THE MARGINAL PRICE INDEX METHOD: CONSTRUCTION FROM STATISTICS CANADA'S CANADIAN SURVEY OF FAMILY EXPENDITURE DATA A ROBUST APPROACH

by

Grant Cameron

Supervisor: Sydney Afriat
Defence Held: April 6, 1987

UNIVERSITÉ D'OTTAWA

FACULTÉ DES SCIENCES SOCIALES

DÉPARTEMENT DE SCIENCE ÉCONOMIQUE

UNIVERSITY OF OTTAWA

FACULTY OF SOCIAL SCIENCES

DEPARTMENT OF ECONOMICS

The Marginal Price Index Method: Construction from Statistics Canada's Canadian Survey of Family Expenditure Data - A Robust Approach

by

Grant J. Cameron

#277307

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: S.N. Afriat

ECO 7997

Ottawa, Ontario

March 16, 1987

Abstract

The objective of this work is to estimate the inflation map of the Canadian economy from 1969 to 1982. This map determines whether the Consumer Price Index is a good indicator of the inflationary effects felt by all consumers. The Marginal Price index is estimated to quantify the size of income inflation bias.

This paper has two major thrusts: price theory and a statistically robust method for matching data from two independent sources. The price theory comes from the work of Afriat and others, little new has been added here. The matching procedure is a new application in the field of economics. It is shown that a particular specification of the Marginal Price index not only measures the inflation bias but also determines the efficiency of the matching procedure.

Introduction

Measuring the redistributional effects of inter-temporal price changes of goods and services in the economy is an important social objective. If one were to find that instruments of maintaining purchasing power parity from one time period to another (such as cost of living allowances in contractual arrangements) did not have an equivalent impact on all individuals in society, then the society itself may undergo some structural changes. The redistribution effect, which has been called Income Inflation Bias[1], could alter the effectiveness of income equalization programs like the U.I. scheme.

The objective of index numbers over the past 300 years has been the measurement of purchasing power parity at the micro level. As Samuelson[1983, page 146] describes in his Foundations text: " *The fundamental problem on which all analysis rests is that of determining merely from price and quantity data which of two situations is higher up on the individual preference scale.* " Premises such as time invariant utility and the existence of a representative consumer are used by economists to define a more precise region in which the agent's true preference lies. Samuelson's Revealed Preference Theory was formulated to explain portions of a consumers "region of ignorance" (a set in commodity space of which one was not able to reduce by shifts in the budget line), a problem which was first exposed by index number theorists. However, little attention has been paid to the consequences of representing the collection of all individual price indexes (the inflation map) with a single number, namely the Consumer Price Index.

The Marginal Price Index Method outlined in Afriat [1974] presents a way of exposing possible biases in the C.P.I. estimate via the inflation map. Historically, economists required longitudinal microdata or "panel data" to construct this map[2]. This paper employs a statistically robust[3] technique for matching observations from two independent surveys of consumer costs, namely the 1969 and 1982 Canadian Survey of Family Expenditures databases, whose measure of efficiency is closely tied to a linear specification of the Marginal Price Index. Therefore, the conceptualization of price changes faced by society as a bivariate distribution is possible. This is preferred to the single numbers achieved through price index construction.

The structure of this paper is as follows. The first section gives a brief theoretical overview of price indexes in general. The second portion discusses the Marginal Price Index Method in some detail. The nature of non-parametric statistics follows with an analysis of the matching procedure. The paper concludes with results and possible extensions of the work.

Theoretical Overview

(i) General Concept and Notation

A price index can be described as a measure of price changes faced by one or more consumers from one time period to another. Consider a row vector of prices for m commodities and a corresponding column vector of commodities purchased. The product of these two vectors yields a scalar describing the total cost of the consumer's aggregate consumption. Once one has obtained a set of these vectors for different time periods then one may begin the construction of price indexes.

The general form of the price index could be expressed as follows:

let $p_t = (p_t^1, p_t^2, \ldots , p_t^m)$ -> price vector at time t

let $x_t = (x_t^1, x_t^2, \ldots , x_t^m)^T$ -> quantity vector time t

and $c_t = p_t * x_t$ -> current consumption

Introducing time subscripts for the periods t=0,1 would render a possible index form:

$$I_{10} = c_1 / c_0$$

Now that the basic ideas of index number construction have been discussed (and this papers notation has been described) a brief historical perspective will be pursued.

(ii) Historical Developments

The construction of price indexes has been underway since the time of Fleetwood who measured the price differential of a bundle of goods consumed by a typical university student in 1707 relative to the same bundle consumed 200 years previous. This simple measure began the long history of index number usage.

It was not until the 1920s that the index number problem underwent a thorough review from two different aspects. Fisher[1922] published his renowned book in which comparisons of a variety of index formulations could be made through a specific set of mathematical tests. These axioms of index refinement rendered an easily constructed *ideal index* which adhered to his specifications. The *ideal index*, the geometric mean of the commonplace Paasche and Laspeyres indexes, became an economic convention.

Konyus[1924], Byushgens[1926], and others approached the index number question from a theoretical point of view which relied upon the concept of ordinal utility. The Konyus *true cost* of living index is described as follows:

The true cost of living index is the scalar which measures the minimum cost a consumer must pay to obtain a reference utility level at time 1 relative to the cost faced at time 0 required to achieve the same reference utility level.

The assumption that all consumers have the identical family of utility maps is also necessary for aggregate price indexes to have significance. This implies that utility is homothetic so proportionate increases in prices and incomes do not change the expenditure shares of the consumer. Deaton and Muellbauer [1980, page 170] express this index as:

$$\hat{I}_{10}(p_1, p_0, x) = c(u^R(x) , p_1) / c(u^{R'}(x) , p_0)$$

where $p_1, p_0 > 0$ are prices vectors at times 1,0

$u^R = u^R(x)$ is a reference indifference level for some commodity vector $x > 0$.

c some general form of the consumer's cost function.

