}{E(Y_t)}, \tag{10}$$

and in practice can be evaluated at the sample mean values of the explanatory variables. The total elasticity can be divided into the elasticity of the purchasing probability (*EPB*) and the elasticity of the amount of insurance demanded (*EID*). Using equation (9), the decomposition can be calculated in the following manner:

[1] X_{it}: i indexes variables and t indexes observations.

$$EPD = E(Y_t \mid Y_t > 0) * \frac{\partial F(z)}{\partial X_{it}} * \frac{X_{it}}{E(Y_t)}, \tag{11}$$

$$EID = F(z_t) * \frac{\partial E(Y_t \mid Y_t > 0)}{\partial X_{it}} * \frac{X_{it}}{E(Y_t)}. \tag{12}$$

Heteroscedasticity is a common issue when analyzing microeconomic data with the Tobit model. Two approaches to testing for the problem are applied in this paper: one is a likelihood ratio (LR) test against the null hypothesis of homoscedasticity and the other is a Lagrange Multiplier test (LM).

The LR test is quite direct. It requires estimating the Tobit model both with and without heteroskedasticity. The log-likelihood function in the former case is similar to equation (3), except that σ is replaced by σ_t. For both tests, the variance is assumed to be of the form:

$$\sigma_t^2 = \sigma^2 e^{\alpha' w_t} \tag{13}$$

where the existence of heteroscedasticity on the estimates is tested against the null hypothesis that $\alpha = 0$ based on the likelihood ratio statistic. The LM test is carried out without having to estimate the unrestricted model. The formulas for the LM statistic and its components can be found in Greene (1997, 968-969). To derive the LM statistic, one first needs to differentiate the log-likelihood function for the model with heteroskedasticity with respect to all its parameters, and then substitute in the values of those parameters under the null hypothesis of homoskedasticity. In such a way, the following expressions for the derivatives:

$$\ln L_\beta = \sum_{t=1}^{n} \partial_t X_t, \quad \ln L_\sigma^2 = \sum_{t=1}^{n} b_t, \quad \ln L_\alpha = \sum_{t=1}^{n} \sigma^2 b_t w_t \tag{14}$$

where

$$z_t = \begin{cases} 1, if\, Y_t > 0 \\ 0, if\, Y_t = 0, \end{cases} \tag{15}$$

$$\alpha_t = z_t(\tfrac{\varepsilon_t}{\sigma^2}) + (1 - z_t)(\tfrac{-\lambda_t}{\sigma}), \tag{16}$$

$$b_t = z_t(\tfrac{\varepsilon_t^e / \sigma^2 - 1}{2\sigma^2}) + (1 - z_t)(\tfrac{(\beta' X_t)\lambda_t}{2\sigma^3}), \tag{17}$$

$$\lambda_t = \tfrac{F(\beta' X_t / \sigma)}{1 - F(\beta' X_t / \sigma)}. \tag{18}$$

Therefore the LM statistic can be computed as

$$LM = i' G[G' G]^{-1} G' i, \tag{19}$$

where the t^{th} row of G is given by

$$G_t = [a_t W_t, b_t, b_t W_t]. \tag{20}$$

4. Empirical Model

4.1 Data Set

The data used for the empirical study were obtained from the *Survey of Family Expenditures in 1996* released by Statistics Canada in May 1999. Although there are some newer datasets, they do not contain variables considered important to the analysis such as occupation and country of birth. The 1996 survey of family expenditures is the most recent national survey since 1992 that contains a complete list of the required variables. The survey was designed to provide information on persons living in private households in the ten provinces of Canada as well as Whitehorse and Yellowknife. The survey was carried out in January, February and March 1997 and refers to calendar year 1996. The original data include 10,417 households who responded to the survey and in total 239 variables are available. For this study, nineteen life-insurance-related variables are selected. Of the 10,417 observations, after excluding those with missing data, there are 8,931 observations available for estimation.

4.2 Dependent variable

Anderson and Nevin (1975) use the amount of life insurance purchased and the type of life insurance (with or without cash value) as dependent variables. The quantity of life insurance demanded is the dependent variable in the model of Truett and Truett (1990). However, most previous studies on household life insurance purchasing behavior use premium expenditures as the dependent variable (for example, Hammond, Houston, and Melander 1967; Duker 1969; Ferber and Lee 1980; Showers and Shotick 1994), mostly because it is readily available in most datasets. Data on the type of life insurance and the quantity of insurance are less often available.

Since the life insurance premium is positively related to the quantity demand and is available in the dataset, in the empirical model the life insurance premium (LIP) is used as the dependent variable. A histogram of this variable is shown in Figure 1, from which it can be seen that there are many zero values in the dataset. This explains why it is appropriate to use the Tobit model, which allows one to analyze the purchasing decision and the quantity demanded together.

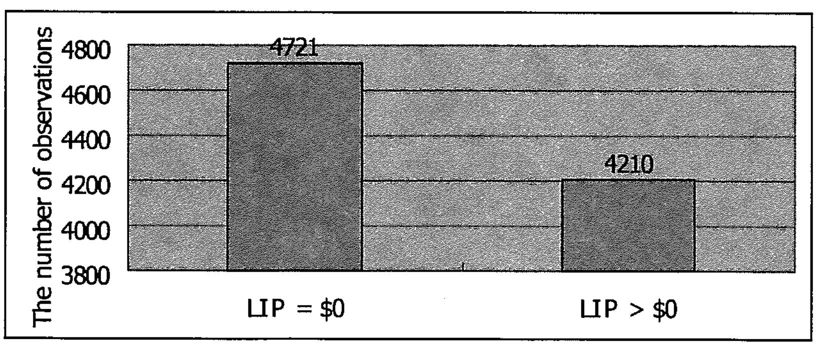

Figure 1 Histogram of LIP

4.3 Independent Variables

The empirical model was constructed using household-related variables that were available in the dataset, including demographic and economic factors.

Canada has few people, but a relatively uneven geographic distribution of its population. Consequently, different conditions, such as provincial regulations, the crime rate, air pollution, transportation problems, etc., exist in the different provinces, which probably results in different risks and different attitudes towards life insurance. Considering this factor, the estimating equation includes four dummy variables to represent four geographic zones: Ontario (ONT), Quebec (QUE), the three Prairie provinces (PRA), and British Columbia (BC).[2] Whether the household resides in an urban (URBA) or rural area is another factor that will affect the demand and it can be investigated using one dummy variable. According to this survey, urban areas are defined as follows: "urban areas have minimum population concentrations of 1000 and a population density of at least 400 per square kilometer."[3]

As prior studies show, income is positively related to the demand for life insurance. If income is low, households cannot afford the premium even though they need insurance to avoid the risk of financial loss. Thus after-tax income of the respondent (IAT) is used and expected to be positively related to the demand for life insurance. The spouse's after-tax income (SIAT) should have the same effect as the respondent's income. In a household with two potential earners, members can share risks, such that the potential future earnings of one spouse can affect the need for life insurance on the other. Given the rise in the number of two-earner couples in Canada, the role of the spouse's earnings has become increasingly important. The wealth level is also assumed to be positively related to the demand for life insurance. Whether the family owns or rents the house (OWNR), as well as the number of rooms (NUMR) can be used to represent the wealth.

Different families have different living styles that give us other indications for the

[2] Atlantic Canada is the reference region.
[3] Quoted from *1996 Survey of Family Expenditures PUBLIC-USE MICRODATA FILE*, version 4, pp. 15

study, such as the total balance owing on principal mortgage (BOPM) and expenditure on cigarettes (CIGS). If people like more loans or mortgages in their lives, they usually need to buy more insurance to protect their beneficiaries from this liability. Since there is a strong link between cigarette smoking and health, expenditure on cigarettes is an indicator of potential heath problems that would increase the probability of death. On the other hand, the premium for the smoker is higher than the non-smoker for the same amount of life insurance, thus the higher cost will decrease their demand. The total effect on the demand is unclear.

Some religions have a strong source of cultural opposition to life insurance, such as Christianity, Islam and Buddhism. People who believe in God always believe that God will protect them forever; purchasing life insurance implies no trust in God. So the percentage of total income on religions (REOR) is included as an explanatory variable in the model, which is defined as expenditure on religions divided by the sum of respondent's and spouse' after-tax income. This variable is expected to be negatively related to the demand for life insurance.

The level of education is another element to be studied in this model. People with higher education may have more knowledge of risk and thus more readily accept the concept of life insurance, which leads us to expect that the level of education is positively related to the demand for life insurance. Education is included in the model in the form of four dummy variables that are used to categorize five classes: less than 9 years of education (reference variable), secondary education (HE), some post-secondary non-university (PSN), post-secondary certificate or diploma (PSU) and university degree (UNV). For much the same reason, occupation is another important element that cannot be ignored in this study. High-risk occupations such as construction are likely to make people think more about safety and the impact of financial loss on their family. That is why I divided occupation into three categories. The reference group is low-risk occupations including managerial and administrative, professional and technical, teaching, sales and service. One dummy variable is used for higher risk occupations (OCCU) such as farming, fishing, logging, mining, construction, machining, assembling and repair. Another one (RET) is used for retired people who do not need to work.

Canada is a country of immigrants, such that residents come from many different cultures and regions across the world. These different traditions and backgrounds may have a significant effect on life insurance demand. People who come from different cultures and countries may have different perspectives on life insurance. In addition, there are differences between developed and developing countries in the level of development of the insurance industry.

In the Survey of Family Expenditures, the "country of birthplace" variable indicates the respondent's birthplace. Therefore, this variable can be used in the regression model to identify immigrants. According to the criteria of the Survey, four dummy variables are defined and included to differentiate five regions of origin: native-born Canadian (reference variable), North America and Western Europe (NAWE); Southern and Eastern Europe (SEU); Asia and Oceania (ASOC); and other countries (OTHR).

Even though attitudes towards risk are difficult to measure, the variable called "Games of chance expenses" (GCEX) is included in the model. GCEX is the percentage of total income (IAT plus SIAT) the household spends on games of chance, which represents their attitude toward fortune and risk. A higher level of GCEX may indicate more demand for life insurance, because they know there is risk in their lives and believe that even accidents with a small probability may happen to them.

Since a family with multiple earners usually has more total assets and lives in a higher-valued house, the households will probably pay more consideration to protecting their assets through the purchase of life insurance. Compared to households with no wage earners, the estimated coefficients of both the one-earner (NFTE1) and two or more earners (NFTE2) dummy variables are expected to be positive.

The remaining explanatory variables are marital status (MARR and DIV), family size (SIZE), gender (MALE) and the age (AGE) of the respondent, which will also affect the demand for life insurance, since people who are married or who have more children have more obligations to protect family members including their spouse and children than the single person. Generally speaking, males have more responsibility than females to protect their family. Furthermore, as people's age increases, their income also increases, and thus they may demand more life insurance.

As a result, the life insurance demand empirical model for this study can be summarized in the following equation:

$$
\begin{aligned}
LIP = {} & \beta_0+\beta_1 ONT+\beta_2 QUE+\beta_3 PRA+\beta_4 BC+\beta_5 URBA+\beta_6 NUMR+\beta_7 BOPM \\
& +\beta_8 IAT+\beta_9 MARR+\beta_{10} DIV+\beta_{10} AGE+\beta_{12} MALE+\beta_{13} SE+\beta_{14} PSN \\
& +\beta_{15} PSU + \beta_{16} UNV + \beta_{17} OCCU + \beta_{18} RET+\beta_{19} NAWE+\beta_{20} SEU \\
& +\beta_{21} ASOC+\beta_{22} OTHR+\beta_{23} SIAT+\beta_{24} SIZE+\beta_{25} NFTE1+\beta_{26} NFTE2 \\
& +\beta_{27} CIGS+\beta_{28} GCEX +\beta_{29} REOR+\beta_{30} OWNR+\mu \, . \qquad (21)
\end{aligned}
$$

Compared with the Engel functions used by Bernstein and Geehan (1988), there are more independent variables included in equation (21), such as education, occupation and country of birth. Moreover, they use total insurance, not just life insurance as the dependent variable.

The exact definitions of all the variables are given in table 1. The expected sign of the coefficient of each variable is also indicated in the table.

Some descriptive statistics for the thirty variables in the empirical model are given in table 2. There are twenty-one dummy variables and nine numeric variables. Approximately 52% of the sample is male, which is reasonable. We also notice that the after-tax income (IAT or SIAT) is sometimes negative. That is because the income is the after-tax cash value in the survey year and could be negative if the respondent is self-employed. The mean of each dummy variable, which measures the percentage of the sample with the given characteristic, is also given in the following figure 2.

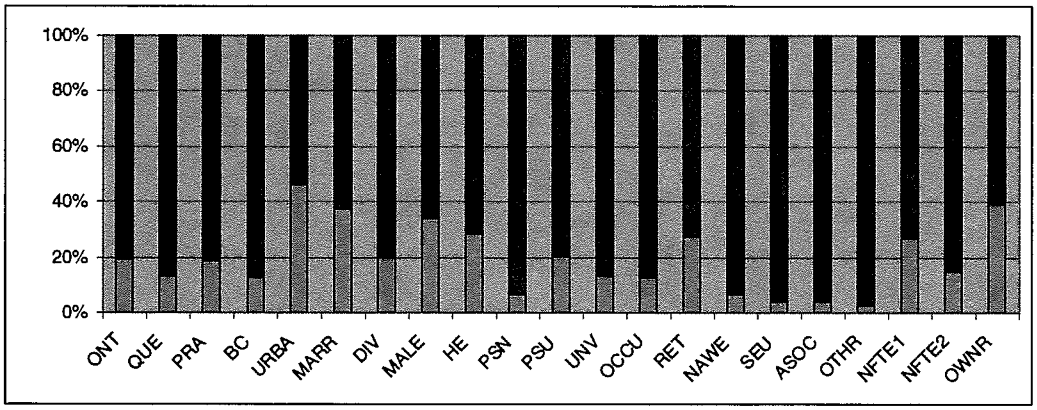

Figure 2. Representation of Household Characteristics in Sample

Figure 2 shows that nearly 20% of the population are immigrants, which includes 7% from North America and Western Europe, 4% from southern and Eastern Europe, 4% from Asia and Oceania, and 3% from other regions. Among all of the occupations, about 15% are high-risk occupations; 30% of respondents are retired.

5. Analysis of Results[4]

5.1 Test for Heteroskedasticity

In an attempt to correct for heteroskedasticity, Showers and Shotick (1994) used the square root of insurance premiums as the dependent variable in their Tobit regression model, instead of the actual level of premiums. Therefore, in this paper the

[4] All estimation was carried out using SHAZAM 9.0.

LM test for heteroscedasticity is carried out in the Tobit regression model with LIP, the square root of LIP and the natural logarithm of LIP as the dependent variable.[5] After defining the vector W to include all the variables in equation (21) (except the one that multiplies the constant term), the result of the LM test for heteroskedasticity is given in table 3. In all cases, the chi square test rejects the null hypothesis of homoskedasticity at the 1 percent level of significance.

Table 3. Tests for heteroskedasticity

Dependent variable	Chi Square Value	Degree of Freedom	0.01 Level Significant
LIP	2818.7	30	Yes
SQRT(LIP)	1869.7	30	Yes
Log(LIP)	2522.9	30	Yes

I also tried the Likelihood Ratio (LR) test with equation (13) to estimate the model with heteroskedasticity, since according to wooldridge (2002) it is more reliable. However unless the likelihood function converges when estimating the model with heteroskedasticity, the LR test for heteroskedasticity is not valid. In my case, although many options were tried, including changing the convergence criterion and step size, using different algorithms, and even reducing the vector W to just IAT and SIAT, convergence was not achieved when trying to estimate the Tobit model with heteroskedasticity. Hence the likelihood ratio test for heteroskedasticity could not be carried out.