Effort has been spent refining index number construction as to improve the specification of the utility reference. Two particularly important theorems are as follows[4]:

1)(Konyus) If $u^R$is continuous from above and the formula for $\hat{I}_{10}$ holds then:

$$\hat{I}_{10}(p_1, p_0, x_0) \leq p_1{*}\, x_0 / p_0{*}\, x_0 = \text{Laspeyres Index}$$

and

$$\hat{I}_{10}(p_1, p_1, x_1) \geq p_1{*}\, x_1 / p_0{*}\, x_1 = \text{Paasche Index}$$

2)(Frisch) If in addition, u^R is homothetic, then:

$$\text{Paasche Index} = p_1{*}\, x_1 / p_0{*}\, x_1 \leq \hat{I}_{10}(p_1, p_0, x)$$
$$\leq p_1{*}\, x_0 / p_0{*}\, x_0 = \text{Laspeyres Index}$$

Theorem 2) states that the Laspeyres and Paasche indexes provide upper and lower bounds for *the* price index if the utility map is homothetic. One cannot improve upon these bounds without making some specific assumptions about the functional form of the reference utility surface u^R.

There exists evidence to reject the assumption of homothetic utility. Afriat[1977] cites a study by the National Council of Welfare which states, among other things, that families in 1969 with less than $3,000 spent 27.9% of their income on food while those with over $15,000 spent 13.4%. Given these results Afriat suggests that " *It is necessary to go back to the original detailed, disaggregated data and start again.* " The estimation of the inflation map, the basis from which the Marginal Price Index method is constructed, seems a logical way of returning to this level of disaggregation.

Inflation Maps and the Marginal Price Index

(i) The General Method

The total cost for the ith individual to attain some utility level at time period t has been defined as:

$$c_{it} = p_{it} * x_{it}$$

Mappings of two c_t's of different time periods could be constructed and represented as:

$$e_{i10} = \{ c_{i1} , c_{i0} \}$$

If the ith consumer is considered representative of some subset of the population, then this mapping would be entirely analogous to the concept first used by Fleetwood which was described earlier.

The collection of Fleetwood points for all N consumers gives a bivariate distribution of price change effects for the society. Assuming that all individuals have non-negative aggregate consumption at every point in time then the entire collection of points would lie in the positive orthant. The nature of the inflationary impact could be observed by examining the type of dispersion the distribution would exhibit. By calculating the mean point of the collection as:

$$\bar{e}_{10} = \{ \bar{c}_1 , \bar{c}_0 \}$$

where $$\bar{c}_t = \frac{1}{N} \sum_{i=1}^{N} \{ c_{it} \}$$

one could define the AVERAGE PRICE INDEX as:

$$API_{10} = \bar{c}_1 / \bar{c}_0$$

which is geometrically described as the ray passing through the origin and the point $\bar{e}_{10}$. If utility is homothetic then every point in the inflation map will lie along this ray. The ray is theoretically equivalent to the CONSUMER PRICE INDEX.

One can define the MARGINAL PRICE INDEX as:

$$MPI_{i10} = c_{i1} / c_{i0}$$

for the ith individual. In the context of purchasing power equivalency the definition becomes:

$$c_{i1} = MPI_{i10} * c_{i0}$$

From this construction:

$$API_{10} = \bar{MPI}_{10} = \frac{1}{N} \sum_{i=1}^{N} \{ MPI_{i10} \}$$

If one ranked the set of e_{10}'s with respect to income levels at time 0 then the collection of MPI_{i10}s in this set might approximate some functional form. If this form was significantly different from the ray described by the API point then, inflationary bias would exist between the two time periods under consideration.

(ii) The Original Application

Afriat & Kuiper[1975] imposed the further restriction of linearity on the ordered data by bisecting the data into k-classes and calculating the corresponding set of k means. Assuring that k was always of the order 2^n the method could be iterated to render three mean estimates: one for the population mean; one for the class of individuals with incomes less than the population mean; one for the class with incomes higher than the population mean. Their estimation of the MPI was derived by joining these three mean estimates together. One could then compute slope and intercept parameters for this function.

The authors then calculated, from the mean of the entire population, the API curve as it is described above. By calculating the slope of this ray some estimates of income inflationary bias could be achieved. Mathematically:

bias = {slope of MPI ray} / {slope of API ray}

therefore,

bias < 1 => income re-dist to high income groups

bias > 1 => income re-dist to low income groups

bias = 1 => no redistribution effect

The authors observed that the bias favors the lower income groups in some periods and favors higher income groups in others.

(iii) Theoretical Linkages

In order to conceive of price changes in the context of the inflation map one requires evidence of consumer utility maximization. Without confirmation of this hypothesis this analysis is weakened.

Such evidence was presented by Varian[1982] who by applying the results of Afriat's Theorem[5] found that finite sets of price/quantity observations were consistent with preference maximization. Varian concludes his study of the hypothesis by saying " *The observation [the data are consistent with the maximization hypothesis] implies that those studies which have rejected the preference maximization using conventional parametric techniques are rejecting only their particular choice of parametric form.* "

Given this testimony, one can now proceed with confidence that the observations of the distributions will bear results equivalent with that of traditional index number theory. However, the construction of reliable price/quantity observations from two independent surveys for two distinct time periods remains a problem. The following section examines this issue fully.

Robust Matching

Before outlining the method of matching cases from the two independent samples one should be familiar with the contents of the survey data.

The major objective of the Statistics Canada " Canadian Survey of Family Expenditures "[1983a] is to estimate changes in household expenditure patterns at a very fine level of detail. Its findings contribute greatly to the composition of the Consumer Price Index commodity basket which directly affects the official measure of our inflation rate.

The survey is conducted approximately every two years and draws respondents from both urban and rural settings all across Canada[6]. These respondents, which are equivalent to economic households, are drawn to reflect particular income, geographical, and demographic compositions for which they are suitable proxies for a larger class of Canadian households. On the basis of this information, participants are issued a survey weight which reflects the proportion of the Canadian population their particular household represents (this weight will play a large role in the observation matching process described below).