As shown above, using the square root or logarithm of LIP as the dependent variable will not eliminate the heteroskedasticity and will make interpretation of the results more complex. Consequently it is safe and convenient to use LIP as the dependent variable. However, future research to solve the heteroskedasticity problem would be of interest.

[5] In the third case, only nonzero values of LIP are converted to natural logs.

5.2 Coefficient Estimates

The Tobit model regression results are summarized in table 4. Because it proved impossible to estimate the Tobit model with heteroskedasticity, these estimates are for a model that assumes no heteroskedasticity. There are two coefficient estimates shown in the table: the normalized coefficient (α) and the regression coefficient (β). The relationship between these two estimates is reported in the following equation:

$$\alpha = \frac{\beta}{\sigma} \tag{22}$$

where σ is standard error of the estimate.

As indicated in the notes to the table, asterisks are used to indicate the coefficients that are significant at the 0.1, 0.05 and 0.01 levels. Among the 30 coefficients, eighteen are significant at the 1% level, two are significant at the 5% level, five are significant at the 10% level and just five coefficients are not significantly different from zero. Among the variables with significant coefficients, the respondent's after-tax income (IAT), age (AGE), educational level (SE, PSN, PSU, UNV), number of earners (NFTE1, NFTE2), family size (SIZE) and home ownership (OWNR) have positive relationships with life insurance purchases, which is consistent with expectations.

The respondent's after-tax income (IAT) is an important factor in the study of life insurance demand. A family with a higher income level has a stronger desire to maintain its higher level of quality of life, such that they would like to purchase more life insurance to protect their financial security. The estimated coefficient of IAT is significant at the 1% level. This result is consistent with prior studies, for example, Hammond, Houston and Melander (1967), Duker (1969), Ferber and Lee (1980), Truett and Truett (1990) and Showers and Shotick (1994).

In earlier studies, Hammond, Houston and Melander (1967) and Mantis and Farmer (1968) found that marriage is negatively related to life insurance demand. Later, Burnett and Palmer (1984) found that it does not have a significant impact on life insurance demand. But in table 3, marriage is positively related to life insurance demand, at the 1% level of significance. This result implies that married individuals feel a greater obligation to protect their spouse in the case of his or her death than a single individual who has no such obligation.

In practice, life insurance premiums depend partly on the age of the insured person. In addition, the age of the survey respondent is also used to represent the stage of the family. Consequently, age (AGE) is always considered an important explanatory variable in studies of life insurance. In this study, age is significantly positive as related to life insurance demand at the 1% level, which is consistent with the results of Berekson (1972), Showers and Shotick (1994) and Truett and Truett (1990). They also found that age is positively significant in determining life insurance demand. However, studies by Ferber and Lee (1980) and Chen, Wong and Lee (2001) find a different result: that age is negatively related to life insurance purchases. Hammond and Houston (1967) and Anderson and Nevin (1975) found that age does not have a significant effect.

Generally speaking, the demand for life insurance can be divided into different stages according to the respondent's age. Before marriage, a single person needs little insurance. Then, after marriage a family with a house and children needs to purchase more life insurance to ensure the financial security of the whole family. Finally, when people are close to retirement age, their mortgage has been paid off and fewer dependents need to be taken care of. Therefore, they just need a limited amount of life insurance to cover funeral costs. Different datasets that include different age groups are analyzed in the previous studies, which is why different results and conclusions are obtained.

Common sense suggests that a higher level of education will give a person more rational thinking. They will more easily recognize risks and have a stronger desire to protect dependents. Hammond, Houston and Melander (1967), Ferber and Lee (1980) and Truett and Truett (1990) found that education is significantly positively related to the demand for life insurance. The estimates obtained here are almost consistent with their results, since all of the coefficients are positive, especially the coefficient estimate for the group with post-secondary education (PSN) which is significant at the 1% level and the coefficient estimate for individuals with a university degree (UNV) which is significant at the 10% level. These results that these groups indicate have a higher demand for insurance than the group with just an elementary education. However, one surprising thing is that the coefficients do not seem to increase with the level of education. The group with some post-secondary non-university (PSN) seems to require the most life insurance. In addition, the demand of those with a university degree (UNV) is lower than that of those with a post-secondary certificate or diploma (PSU).

Compared to households with no wage earners, the estimated coefficients of both the one-earner (NFTE1) and two or more earners (NFTE2) dummy variables are strongly positive and are significant at the 1% level in this study. A family with multiple earners usually has more total assets and lives in a higher-valued house. Due to a higher legacy tax, they will probably pay more consideration to protecting their assets intact for their children through the purchase of life insurance. That is also the reason why the coefficient of the after-tax income of the spouse (SIAT) has a significantly positive effect on life insurance demand. Both estimates are consistent with expectations. Other studies such as Ferber and Lee (1980), and Showers and Shotick (1994) obtain the same result. However, the results of Duker (1983) and Gandolfi and Miners (1996) are the opposite. They explain their result by arguing that a family with multiple earners has a higher income level, and thus has the ability to control its financial security in the event of risk with less life insurance.

Earlier studies found that family size (SIZE) has an impact on life insurance demand. The coefficient of family size (SIZE) in this model is strongly significant at the 1% level, with a positive sign, which is consistent with previous expectations. Berekson (1972), Ferber and Lee (1980), Showers and Shotick (1994) and Bernstein and Geehan (1988) also found that family size is significantly positively related to life insurance purchases. The family with more dependants perhaps buys more life insurance to protect its members until they are no longer dependent on the family.

The coefficient of the home ownership variable (OWNR) is highly significant with a positive sign. People who own property usually have a strong desire to protect the asset and maintain their quality of life with life insurance. On the other hand, people who are tenants usually do not have many assets to be protected with life insurance. In relevant studies, Anderson and Nevin (1975) and Bernstein and Geehan (1988) found that homeownership and type of housing have significant positive effects on life insurance demand. Other studies by Ferber and Lee (1980) and Gandolfi and Miners (1996) obtained similar results.

The coefficient of the number of rooms in the house (NUMR) is also positive and highly significant at the 1% level, which is consistent with expectations. A house with more rooms is an indicator of more wealth, suggesting that the households need to purchase more life insurance to protect their assets and quality of life in case of an income earner's death. Hammond, Houston and Melander (1967), Anderson and

Nevin (1975) and Hau (2000) all found that the level of family wealth has a positive relationship with life insurance consumption.

The coefficient of the total balance owing on the principal mortgage (BOPM) is significantly negative, which is not consistent with expectations. Generally speaking, people with more loans or mortgages need to buy more insurance to protect their beneficiaries from this liability. One possible explanation for the negative effect is that maybe they do not have enough money to buy the additional insurance to protect the family. Another explanation for the negative effect is that they already have some mortgage insurance from a bank or a mortgage company because of the creditor's requirement, so that they do not need life insurance to cover the balance outstanding on their mortgage. Previous studies did not include such a variable. This factor should be investigated in further empirical research so that its effects can be better understood.

Some prior studies took into account geographic location factors in their models. Such factors are also investigated in this study. The regression results show that geographic factors have a significant impact on the life insurance demand. Surprisingly, residents of Ontario (ONT), the Prairies (PRA) and B.C. (BC) purchase less life insurance than residents of Atlantic Canada, while Quebec (QUE) residents purchase more. These differences may be due to differences in industries and occupations among the provinces. Commercial and financial industries are particularly important to the Ontario economy. A higher proportion of people in Ontario may work in low-risk occupations compared to provinces where fishing, mining, or logging is more important. At the same time, people in Ontario may prefer to use financial investments to avoid some risk and diversify their assets. That may be why the demand for life insurance in Ontario is less than that in Atlantic Canada. In

B.C. and the Prairies, the occupations of people may also be less risky. Many manufacturing industries located in Quebec, where people may need more life insurance to avoid the occupation-related risk. By way of comparisons Bernstein and Geehan (1988) also found that Quebec residents spent the most on life insurance, but in their study Ontario residents spent more than those of the remaining provinces.

Another geographic location factor is whether the place of residence is urban or rural. The results indicate that there is no significant difference in life insurance purchases between people who live in urban (URBA) and rural areas. Theoretically speaking, in an urban area with a large population, the economy develops fast and business activities are frequent, leading to problems such as a higher crime rate, air pollution, etc. Such factors may encourage people to purchase more insurance. Besides, many insurance salespeople are active in urban areas, which should also increase sales of insurance. Bernstein and Geehan (1988) also show that inhabitants of larger urban areas spend more on insurance. The explanation for the insignificant coefficient of URBA is still unclear and needs to be examined in the future.

The coefficient estimates for the occupational dummies are also different than were expected. The coefficient of the variable for high-risk occupations (OCCU) is negative, but it is not statistically significant. This is surprising, because usually people in high-risk occupations would recognize its effect on their family, and therefore purchase more life insurance because it is one of the best ways to reduce the loss if something happens to them. Being retired with no occupation (RET) has a significant negative impact on life insurance demand compared with people in low-risk occupations. This result is reasonable because after people retire, they face less work-related risk and have fewer liabilities and dependents, so they need less life insurance.

To check whether there is a correlation between the province variables, URBA and the occupation variables, the condition index for this subset of variables was computed. The largest condition index is 14.661, which means there is not a strong linear relationship between these variables. Thus multicollinearity between these variables cannot explain the insignificance of the coefficients of UBRA and OCCU.

Four dummy variables (NAWE, SEU, ASOC, OTHR) were derived from the country of birth variable and included in the model to study the effect of immigration on life insurance. Most of them have significant coefficients. The coefficient of NAWE implies that people from developed regions such as the UK, North America and Western Europe need less life insurance to cover risk than native-born Canadians; its coefficient is significant at the 5% level. Although these countries are developed countries, those immigrants are usually not very rich and thus cannot afford too much life insurance. Families from Southern or Eastern Europe (SEU) also need less insurance than native-born Canadians. Only people from Asia or Oceania (ASOC) demand more life insurance than native–born Canadians, but the coefficient of ASOC is not very significant. People who come from other countries such as the Republic of South Africa, Kenya or Niger (OTHR) demand less life insurance than native-born Canadian and the coefficient is significant at the 5% level.

The coefficient of expenditure on cigarettes, cigars and similar products (CIGS) is negative but not significant in this model. Insurance companies always investigate the health of the life-insured by asking medical questions and for information about lifestyle. Smokers may face a higher premium for life insurance than non-smokers. Consequently, smokers have less desire to buy life insurance because they face a higher price level, even though they actually need more life insurance to protect their family from their health problems.

In this paper, we use GCEX to represent how much risk and chance they like. That is, the higher GCEX, the more money spent on games of chance, and the greater the awareness of risk. The regression results show that people who spend a higher proportion of total income on games of chance such as lotteries and bingo also purchase more life insurance. The coefficient of GCEX is positive and significant at the 10% level, which is consistent with expectations. However, the explanation of this result is much more complicated. According to prospect theory, which was proposed by Kahneman and Tversky (1979), people tend to overweight small probabilities. That is, they behave as if the probability of an event with a small probability is higher than it actually is, whether they are facing a gain or a loss. Since the chance of winning a lottery is quite small and the chance of dying is also very small, the conditions for overweighting are satisfied in the cases of both gambling and insurance. Therefore, people who buy lottery tickets also tend to spend more money on insurance.

Zelizer (1979) notes that religion has historically provided a strong source of cultural opposition to life insurance; many religious people believe that a reliance on life insurance results from distrust of God's protecting care. Religious antagonism to life insurance still remains in several Islamic countries. Browne and Kim (1993) found that religion does not have a significant effect on life insurance purchases, but Burnett and Palmer (1984) found that it is negatively related to life insurance demand. However, Fitzgerald (1989) found that the bequest motive has a significant positive effect on life insurance demand. The regression results in this study show that families that spend more on religious organizations (REOR) will buy less life insurance. The coefficient of REOR is negatively related to the demand for life insurance but not significant. The sign of the result is consistent with expectations. Since religious individuals have the strong belief that God will protect them very well, they may not need life insurance for their financial security. However, the coefficient is insignificant. The possible reason is that the bequest motive counteracts the effect of strong belief, and thus the religion effect is neutralized.

According to the standard theoretical model of consumer behaviour, consumers choose expenditures on all goods and services simultaneously to maximize their utilities. Therefore from an econometric point of view, the variables CIGS, GCEX and REOR are likely to be correlated with the error term in the equation estimated by this study. To examine the effect of these three variables, the regression model was re-estimated with the three variables excluded completely. Table 5 presents the estimates for this Tobit regression. Compared to the results in table 4, the coefficient of male (MALE) in the new model changes sign to become positively related to life insurance demand, but it is still not statistically significant. At the same time, the coefficient of MARR in the new model is still positively related to life insurance demand, but significant at the 5% level. However, the coefficient of MARR in the original model with all variables is significant at the 1% level. Aside from the above changes, there are no other important changes in the results.

It is surprising that in our regression results the gender of the respondent (MALE) has no significant effect on life insurance demand. A similar study by Gandolfi and Miners (1996) found that gender is significantly positively related to life insurance demand. Bernstein and Geehan (1988) also show that female-headed households spend a lower share on life insurance than male-headed households.

5.3 Marginal Effect Analysis

Table 6 provides estimates of the marginal effects or partial derivatives of insurance expenditures. The decomposition is calculated according to equations (7) and (8) in section 3. The marginal effect when considering only purchasers of insurance is represented by $\partial E(LIP \mid LIP > 0) / \partial E(X_i)$, and the marginal effect on the probability of purchasing insurance is expressed by $\partial E(F(z) / \partial E(X_i)$. The predicted valued of total insurance demand at the mean values of the variables for current purchasers is \$1084.73. On the other hand, there is an estimated 38 percent likelihood that an individual will make the decision to purchase life insurance to reduce the cost of unexpected risk.

The computed elasticity at the sample mean values and its decomposition according to equations (11)-(12) in section 3 are also reported in table 6. It provides estimates of the marginal change in the demand for insurance and the change in the probability of purchasing insurance. According to the table, the total elasticity estimates in the rightmost column are all smaller than 1, which implies an inelastic response to changes in all of the dependent variables.

The respondent's after-tax income (IAT) has the highest impact on the probability of purchasing life insurance and the demand for life insurance. The total elasticity value is 0.32, which is decomposed into 0.1 for the elasticity of purchasing probability and 0.22 for the elasticity of expected life insurance demand. This suggests that a 10% increase in income would result in an increase of about 3.2% in total life insurance demand. Specifically, the probability of purchasing life insurance will increase by 1%, while the expected amount of life insurance demand from previous purchasers will increase by 2.2%. However, the elasticity with respect to the spouse's after-tax income (SIAT) is only 0.156, which is smaller than the elasticity with respect to IAT. This is due to the fact that multiple-income earners may respond less to their spouse's income increase, since they have multiple income sources and need less life insurance to protect the family members.

The second highest impact comes from family size (SIZE). The total elasticity is 0.258, which is divided into 0.078 for the elasticity of the probability of purchasing and 0.18 for the elasticity of purchasers' demand. Being married (MARR) has an estimated total elasticity of 0.12, which is decomposed into 0.04 for the elasticity of the probability of purchasing life insurance and 0.08 for the expected life insurance demand.

A one percent increase in age (AGE) is estimated to increase insurance demand by 0.22% in total, which is decomposed into 0.067 for the elasticity of the probability of purchasing life insurance and 0.156 for the elasticity of the expected insurance demand. The estimated total elasticity with respect to home ownership (OWNR) is 0.18, which is decomposed into 0.05 for the elasticity of the probability of purchasing life insurance and 0.13 for the elasticity of the expected life insurance demand.