Each household is asked to itemize its expenditure on an extremely detailed vector of goods and service commodities so the sample is rotated to ensure no respondent completes a questionnaire in two consecutive surveys.

The objective of the matching technique is two-fold. One is interested in matching observations from two independent samples which, when combined, approximate the inflationary effect experienced by a hypothetical family unit of a certain income order. Secondly, the technique must maximize the use of the data in both surveys. The method described below achieves both these objectives.

Consider a random variable X drawn from some empirical mass function f(x) where x is the finite set of all X. One can construct the respective empirical distribution function as:

$$F(x \leq X) = \int_0^X f(x)\, dx$$

Parzen[1979] defined a Quantile function, Q(u), which is intrinsically related to this e.d.f. in the following fashion:

$$Q(u) = F^{-1}(u) = \inf\{ x : F(x) \geq u , 0 \leq u \leq 1 \}$$

From this construction fall two interesting identities:

Correspondence Identity: $F(x) \geq u \iff Q(u) \leq x$

Inverse Identity: $F(Q(u)) = u$

The sample quantile function, $\tilde{Q}(u)$, can be directly calculated explicitly in terms of the order statistics $X_{(1)} < X_{(2)} < X_{(3)} < \ldots < X_{(n)}$. The function fundamentally relates the position of these ordered statistics with respect to the form of F(x). For example,

$$\tilde{Q}(u) = X_{(j)}, \quad (j-1)/n < u \leq (j/n)$$

From this, $\tilde{Q}(0.5)$ is defined to be the sample median,

$$\begin{aligned} \tilde{Q}(0.5) &= X_{(m+1)} && \text{if } n = 2m + 1 \text{ is odd} \\ &= \tfrac{1}{2}\,(X_{(m)} + X_{(m+1)}) && \text{if } n = 2m \text{ is even} \end{aligned}$$

The quantile functions for the two samples become the instruments with which consumption observations were matched. They were constructed as follows.

Let the ordered set of sampled households by their total income before taxes from lowest to highest be denoted by x. The e.m.f. value for any particular observation in the set $f(X_{(i)})$ is represented by the households sample survey weight. Therefore, for all of x we have defined a respective empirical mass function f(x).

One can construct an analogous e.m.f by using the set of normalized survey weights. The normalized values are calculated by dividing each observation by the sum of the weights.

The new set of mass observations g(x) is defined as follows:

$$g(\ X_{(i)}) = f(\ X_{(i)}) \ / \ \sum_{i=1}^{N} f(\ X_{(i)}) \qquad i = 1, \ldots, N$$

$$g(x) = \{\ g(\ X_{(1)}), \ \ldots \ , g(\ X_{(N)})\ \}$$

The e.d.f of g(x), G(x), is calculated as follows:

$$G(\ x \le X_{(i)}) = \sum_{i=1}^{X_{(i)}} g(\ x\ ) \qquad 0 \le G(x) \le 1$$

From G(x) we can easily calculate its quantile function $\tilde{Q}(u)$.

$$\tilde{Q}(u) = G^{-1}(u)\ , \qquad 0 \le u \le 1$$

The sample matching was carried out based on the following formula. (Note that the sample quantile functions now carry subscripts to indicate the survey year.)

$$X_{82(i)} = \tilde{Q}_{82}(\ u_i) \ \ \textit{matched with} \ \ X_{69(j)} = \tilde{Q}_{69}(\ u_j) \ \ \textit{if}:$$

$$u_{j-1} < u_i < u_j \qquad \textit{and} \qquad u_j - u_i < u_i - u_{j-1}$$

where: $i = 1, \ldots, N$ (number of observations in 1982 survey)
$j = 1, \ldots, M$ (number of observations in 1969 survey)
$M > N$

The premise of the algorithm is to match each case from the 1982 sample with the observation in 1969 which most closely approximates its position on the quantile function. This method is entirely robust since the quantile functions need not share any common characteristics with the exception of the cumulative property and their equivalent ranges. Each function could be of an entirely different distribution family and the matching still takes place. Therefore, one is able to match cases from surveys of great size discrepancy.

Before analyzing the sample quantile functions some points are worth mentioning. First, this procedure will produce N pairs of matched observations and secondly, a one-sided test is sufficient in the matching algorithm since both quantile functions have the property of being cumulative since the bounds of $\tilde{Q}_{82}(u)$ and $\tilde{Q}_{69}(u)$ both span the range 0 through 1 in u.

Finally, matching solely as a function of the income order and survey weight becomes more preferable as the two time periods are further distanced. This is due to inequalities of identical descriptive statistic categorizations. For example, white collar employment in the early 1960s may not occupy the same relative position in the income order as its mid-1980s counterpart. Conversely, as the time interval shortens, it may be better to partition the data by some subset of the descriptors and proceed with the matching within these classes. Therefore, introducing a priori restrictions before matching becomes dependent on the time interval under consideration.

Sample Quantile Function Analysis

For the matching algorithm to be meaningful it must be shown that the two independent samples are from same the location/scale family. Therefore, a method must be devised from which one can determine whether these two distributions are equivalent in the first two moments. The following derivation shows how a linear specification for the MPI is analogous for this necessary test.

From theoretical statistics we know that a bivariate distribution can be expressed entirely by its two marginal distributions and the correlation coefficient. To put the problem into a regression perspective define two standardized random variables Z_1 and Z_2 of X and Y such that:

$$Z_1 = ((X - \mu_X)/ \sigma_X) \qquad Z_2 = ((Y - \mu_Y)/ \sigma_Y)$$

where: μ_X μ_Y are the population means of distributions X and Y

σ_X σ_Y are the population standard deviations of X and Y

We can now describe the general regression of Y on X as:

$$E[Z_1 \mid Z_2] = E[Y \mid X] - \mu_Y$$
$$= p(\sigma_Y/ \sigma_X)(X - \mu_X)$$

Kendall and Stuart[1967] describe this in the context of the normal family, however this specification is quite appropriate for the analysis of quantile functions.

One can define the conditional probability of the ith quantile observation in 1969 to be equivalent in order to the jth quantile in 1982 in the following manner:

$$Q_{82}(u_j) = p * Q_{69}(u_i)$$

where p represents the conditional autocorrelation coefficient[7].

By standardizing the quantiles we can derive an analogous general regression equation:

$$Q_{82}(u) = (\mu_{82} - p * \mu_{69}) + p * (\sigma_{82} / \sigma_{69}) Q_{69}(u)$$

Since we are dealing with order statistics then, by definition, p=1. Therefore, the sample regression simplifies to:

$$\tilde{Q}_{82}(u) = (\mu_{82} - \mu_{69}) + (\sigma_{82} / \sigma_{69}) \tilde{Q}_{69}(u)$$

or

$$\tilde{Q}_{82}(u) = d + s * \tilde{Q}_{69}(u) \quad \blacksquare \quad X_{82(i)} = d + s * X_{69(i)} \quad (**)$$

where d -> the shift parameter
s -> the relative scale parameter

Now, if the relation (**) is indeed linear, then this is both a necessary and sufficient condition for the two quantile functions to be of the same location/scale family. In other words, if relation (**) is found to be linear then one can feel confident about the soundness of the results which are derived from the synthesized bivariate distribution.

The linear formulation of the MPI[8]

$$c_{82} = a + b * c_{69}$$

is analogous to relation (**) under the assumption of a linear relationship existing between consumption and incomes[9]. The pairs of consumption observations are matched via the properties of their respective quantile functions through the implicit correlation of their income order statistics[10]. Therefore, the estimation of the MPI in this form provides a method of identifying location/scale equivalence. In the next section the existence of this important linear property is determined.

Analysis: (1) Data Modifications

Before beginning the regression analysis some deletions to the sample were affected. The Households whose head was less than 25 years old were dropped from the sample because they form and reform under different descriptive categories rapidly. Households not reporting their education or occupation status were deleted because it was preferable to work with cases in which all pertinent information existed. In all, less than 70 cases were deleted for these reasons.

Other sample cases, whose information vector was intact, were deleted because their consumption pattern were outliers. If these observations had not been dropped, then the traditional forms of regression analysis used to estimate the MPI would have been biased[11]. All of these cases tended to be of a extremely high income/consumption order. Table (a) presents the marginal distributions which give an idea of the nature of the problem.

Based on these tables it was decided to delete cases in which the Total Current Consumption in 1969 was greater than $25,000 or greater than $70,000 in 1982. This lead to a further reduction of only 36 cases from the sample as some matched cases possessed both of the above characteristics. The final sample size, on which all subsequent analysis is based, is 9976 cases.

Table (a): Post Matching Marginal Frequencies

(1) Total Current Consumption - 1969

Income Class($)	Frequency	Percentage	Cumulative %age
0 - 5,000	3569	35.6	35.6
5,001 - 10,000	4977	49.7	85.3
10,001 - 15,000	1199	12.0	97.3
15,001 - 20,000	220	2.2	99.5
20,001 - 25,000	30	0.3	99.8
25,001 - 90,000	24	1.0	100.0

(2) Total Current Consumption - 1982

Income Class($)	Frequency	Percentage	Cumulative %age
0 - 10,000	1767	17.6	17.6
10,001 - 20,000	3550	35.4	53.0
20,001 - 30,000	2899	28.9	82.0
30,001 - 40,000	1228	12.2	94.2
40,001 - 50,000	409	4.1	98.3
50,001 - 60,000	97	1.0	99.3
60,001 - 70,000	37	0.3	99.6
70,001 - 80,000	20	0.2	99.8
80,001 - 210,000	12	0.2	100.0

Analysis: (ii) Testing for Non-linearity

The first and simplest test for non-linearity is to visually examine the properties of the inflation map for all cases. This distribution, shown as figure (i), indicates a "cone" shape for the map implying heteroscedasticity. Also, a disproportionate number of cases with low joint consumption levels are apparent. Given these two properties it would seem inappropriate to apply ordinary least squares to the map to estimate the MPI. However, a correlation between total current consumption in 1969 to that in 1982 clearly exists so one should be able to construct some linear form of regression analysis which will capture the essence of the MPI relationship.

Two sets of regression tests were performed on the data to further substantiate the claim of OLS impracticality. The first test was to run the regression equation over progressively larger portions of the sample and check whether the parameter coefficients varied as more of the sample was brought into the analysis. Table (b) presents the results of these tests. The second test was to partition the observations into three equal groups, run identical regressions, and observe differences in the parameters. The results of this test are located in table (c).