An increase in the number of earners from none to one (NFTE1) is estimated to increase the total insurance demand by 0.14, which is decomposed into 0.04 for the elasticity of the probability of purchasing life insurance and 0.10 for the elasticity of the expected life insurance demand. However, increasing the number of earners from one earner to two earners or more in the family (NFTE2) has a smaller impact. The estimated total elasticity is 0.06, of which 0.02 is the elasticity of the probability of purchasing life insurance and 0.04 is the elasticity of the expected life insurance demand. This means that the marginal change decreases with the number of earners. The families with only one earner have a greater probability of purchasing life insurance to protect their financial security. Multi-earner families perhaps have a lower demand for life insurance than single-earner families due to their income being higher. This result is consistent with the results of Showers and Shotick (1994), who explain that multi-earner households have a greater likelihood of receiving some type of employer-paid or subsidized life insurance.

The estimated marginal effect with respect to the number of rooms (NUMR) is 0.01 for the probability of purchasing life insurance, and \$13 for the expected life insurance demand. This means when the number of rooms is increased by 1, the probability that a household without life insurance will buy life insurance will increase 1% and the household with life insurance will spend \$13 more on life insurance premiums.

The estimated marginal effect of being retired and not working (RET) is \$123.69, which means that if a person is retired from his occupation, he will spend \$123.69 less on life insurance.

People who come from the UK, North America, Northern or Western Europe (NAWE) spend about \$48.97 less on life insurance premiums. The marginal effect coming from Southern or Eastern Europe (SEU) is –\$125.53. People who come from Asia and Oceania (ASOC) spend \$9.98 more than native-born Canadians. People who come from all other countries spend \$86.47 less than Canadian natives.

6. Conclusion

This study estimates the economic and demographic factors that influence the demand for life insurance in Canada using Tobit analysis. The quantity demanded of life insurance and the probability of purchasing are studied together. It finds that income, marriage, age, educational level, number of earners, family size and home ownership have a significant positive relationship with the demand for life insurance, which is consistent with our expectations and some prior studies. The results also show that immigrants from different regions of the world have a different demand for life insurance compared with the Canadian-born, partly because different cultures have some impact on attitudes towards purchasing life insurance. In the analysis, the marginal effects of the probability of purchasing life insurance and the amount of expenditure on life insurance are also estimated and discussed. The after–tax income of the respondent has the highest elasticity.

There are a number of ways in which this study could be extended. Since Canada has a high level of social security benefits, considering the effect of this factor on life insurance demand will provide a more complete perspective on the demand for life insurance, if a future dataset supports this option. This study also ignored some other factors that may affect the demand for life insurance, such as inflation, interest rates, stock market behaviour, the bequest motive and the price of insurance. Some of the estimation results including those with respect to geographic factors, occupation and gender are not consistent with my expectations and need further investigation. In addition, only the premium of life insurance is available in the dataset used, so it would be worthwhile to examine the effects of these factors on the type of life insurance and the amount of life insurance. Finally, even though the study tests for heteroskedasticity in the Tobit model and tries to resolve the problem with different methods, it was not successful in doing so. Therefore, a future study that resolves this problem would be very interesting.

References

Anderson, D. R., and J. R. Nevin, 1975, Determinants of Young Marrieds' Life Insurance Purchasing Behavior: An Empirical Investigation, *Journal of Risk and Insurance*, 42: 375-387.

Berekson, L. L., 1972, Birth Order, Anxiety, Affiliation and the Purchase of Life Insurance, *Journal of Risk and Insurance*, 39: 93-108.

Bernstein, Jeffrey I. and Randall R. Geehan, 1988, *The Insurance Industry in Canada*, Vancouver: The Fraser Institute.

Browne, M. J., and K. Kim, 1993, An International Analysis of Life Insurance Demand, *Journal of Risk and Insurance*, 60: 616-634.

Burnett, J. J., and B. A. Palmer, 1984, Examining Life Insurance Ownership Through Demographic and Psychographic Characteristics, *Journal of Risk and Insurance*, 51: 453-467.

Campbell, R. A., 1980, The Demand for Life Insurance: An Application of the Economics of Uncertainty, *Journal of Finance*, 35: 1155-1172.

Chen, R., K. A. Wong, and H. C. Lee, 2001, Age, Period and Cohort Effects on Life Insurance Purchases in the U.S., *The Journal of Risk and Insurance*, 68: 303-327.

Duker, J. M., 1969, Expenditures for Life Insurance Among Working-Wife Families, *Journal of Risk and Insurance*, 36: 525-533.

Ferber, R., and L. C. Lee, 1980, Acquisition and Accumulation of Life Insurance in Early Married Life, *Journal of Risk and Insurance*, 47: 713-734.

Fischer, S., 1973, A Life Cycle Model of Life Insurance Purchases, *International Economic Review*, February, 14: 132-152.

Fitzgerald, J. M., 1989, The Taste for Bequests and Well-Being of Widows: A Model of Life Insurance Demand by Married Couples, *Review of Economics and Statistics*, 71: 206-214.

Gandolfi, A.-S., and L. Miners, 1996, Gender-Based Differences in Life Insurance Ownership, *Journal of Risk and Insurance*, 63: 683-693.

Greene, William H., 1997, *Econometric Analysis*, Third Edition, Prentice Hall

Hammond, J. D., D. B. Houston, and E. R. Melander, 1967, Determinants of Household Life Insurance Premium Expenditures: An Empirical Investigation, *Journal of Risk and Insurance*, 34: 397-408.

Hau, A., 2000, Liquidity, Estate Liquidation, Charitable Motives, and Life Insurance Demand by Retired Singles, *Journal of Risk and Insurance*, 67: 123-141.

Kahneman, D., and A. Tversky, 1979, Prospect theory: an analysis of decision under risk, *Econometrica*, 47:263-91.

Lewis, F. D., 1989, Dependents and the Demand for Life Insurance, *American Economic Review*, 79: 452-467.

Mantis, G., and R. Farmer, 1968, Demand for Life Insurance, *Journal of Risk and Insurance*, 35: 247-256.

Mason, Karen O., William E. Manson, H. H. Winsborough, and W. Kenneth Poole, 1973, Some Sociological Issues in Cohort Analysis of Archival Data, *American Sociological Review,* 38: 242-258

Matteo, Livio Di and J.C. Herbert Emery, 2002, Wealth and the demand for life insurance: evidence form Ontario, 1892, *Explorations in Economic History*, 39: 446-469

McDonald, John F. and Robert A. Moffit, 1980, The Uses of Tobit Analysis, *Review of Economics and Statistics*, 62: 318-321

Showers, V. E., and J. A. Shotick, 1994, The Effects of Household Characteristics on Demand for Insurance: A Tobit Analysis, *Journal of Risk and Insurance*, 61: 492-502.

Tobin, James, 1958, Estimation of Relationships for Limited Dependent Variables, *Econometrica*, 26:24-36

Truett, D. B., and L. J. Truett, 1990, The Demand for Life Insurance in Mexico and the United States: A Comparative Study, *Journal of Risk and Insurance*, 57: 321-328.

Wooldridge, Jeffery M., 2002, *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press

Yaari, M. E., 1965, Uncertain Lifetime, Life Insurance, and the Theory of the Consumer, *Review of Economic Studies*, April, 32: 137-150.

Zelizer, Viviana A. Rotman, 1979, *Morals and Markets: The Development of Life Insurance in the United States*, Transaction Books, Columbia University Press.