FIGURE (1): INFLATION MAPS – TOTAL CURRENT CONSUMPTION

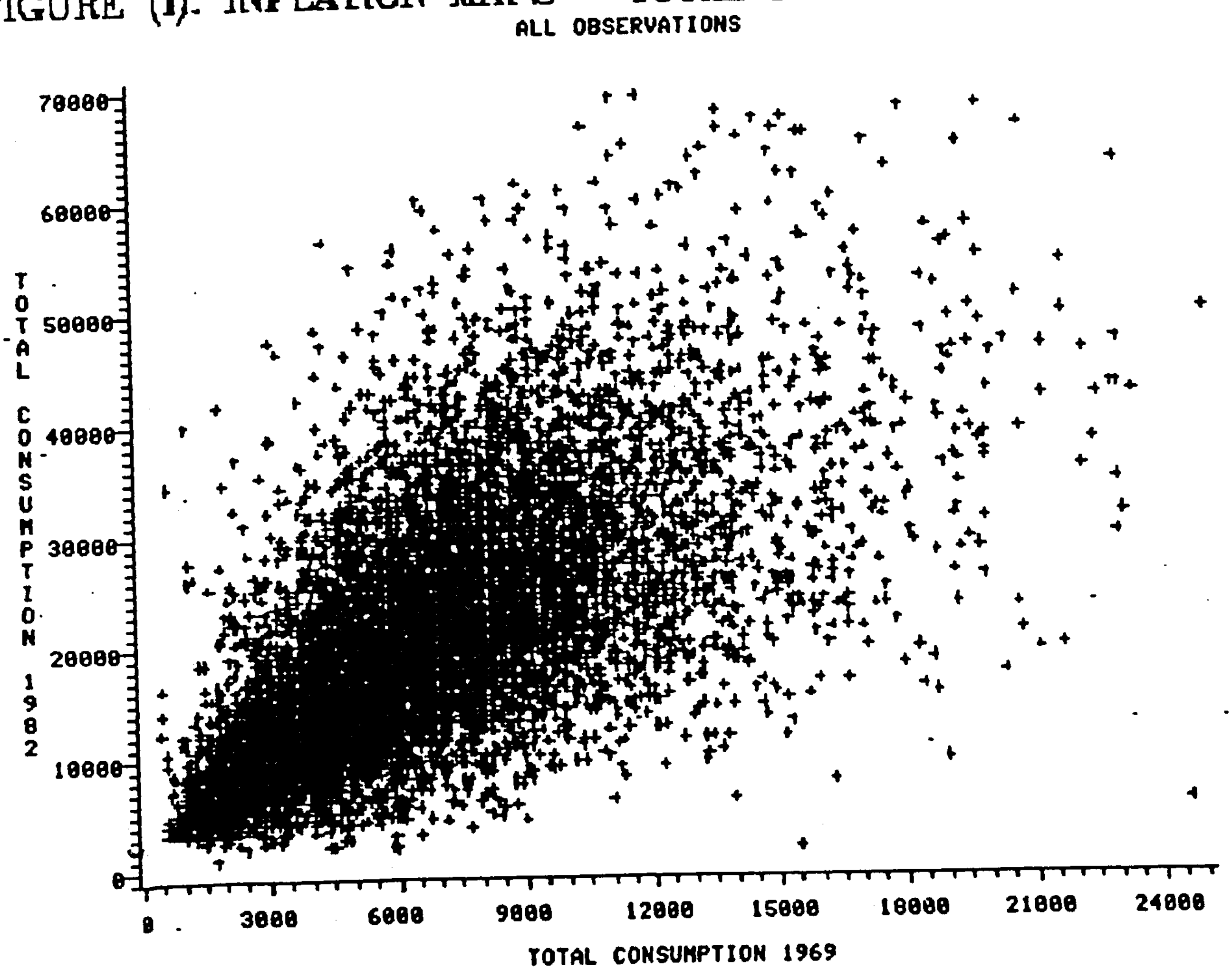


Table (b): Results for Progressive Regression Test

The values in the ( ) and < > indicate the standard error and t statistic of the coefficient respectively.

Sample Size	*Intercept*	*Slope*	*Rsquare*
1st 2500 cases	6399.3 (176.5) < 36.3 >	0.996 (0.053) < 18.5 >	0.1211
1st 5000 cases	6446.0 (165.8) < 38.9 >	1.574 (0.035) < 45.1 >	0.2894
1st 7500 cases	6466.1 (166.2) < 38.9 >	1.891 (0.028) < 67.2 >	0.3759
All 9976 cases	6780.5 (170.0) < 39.9 >	2.077 (0.023) < 91.0 >	0.4541

Table (c): Results of Partitioned Regression Test

Sample Size	*Intercept*	*Slope*	*Rsquare*
1st 3300 cases	6324.9 (171.8) < 36.8 >	1.262 (0.045) < 27.9 >	0.1911
middle 3300	16891.1 (362.9) < 46.5 >	0.4797 (0.054) < 9.0 >	0.0237
final 3376	22650.4 (515.4) < 43.9 >	0.802 (0.050) < 16.0 >	0.0701

One can conclude from these checks that this type of regression analysis will render meaningless results since tables (b) and (c) show continually changing values of the slope parameters and the Rsquare statistics. Further examination of figure (i) indicates that the third quartile of the distribution displays no real trend at all when examined independently of the other cases in the inflation map. Since so many data assumptions regarding the use of OLS have been violated these results cannot be used as evidence of non-linearity in the data.

Since OLS is an inappropriate form of estimation of the MPI some other method was required which would reduce the heteroscedasticity problem and re-weight the data such that no disproportions in representation would exist. This objective was achieved by taking a subsample of the 9976 cases which did not express the drawbacks of the entire set of points, yet maintained the underlying relationship.

To remove the growth in the spread of the independent variable the subsample would have to be a product of some form of averaging. Furthermore, to render a proportionate representation, this new set of points would have to be re-drawn based on some equitable distribution of the 1969 quantile function. Both these goals were achieved by constructing pairs of consumption observations for the two years in the following two step method.

The first step was to find the median value for equal areas lying under the quantile function of the independent variable for the regression (in this analysis 1969). Approximately 500 observations were to be used in the estimation so the interval chosen to increment along the quantile function was 0.0015. The 1969 observation of the pair can be described as:

$$\tilde{c}_{i69} = \text{MEDIAN}\{ c^1_{i69}, c^2_{i69}, \ldots , c^k_{i69}\} \quad i = 1, \ldots , 580$$

where: $u_i - u_{i+d} \Rightarrow \{ c^1_{i69}, c^2_{i69}, \ldots , c^k_{i69}\}$ $d=.0015$

Note that k can vary in size from one bisection of the quantile function to the next. Unevenness in the normalized sample weights rendered 580 observations and not the expected 667 (1.0/0.0015).

The second step was to match a representative 1982 consumption figure for each of the medians calculated above. For each set of k observations for 1969 associated with the ith quantile bisection there exists a median corresponding to the set of k 1982 consumption observations. This value is the conditional median of 1982 on the 1969 bisections. More succinctly:

$$\{ c^1_{i69}, c^2_{i69}, \ldots , c^k_{i69}\} \Leftrightarrow \{ c^1_{i82}, c^2_{i82}, \ldots , c^k_{i82}\}$$

for all $i = 1, \ldots , 580$

$$\tilde{c}_{i82} = \text{MEDIAN}\{ c^1_{i82}, c^2_{i82}, \ldots , c^k_{i82}\} \quad i = 1, \ldots , 580$$

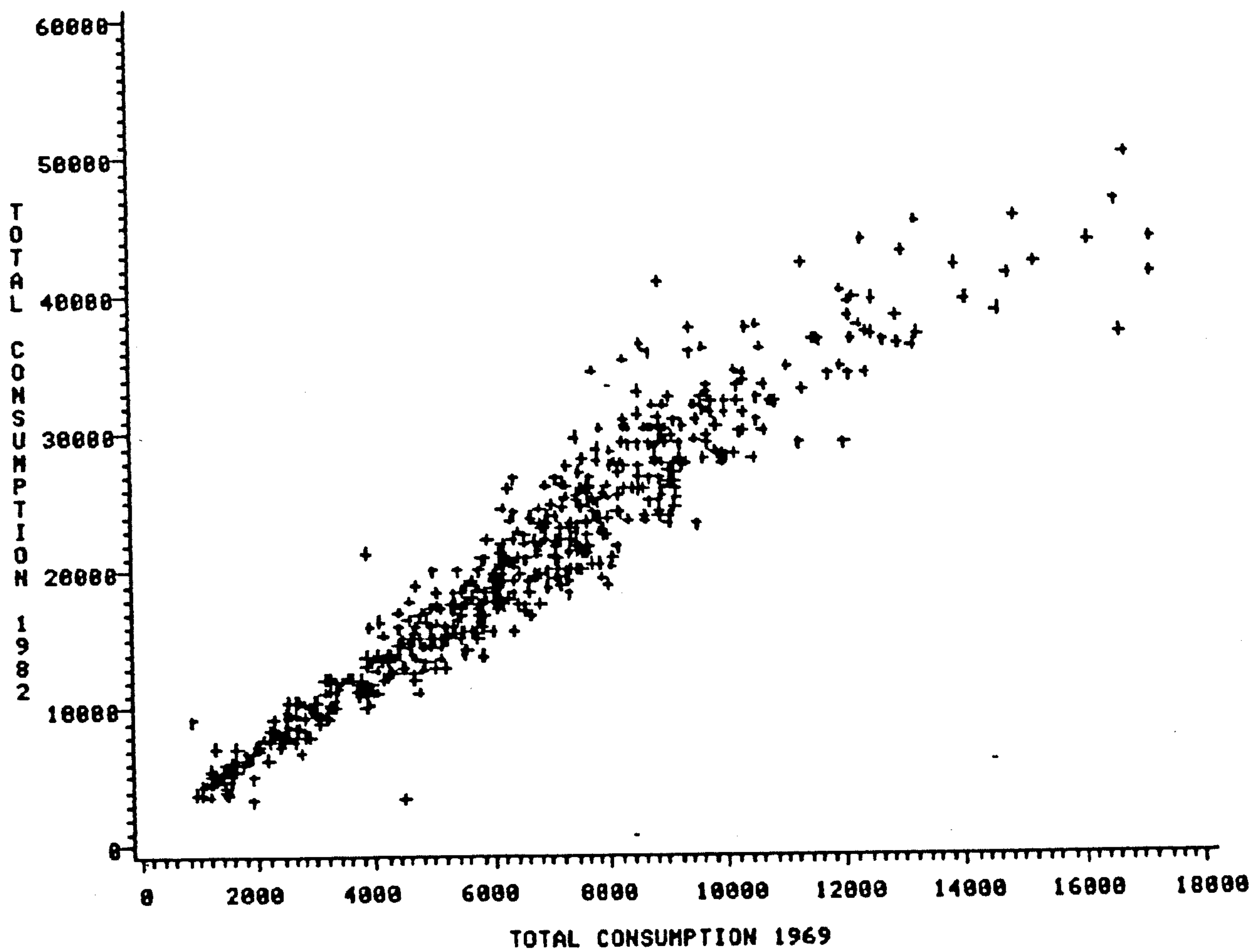
Figure (ii) shows the resulting subsample inflation map where the heteroscedastic phenomenon has been reduced and the points are equivalent with respect to the original survey weight distribution. A truly linear relation is present so one can express the regression in terms of the MPI as:

$$\tilde{c}_{i82} = a + b * \tilde{c}_{i69}$$

Figure (ii) serves the additional purpose of being a proxy for the probability plots described by Hoaglin[1985] which can be used to examine relative differences in the shape of the two quantile functions. Hoaglin states that if the two quantile functions are identical in all respects then their corresponding scatter plot should be linear with a slope of one and an intercept of zero. Figure (ii) is not quite analogous to the plots described above since the consumption estimates are a function of the original order statistic (before-tax income). Nonetheless, it does not reveal any obvious non-linearities[12]. From this one can conclude that the two quantile functions are indeed from the same location/scale family.