Zietz, Emily Norman, 2003, An Examination of the demand for life insurance, *Risk Management and Insurance Review*, Vol. 6, No. 2, 159-191

Table 1. Definitions of variables in the empirical model

Dependent Variable		
LIP		Life Insurance Premium, measured in dollars
Independent Variables		
ONT	+	Dummy variable: 1 if resides in Ontario, 0 otherwise.
QUE	+	Dummy variable: 1 if resides in Quebec, 0 otherwise.
PRA	+	Dummy variable: 1 if resides in Prairie province, 0 otherwise.
BC	+	Dummy variable: 1 if resides in British Columbia, 0 otherwise.
URBA	+	Dummy variable: 1 if resides in Urban, 0 if resides in Rural
NUMR	+	Number of rooms in principal dwelling
BOPM	+	Total balance owing on principal mortgage (dollars)
IAT	+	Respondent's income after tax (dollars)
MARR	+	Dummy variable: 1 if married, 0 otherwise
DIV	?	Dummy variable: 1 if divorced or widowed, 0 otherwise
AGE	+	Age of respondent.
MALE	+	Dummy variable: 1 if the respondent is male, 0 if female
HE	+	Dummy variable: 1 if secondary education is highest level attained, 0 otherwise
PSN	+	Dummy variable: 1 if some Post-secondary non-university is highest level attained, 0 otherwise
PSU	+	Dummy variable: 1if Pose-secondary certificate or diploma is highest level attained, 0 otherwise
UNV	+	Education level: 1 if the university degree is highest level attained, 0 otherwise
OCCU	+	Dummy variable: 1 if high–risk occupations, 0 otherwise.
RET	-	Dummy variable: 1 if retired and not working, 0 otherwise.
NAWE	?	Dummy variable: 1 if country of birth is in UK, North America, North or West Europe, 0 other wise
SEU	-	Dummy variable: 1 if country of birth is in South or East Europe, 0 otherwise
ASOC	-	Dummy variable: 1 if country of birth is in Asia or Oceania, 0 otherwise
OTHR	-	Dummy variable: 1 if country of birth is not in NAWE, SEU, or ASOC, 0 otherwise
SIAT	+	Spouse's income after tax (dollars)
SIZE	+	Family size
NFTE1	+	Dummy variable: 1 if number of full- time earners less or equal to one, 0 otherwise
NFTE2	+	Dummy variable: 1 if number of full-time earners equal or more than two, 0 otherwise
CIGS	?	Expenditure on tobacco products (dollars)
GCEX	+	Expenditure percentage of total income on games of chance (percentage)
REOR	?	Expenditure percentage of total income on religious organization (percentage)
OWNR	+	Dummy variable: 1 if owner of home, 0 if tenant

Note:
+: Expected positive effect on life insurance demand
- : Expected negative effect on life insurance demand
? : Unknown expected effect (positive or negative) on life insurance demand

Table 2. Statistics of variables in the empirical model

NAME	*NUM*	*MEAN*	*ST. DEV*	*VARIANCE*	*MINIMUM*	*MAXIMUM*
ONT	8931	0.2381	0.4259	0.1814	0.0000	1.0000
QUE	8931	0.1528	0.3599	0.1295	0.0000	1.0000
PRA	8931	0.2286	0.4200	0.1764	0.0000	1.0000
BC	8931	0.1451	0.3522	0.1241	0.0000	1.0000
URBA	8931	0.8670	0.3396	0.1153	0.0000	1.0000
NUMR	8931	6.0319	2.1867	4.7816	1.0000	11.000
BOPM	8931	2.0184	3.9060	1.5257	0.0000	38.818
IAT	8931	2.3788	1.6384	0.2684	-3.5000	23.000
MARR	8931	0.6049	0.4889	0.2390	0.0000	1.0000
DIV	8931	0.2450	0.4301	0.1850	0.0000	1.0000
AGE	8931	32.300	16.500	54.595	15.000	85.000
MALE	8931	0.5189	0.4997	0.2497	0.0000	1.0000
HE	8931	0.3969	0.4893	0.2394	0.0000	1.0000
PSN	8931	0.0690	0.2534	0.0642	0.0000	1.0000
PSU	8931	0.2538	0.4352	0.1894	0.0000	1.0000
UNV	8931	0.1519	0.3590	0.1289	0.0000	1.0000
OCCU	8931	0.1441	0.3512	0.1234	0.0000	1.0000
RET	8931	0.3738	0.4838	0.2341	0.0000	1.0000
NAWE	8931	0.0719	0.2583	0.0667	0.0000	1.0000
SEU	8931	0.0401	0.1962	0.0385	0.0000	1.0000
ASOC	8931	0.0437	0.2044	0.0418	0.0000	1.0000
OTHR	8931	0.0261	0.1594	0.0254	0.0000	1.0000
SIAT	8931	1.1679	1.5939	0.2540	-1.4000	19.000
SIZE	8931	2.5232	1.3230	1.7503	1.0000	6.0000
NFTE1	8931	0.3652	0.4815	0.2319	0.0000	1.0000
NFTE2	8931	0.1753	0.3803	0.1446	0.0000	1.0000
CIGS	8931	521.60	923.24	852380	0.0000	9034.0
GCEX	8931	0.0103	0.0566	0.0032	-0.1078	4.0000
REOR	8931	0.0133	0.4243	0.1800	-0.1310	40.000
OWNR	8931	0.6484	0.4775	0.2280	0.0000	1.0000
LIP	8931	420.00	1667	2778619	0.0000	130000

Note:
To make the table more readable, the measurement unit of BOPM, IAT and SIAT is changed from dollars to tens of thousands of dollars only for the display of this table.

Table 4. Maximum likelihood estimation results of the life insurance demand Tobit model

VARIABLE Name	NORMALIZED COEFFICIENT	STANDARD ERROR	T-RATIO	REGRESSION COEFFICIENT
ONT	-0.1059	0.0383	-2.7666***	-165.1500
QUE	0.2570	0.0406	6.3262***	400.6900
PRA	-0.0494	0.0368	-1.3419	-77.0570
BC	-0.2399	0.0450	-5.3250***	-373.9300
URBA	0.0094	0.0388	0.2434	14.7100
NUMR	0.0277	0.0078	3.5658***	43.1050
BOPM	0.0000	0.0000	-2.0836**	-0.0012
IAT	0.0000	0.0000	10.1030***	0.0146
MARR	0.1350	0.0512	2.6389***	210.4500
DIV	0.0492	0.0487	1.0113	76.7160
AGE	0.0241	0.0063	3.8499***	37.5790
MALE	-0.0010	0.0308	-0.0329	-1.5814
HE	0.0805	0.0463	1.7404*	125.5600
PSN	0.1710	0.0643	2.6589***	266.6100
PSU	0.1276	0.0513	2.4870**	198.8800
UNV	0.1011	0.0579	1.7459*	157.6100
OCCU	-0.0577	0.0391	-1.4749	-89.9630
RET	-0.2630	0.0426	-6.1694***	-409.9900
NAWE	-0.1041	0.0508	-2.0509**	-162.3100
SEU	-0.2669	0.0706	-3.7809***	-416.0900
ASOC	0.0212	0.0636	0.3336	33.0820
OTHR	-0.1839	0.0855	-2.1496**	-286.6100
SIAT	0.0000	0.0000	8.9040***	0.0145
SIZE	0.0712	0.0132	5.3993***	110.9700
NFTE1	0.2685	0.0357	7.5180***	418.6100
NFTE2	0.2371	0.0453	5.2312***	369.6800
CIGS	0.0000	0.0000	-1.6204	-0.0346
GCEX	0.4164	0.2183	1.9080*	649.1700
REOR	-0.0300	0.0558	-0.5375	-46.7350
OWNR	0.1944	0.0368	5.2822***	303.0500
CONSTANT	-1.4156	0.0846	-16.7340	-2206.8000
LIP	0.0006	0.0000	87.0850	

Variance of the estimate = 0.2430E+07
Standard error of the estimate = 1558.9
The observed frequency of Y > limit is = 0.4714
At mean values of all X(I), E(Y) = 411.2619
Log-likelihood function = -39243.500
Squared correlation between observed and expected values = 0.088701

Note:

* Significant at 0.1 level, ** Significant at 0.05 level, *** Significant at 0.01 level

Table 5. Maximum likelihood estimation results of the life insurance demand Tobin model (with 3 independent variables deleted)