FIGURE (II): INFLATION MAPS – TOTAL CURRENT CONSUMPTION

MEDIAN SCATTER PLOT



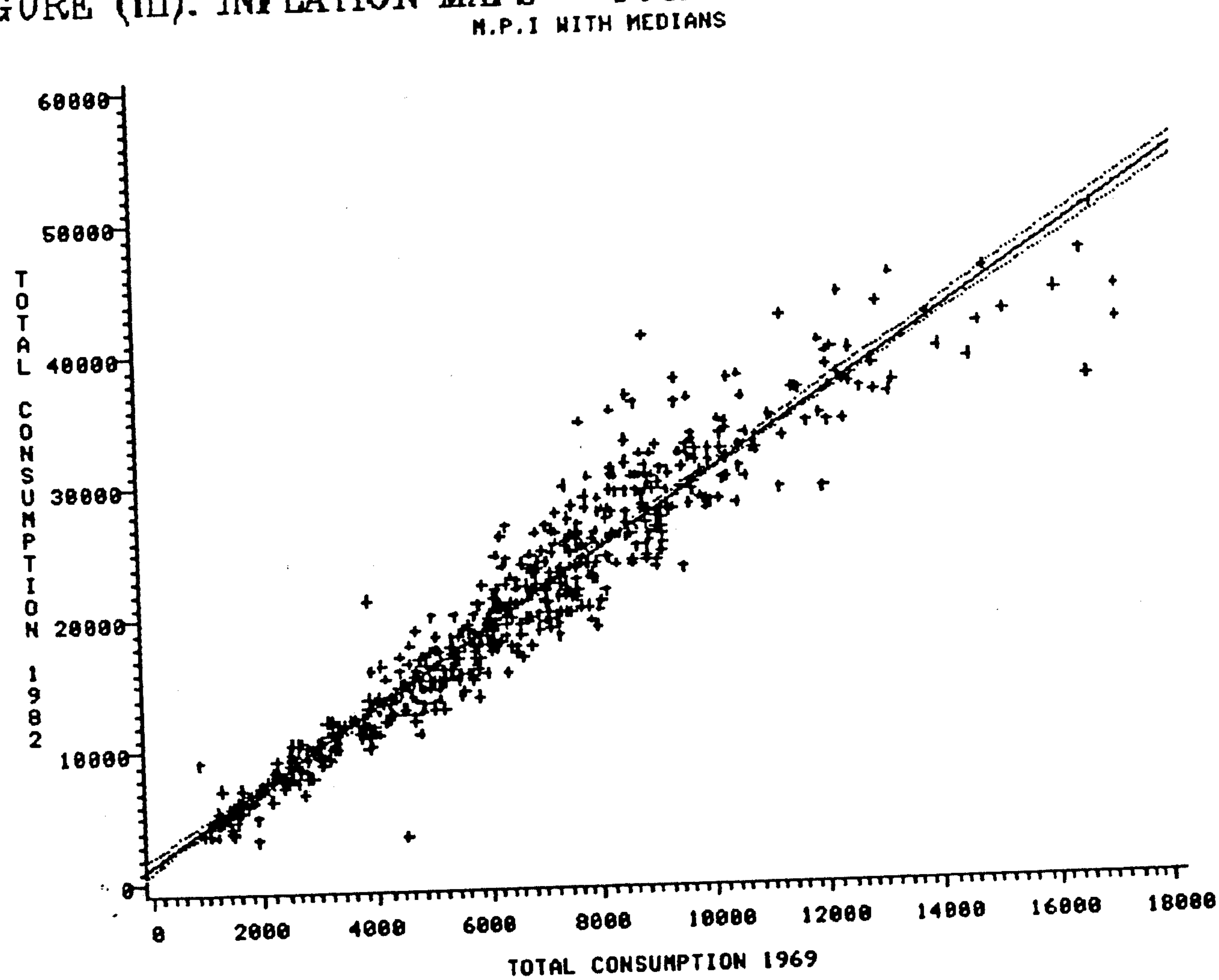
The Results and Conclusions

The subsample, shown in Figure (ii), confirms that a linear specification for the MPI, allowing easy estimation of this relationship, is sensible. The results of this regression renders a much stronger statistical result than that of the original (see table (b)). Not only have the parameters stayed significant at the 1% level but the Rsquare statistic has vastly improved. The following table presents the parameter estimates of the MPI along with the findings for the API and CPI[13]. Figure (iii) shows the MPI estimation with the subsample inflation map.

Index	*Intercept*	*Slope*	*Rsquare*
CPI	0	2.79	n.a.
API	0	3.11	n.a.
MPI	1457.6 (249.4)	2.929 (0.035)	0.924

A most interesting result lies in the discrepancy between the CPI estimate and its theoretical equivalent, the API. This spread suggests that the urban samples of the FAMEX, on which this CPI result is partially based, is not sufficient in providing accurate information regarding the inflationary experiences of all Canadians. Perhaps the CPI could be improved by expanding its coverage to some rural areas. However, recalling the shape of the bivariate inflation map (figure (i)) it is clear that any single number can not hope to represent the

FIGURE (III): INFLATION MAPS - TOTAL CURRENT CONSUMPTION



heterogeneous nature of the set of inflationary experiences the Canadian consumers face from one time period to the next.

Comparison of the slope estimates of the MPI and API indicate an inflation bias in favor of the high income classes for the time period under examination. Conversely, when comparing the value for the MPI with the CPI, one concludes that the inflation bias is in favor of the lower income cases. However, given the length of time between the two survey years, the magnitude of either measure of inflation bias is fairly small.

An final conclusion of this paper springs from the change observed in the spread of cases as one tends towards the high end of the income scale. This infers a relative lack of variability in poorer households when compared to more affluent households.

Possible Extensions of this Analysis

A subsequent work which measures the stability of these results is most important. Through re-estimation of all three indexes given different time intervals one could identify whether effects such as the business cycle change the conclusions significantly. One could also experiment with modifications to the matching procedure through the a priori restrictions discussed earlier. Does collapsing the time increment while partitioning the data change the index values significantly?

A work related to the first and of interest to analysts of the life cycle model is can be conceptualized. Researchers often construct hypothetical panel data for life cycle analysis so a test of which descriptive statistics are important would be of great use. The robust matching technique would enable one to examine the two independent sets of these statistics which are implicitly joined. Those characteristics which were found to be identical could then be used to partition the data.

Examining the relationship between the marginal distributions through scatter plot techniques for different sets of years is another possible work. This type of analysis would enable researchers to explore potential relative changes in skewness, modality, and kurtosis giving insight into distributional income dynamics.