VARIABLE Name	NORMALIZED COEFFICIENT	STANDARD ERROR	T-RATIO	REGRESSION COEFFICIENT
ONT	-0.1016	0.0382	-2.6575***	-158.3500
QUE	0.2595	0.0406	6.3950***	404.6300
PRA	-0.0487	0.0368	-1.3229*	-75.9820
BC	-0.2387	0.0450	-5.3002***	-372.1600
URBA	0.0099	0.0388	0.2559	15.4680
NUMR	0.0279	0.0078	3.5966***	43.4690
BOPM	0.0000	0.0000	-2.1334**	-0.0012
IAT	0.0000	0.0000	10.0630***	0.0145
MARR	0.1299	0.0511	2.5436**	202.5400
DIV	0.0464	0.0486	0.9547	72.3550
AGE	0.0242	0.0063	3.8620***	37.6800
MALE	0.0028	0.0308	0.0907	4.3506
HE	0.0773	0.0462	1.6706*	120.4700
PSN	0.1684	0.0643	2.6201***	262.5500
PSU	0.1283	0.0511	2.5079**	199.9900
UNV	0.1039	0.0576	1.8037*	162.0100
OCCU	-0.0608	0.0391	-1.5567	-94.8620
RET	-0.2598	0.0426	-6.1002***	-405.0700
NAWE	-0.1040	0.0508	-2.0491**	-162.1800
SEU	-0.2670	0.0706	-3.7836***	-416.3600
ASOC	0.0335	0.0633	0.5292	52.2530
OTHR	-0.1770	0.0853	-2.0736**	-275.9300
SIAT	0.0000	0.0000	8.9015***	0.0145
SIZE	0.0696	0.0131	5.2930***	108.5200
NFTE1	0.2665	0.0357	7.4701***	415.6100
NFTE2	0.2330	0.0452	5.1541***	363.3400
CIGS	0.1990	0.0367	5.4170***	310.3000
CONSTANT	-1.4216	0.0839	-16.9410	-2216.8000
LIP	6.41E-04	7.36E-06	87.081	

Variance of the estimate = 0.24314E+07
Standard error of the estimate = 1559.3
The observed frequency of Y > limit is = 0.4714
At mean values of all X(I), E(Y) = 411.5191
Log-likelihood function = -39246.491
Squared correlation between observed and expected values = 0.088613

Note:

* Significant at 0.1 level, ** Significant at 0.05 level, *** Significant at 0.01 level

Table 6. Partial derivatives and elasticity decompositions at the mean value

Variables	Partial Derivatives		Elasticity of		
	$\frac{\partial F(z)}{\partial X_i}$	$\frac{\partial LIP}{\partial X_i}$	Purchasing Probability	Expected Insurance	Total
ONT	-0.0403	-49.8248	-0.0109	-0.0253	-0.0362
QUE	0.0978	120.8847	0.0170	0.0394	0.0565
PRA	-0.0188	-23.2476	-0.0049	-0.0113	-0.0162
BC	-0.0913	-112.8122	-0.0151	-0.0349	-0.0500
URBA	0.0036	4.4380	0.0035	0.0082	0.0118
NUMR	0.0105	13.0043	0.0723	0.1674	0.2397
BOPM	0.0000	-0.0004	-0.0066	-0.0153	-0.0219
IAT	0.0000	0.0044	0.0964	0.2231	0.3195
MARR	0.0514	63.4908	0.0354	0.0819	0.1173
DIV	0.0187	23.1448	0.0052	0.0121	0.0173
AGE	0.0092	11.3373	0.0676	0.1564	0.2239
MALE	-0.0004	-0.4771	-0.0002	-0.0005	-0.0008
HE	0.0306	37.8793	0.0139	0.0321	0.0459
PSN	0.0651	80.4352	0.0051	0.0118	0.0170
PSU	0.0485	60.0021	0.0140	0.0325	0.0465
UNV	0.0385	47.5497	0.0067	0.0154	0.0221
OCCU	-0.0220	-27.1412	-0.0036	-0.0083	-0.0120
RET	-0.1001	-123.6919	-0.0426	-0.0986	-0.1413
NAWE	-0.0396	-48.9673	-0.0032	-0.0075	-0.0108
SEU	-0.1016	-125.5318	-0.0046	-0.0107	-0.0154
ASOC	0.0081	9.9805	0.0004	0.0009	0.0013
OTHR	-0.0700	-86.4674	-0.0021	-0.0048	-0.0069
SIAT	0.0000	0.0044	0.0470	0.1087	0.1557
SIZE	0.0271	33.4793	0.0779	0.1803	0.2581
NFTE1	0.1022	126.2902	0.0425	0.0984	0.1410
NFTE2	0.0902	111.5285	0.0180	0.0417	0.0598
CIGS	0.0000	-0.0104	-0.0050	-0.0116	-0.0166
GCEX	0.1584	195.8493	0.0019	0.0043	0.0062
REOR	-0.0114	-14.0995	-0.0002	-0.0004	-0.0006
OWNR	0.0740	91.4269	0.0547	0.1265	0.1812
E[LIP|LIP>0]	1084.729				
E[F(z)]	0.3791				

[a]: Elasticity estimates at the mean values of LIP and X_i

E[F(z)]: Represent the predicted probability of $Y > 0$ given average X(I).

$\frac{\partial F(z)}{\partial X_i}$: Represent a change in the purchasing probability for a unit change in X_i

$\frac{\partial LIP}{\partial X_i}$: Represent a change in the amount of insurance demand for a unit change in X_i

Appendix 1. Shazam program for Tobin regression analysis

```
PAR 8048
SAMPLE 1 10417
READ (C:\\DATA.TXT) AB AGEGRP   BOPOMO   CIGS  CITY  COB   &
      EDUC GCEX IAT   LIP LQOHB31    MARST    MOPM96   &
      NCAL NFTE NPTE  NUMRM  OCCU  PROV  RELOR  SAGEGRP  &
      SCOB SEDUC SEX  SIAT  SOCC WEIGHT CLTEN
SET NOWARNSKIP
*--ONTARIO
GENR PROV1 = PROV.EQ.35
*--QUEBEC
GENR PROV2 = PROV.EQ.24
*--CENTRE PROVIENCE
GENR PROV3 = (PROV.EQ.46).OR.(PROV.EQ.47).OR.(PROV.EQ.48)
*--BC
GENR PROV4 = PROV.EQ.59
SKIPIF PROV.EQ.00

GENR URBAN = CITY.EQ.1
SKIPIF CITY.EQ.0

GENR MARRIED = MARST.EQ.1
GENR DIVORCED = MARST.EQ.3

GENR MALE = SEX.EQ.1

GENR EDUC1 = EDUC.EQ.2
GENR EDUC2 = EDUC.EQ.3
GENR EDUC3 = EDUC.EQ.4
GENR EDUC4 = EDUC.EQ.5
SKIPIF EDUC.EQ.6

GENR OCCU1 = (OCCU.LE.10).AND.(OCCU.GE.7)
GENR OCCU2 = OCCU.EQ.12
SKIPIF (OCCU.EQ.11).OR.(OCCU.EQ.13)

GENR COB1 = COB.EQ.2
GENR COB2 = COB.EQ.3
GENR COB3 = COB.EQ.4
GENR COB4 = COB.EQ.5

GENR SIZE = MOPM96

GENR NFTE1 = NFTE.EQ.1
GENR NFTE2 = NFTE.EQ.2

GENR OWNER = (CLTEN.EQ.1).OR.(CLTEN.EQ.2)
SKIPIF CLTEN.EQ.4
```

```
STAT PROV1 PROV2 PROV3 PROV4 URBAN &
   NUMRM BOPOMO IAT MARRIED DIVORCED AGEGRP MALE EDUC1 &
   EDUC2 EDUC3 EDUC4 OCCU1 OCCU2 COB1 COB2 COB3 COB4 &
   SIAT SIZE NFTE1 NFTE2 CIGS GCEX RELOR OWNER &
   /MEAN = X

**----------TOBIT REGRESSION --------------------------
TOBIT LIP PROV1 PROV2 PROV3 PROV4 URBAN &
   NUMRM BOPOMO IAT MARRIED DIVORCED AGEGRP MALE EDUC1 &
   EDUC2 EDUC3 EDUC4 OCCU1 OCCU2 COB1 COB2 COB3 COB4 &
   SIAT SIZE NFTE1 NFTE2 CIGS GCEX RELOR OWNER &
   /INDEX = xalpha COEF=alpha piter=0

**-------------print the coefficients---------------------
GEN1 sig=SQRT($SIG2)
PRINT sig
GEN1 NK = $K-2
STAT xalpha /MEAN=mu
SAMPLE 1 1
DISTRIB mu /TYPE=NORMAL PDF=pdf CDF=cdf
GEN1 lambda=pdf/cdf
PRINT lambda cdf
GEN1 adjust = 1 -lambda*(mu+lambda)
PRINT adjust

**-----------Estimate the partial effects ----------------
SAMPLE 1 NK
DELETE SKIP$
PRINT alpha
GENR beta = alpha*sig
PRINT beta
GENR tobit1 = beta*adjust
GENR my1=cdf*tobit1
GENR tobit2=pdf*alpha
GENR my2=beta*pdf*(mu+lambda)
GEN1 Eystar=sig*mu +sig*lambda
PRINT Eystar
GENR total = beta*cdf
PRINT tobit1 my1 tobit2 my2 total

**-------Estimate the elasticity effects ----------------
GEN1 Ey = sig*mu*cdf+sig*pdf
PRINT Ey
GENR e1=my1*X/Ey
GENR e2=my2*X/Ey
GENR etotal = total*X/Ey
PRINT e1 e2 etotal

STOP
```

Appendix II. Shazam program for LM heteroscedasticity test

```
**----------TOBIT REGRESSION ---------------------------
TOBIT LIP PROV1 PROV2 PROV3 PROV4 URBAN &
   NUMRM BOPOMO IAT MARRIED DIVORCED AGEGRP MALE EDUC1 &
   EDUC2 EDUC3 EDUC4 OCCU1 OCCU2 COB1 COB2 COB3 COB4 &
   SIAT SIZE NFTE1 NFTE2 CIGS GCEX RELOR OWNER &
   /INDEX=IND PREDICT=YHAT piter=0

**-------test heteroscedasticity ----------------
GEN1 sig2 = $SIG2
GEN1 sig = SQRT($SIG2)
DISTRIB IND /PDF=P CDF=C
GENR LAMBDA = P/(1-C)
GENR Z = LIP.GT.0
GENR E = LIP - YHAT
GENR A = Z*(E/sig2) + (1-Z)*(-LAMBDA/sig)
GENR X1 = (E**2/sig2) -1
GENR B = Z*(X1/(2*sig2)) + (1-Z)*(IND*LAMBDA/(2*sig2))
GENR ONES=1

MATRIX X =PROV1 |PROV2 |PROV3 | PROV4| URBAN| &
   NUMRM| BOPOMO| IAT| MARRIED|DIVORCED|AGEGRP| MALE| &
   EDUC1|EDUC2|EDUC3|EDUC4|OCCU1|OCCU2|COB1|COB2|COB3|COB4| &
   SIAT| SIZE|NFTE1| NFTE2 |CIGS| GCEX| RELOR| OWNER

MATRIX AX = A*X
MATRIX BX = B*X
MATRIX G = AX|B|BX
MATRIX LM = ONES'*G*INV(G'*G)*G'*ONES
DISTRIB LM/TYPE=CHI DF=30

STOP
```

Appendix III. Shazam program for LR heteroscedasticity test

```
**----------TOBIT REGRESSION ---------------------------
TOBIT LIP PROV1 PROV2 PROV3 PROV4 URBAN &
   NUMRM BOPOMO IAT MARRIED DIVORCED AGEGRP MALE EDUC1 &
   EDUC2 EDUC3 EDUC4 OCCU1 OCCU2 COB1 COB2 COB3 COB4 &
   SIAT SIZE NFTE1 NFTE2 OWNER &
   /INDEX=IND PREDICT=YHAT piter=0 COEF=ALPHA

* Get the regression coefficients
GEN1 SIGMA=SQRT($SIG2)
SAMPLE 1 28
GENR BHAT=ALPHA*SIGMA

* Now use the NL command for maximum likelihood estimation to
* replicate the results of the TOBIT command.
* First, use OLS to set some starting values for the estimation.
* Restrict the sample to those who work
sample 1 10417
skipif lip.eq.0

DIM BCON 29
OLS LIP PROV1 PROV2 PROV3 PROV4 URBAN &
   NUMRM BOPOMO IAT MARRIED DIVORCED AGEGRP MALE EDUC1 &
   EDUC2 EDUC3 EDUC4 OCCU1 OCCU2 COB1 COB2 COB3 COB4 &
   SIAT SIZE NFTE1 NFTE2 OWNER &
   / HETCOV COEF=BETA

GEN1 BCON:29=SQRT($SIG2)
delete skip$
GEN1 P=4210/8941
SAMPLE 1 28
GENR BCON=BETA/P
PRINT BHAT BETA BCON

* Define an equation in a SHAZAM character string.
* Description of character strings is in the chapter SHAZAM PROCEDURES
* in the SHAZAM User's Reference Manual.
* Note that the expression is enclosed by brackets -- this is to
* ensure correct evaluation of the expression when it is used later on
* the EQ command.
XB1:B1*PROV1+B2*PROV2+B3*PROV3+B4*PROV4+B5*URBAN+
XB2:B6*NUMRM+B7*BOPOMO+B8*IAT+B9*MARRIED+B10*DIVORCED+
XB3:B11*AGEGRP+B12*MALE+B13*EDUC1+B14*EDUC2+B15*EDUC3+
XB4:B16*EDUC4+B17*OCCU1+B18*OCCU2+B19*COB1+B20*COB2+
XB5:B21*COB3+B22*COB4+B23*SIAT+B24*SIZE+B25*NFTE1+B26*NFTE2+
   B30*OWNER+B0

SAMPLE 1 10417
skipif (prov.eq.00).or.(city.eq.0).or.(educ.eq.6).or.(occu.eq.11).or.(occu.eq.13).or.
```

```
(clten.eq.4)
GENR const=-LOG(2*$PI)
GENR LIMIT=DUM(LIP)

DIM BTOBIT 56
NL 1 / NCOEF=29 LOGDEN START=BCON COEF=BTOBIT
EQ (1-LIMIT)*LOG(1-NCDF(([XB1] [XB2] [XB3] [XB4][XB5]) /sig))
+LIMIT*(const-LOG(sig**2)-((LIP- ([XB1][XB2][XB3][XB4][XB5]))/sig)**2)/2
END

GEN1 LLF0=$LLF
SIG0:A0*(EXP(A1*PROV1+A2*PROV2+A3*PROV3+A4*PROV4+A5*URBAN+
SIG1:A6*NUMRM+A7*BOPOMO+A8*IAT+A9*MARRIED+A10*DIVORCED+
   A11*AGEGRP+A12*MALE+A13*EDUC1+
SIG2:A14*EDUC2+A15*EDUC3+A16*EDUC4+A17*OCCU1+A18*OCCU2+A19
   *COB1+A20*COB2+A21*COB3+A22*COB4+
SIG3:A23*SIAT+A24*SIZE+A25*NFTE1+A26*NFTE2+A30*OWNER)**(1/2))

* Estimation of the Tobit Model with Multiplicative Heteroskedasticity.
* The starting values are the coefficient estimates from the
* previous Tobit estimation - note that the value of the
* log-likelihood function at the first iteration is identical
* to the Tobit log-likelihood function value.
NL 1 / NCOEF=56 LOGDEN START=BTOBIT
EQ (1-LIMIT)*LOG(1-NCDF(([XB1][XB2][XB3][XB4][XB5])/([SIG0][SIG1]
   [SIG2][SIG3])))+ LIMIT*(const-LOG(([SIG0][SIG1][SIG2][SIG3])**2)-
((LIP-([XB1][XB2][XB3][XB4][XB5]))/([SIG0][SIG1][SIG2][SIG3]))**2)/2
END

GEN1 LLF1=$LLF

* Likelihood ratio test statistic for heteroskedasticity
GEN1 LR=2*(LLF1-LLF0)

* Calculate p-value for the test
SAMPLE 1 1

DISTRIB LR / TYPE=CHI DF=27 CDF=cdf
GEN1 p_value=1-cdf
PRINT LR p_value
STOP
```