Finally, verification of some sub-index consumption theories is possible from the matching technique. Starting from the premise of location/scale equivalence one could construct scatter plots of consumption components. This could help to establish whether a parametric representation between components and the income level of the household is possible.

Footnotes

1 Afriat and Kuiper[1975] may have coined this phrase.

2 Panel data studies, since they track a particular agent over time, have their own set of inherent biases when measuring phenomenon such as the inflationary effect. These biases arise from changes in the descriptive nature of the agent from one period to the next. It is not clear that these biases are smaller than those encountered through the matching of hypothetical income order-equivalent cases.

3 In statistics the term robust means relatively insensitive to distributional assumptions.

4 These proofs are in the form found in Diewert[1979] pp 10-11.

5 The following two results from Afriat's theorem[1967] enabled Varian to directly test the consumer maximization model:

i) The data implies "cyclical consistency" i.e.

$p_r x_r \geq p_r x_s,\ p_s x_s \geq p_s x_t,\ \ldots,\ p_q x_q \geq p_q x_r$ implies
$p_r x_r = p_r x_s,\ p_s x_s = p_s x_t,\ \ldots,\ p_q x_q = p_q x_r$

ii) There exist numbers $U_i, W_i > 0,\ i = 1, \ldots, n$ such that
$U_i \leq U_j + W_j^* p_j(x_i - x_j)$ for $j = 1, \ldots, n$

6 In some years the FAMEX surveys cover urban areas only. The 1969 and 1982 surveys are however both urban and rural in scope. The 1982 survey being the most current at the time of this study.

7 When the regression involves truly identical cases then the autocorrelation coefficient is a index of non-mobility i.e. it represents the conditional change in location of a particular observation. In the case of our hypothetical construction of points its meaning becomes unclear.

8 It is conceivable that an economic interpretation of the parameters in this equation which are linked to the relative location scale parameters of the quantile formulation is possible.

9 There exists a wide body of theoretical work and supporting econometric evidence pointing to this linear relationship.

10 Let the consumption functions be defined as follows:

$$c_{69(i)} = a_0 + b_0 * X_{69(i)} \qquad c_{82(i)} = a_1 + b_1 * X_{82(i)}$$

or

$$c_{69(i)} = a_0 + b_0 * \tilde{Q}_{69}(u) \qquad c_{82(i)} = a_1 + b_1 * \tilde{Q}_{82}(u)$$

by substitution relation (**) becomes,

$$c_{69(i)} = \hat{a} + \hat{b} * c_{69(i)}$$

where $\hat{a} = (b_1 / b_0)[s * b_0 - d * a_0) - a_1]$

$$\hat{b} = (b_1 / b_0) * d$$

11 A large bias in estimates is introduced when using OLS given a number of outliers in the sample. This is due to the greater than unity weight given outliers under the OLS method.

12 If fig(ii) showed signs of non-linear behaviour it could be due to poor location/scale equivalence or a different specification for the model relating consumption and income. This should be kept in mind when considering scatter plots of sub-index items such as food, recreation, or transportation. Non-linear plots in these instances do not necessarily imply poor location/scale equivalence.

13 The CPI value given here has been taken directly from the CANSIM database at Statistics Canada. It can also be calculated from the publication listed in the references([1983b]).

The API value comes from dividing the average value of consumption in 1982 by the average value of 1969. The averages come from the entire population of points (9976) and not the median subsample.

References

Afriat, S.N. [1967]: "The Construction of a Utility Function from Expenditure Data", International Economic Review 8, 67-77.

Afriat, S.N. [1974]: "The Marginal Price Index Method (I and II)", Research Paper # 7415, Department of Economics, University of Ottawa, Ottawa, Canada.

Afriat, S.N. and J. Kuiper [1975]: "Inflation Maps and the Total and Incremental Inflation Rates", Third World Congress of the Econometric Society, Toronto, Canada, Aug.

Afriat, S.N. [1977]: The Price Index, (London: Cambridge University Press).

Byushgens, S.S. [1926]: "Mathematischeskii Sbornik", Recueil Mathematique 32, 625-631 (translation).

Deaton, A. and J. Muellbauer [1980]: Economics and Consumer Behavior, (London: Cambridge University Press).

Diewert, W.E. [1979]: "The Economic Theory of Index Numbers: A Survey", Discussion Paper # 79-09, Department of Economics, University of British Columbia, Vancouver, Canada.

Fisher, I. [1922]: The Making of Index Numbers, (Boston Mass.: Houghton Mifflin).

Frisch, R. [1936]: "Annual Survey of General Economic Theory: The Problem of Index Numbers", Econometrica 4, 1-39.

Hoaglin, D. [1985] : "Using Quantiles to Study Shape" in D.C. Hoaglin, F. Mosteller, J.W. Tukey, eds., Exploring Data Tables, Trends, and Shapes, (New York: Wiley & Sons).

Kendall, M.G. and A. Stuart[1967]: The Advanced Theory of Statistics Vol 2, 2nd Edition, (New York: Hafner Publishing).

Konyus, A.A. [1924]: "The Problem of the True Index of the Cost of Living", Econometrica 7, 10-29.

Parzen, E. [1979]: "A Density-Quantile Function Perspective on Robust Estimation" in R. Launer, G. Wilkinson, eds., Robustness in Statistics, (New York: Academic Press).

Samuelson, P.A. [1983]: Foundations of Economic Analysis, Enlarged Edition, (Cambridge: Harvard University Press).

Statistics Canada [1983a]: "The Canadian Survey of Family Expenditure", Statistics Canada Publication Number 62-544.

Statistics Canada [1983b]: "The Consumer Price Index", Statistics Canada Publication Number 62-010.

Varian, H.R. [1982]: "The Nonparametric Approach to Demand Analysis", <u>Econometrica</u> 50, 945-973.