Using Engel Curves to Estimate Bias in the Canadian CPI as a Cost of Living Index

Timothy Beatty and Erling Roed Larsen

January 2004

Working Paper Number: 2004-05

Food and Resource Economics, University of British Columbia
Vancouver, Canada, V6T 1Z4
http://www.agsci.ubc.ca/fre

Using Engel Curves to Estimate Bias in the Canadian CPI as a Cost of Living Index.

Timothy K.M. Beatty
Canada Research Chair
University of British Columbia

&

Erling Roed Larsen
Research Fellow
Statistics Norway

Abstract

Semi-parametric Engel curves are used to infer bias in the Canadian CPI as a Cost of Living Index. The budget share of food has long been used as an indicator of welfare. We compare households with the same levels of CPI deflated total expenditure over the period 1978-2000. Differences in the expenditure share of food are attributed to the CPI failing to capture changes in wealth. We employ a novel econometric approach using a single index penalized linear spline model. We find that the CPI overstated changes in the cost of living by an average of 1.81% and 0.785% for single and two adult households respectively.

JEL Classification Codes

D1, C1

Corresponding Author

Timothy K.M. Beatty
Canada Research Chair
Food and Resource Economics Group
University of British Columbia
2357 Main Mall
Vancouver, B.C.
V6T 1Z4
timothy.beatty@ubc.ca

Introduction

Consumer Price Indices (CPI) are constructed for a number of reasons[1]. Two of the most important are as a Cost of Living Index (COLI) and as a measure of general inflation. However, these indices may not necessarily serve both purposes equally well. These two phenomena are quite different and arise from different sources. Inflation can be the result of too much printed money pursuing too few goods, which inflates the general price level. Increases in costs of living may be the result of changes in relative prices, with or without a change in the speed of money printing. Moreover, costs of living may change when new goods appear or old disappear or when quality improves or deteriorates. This research asks how well the CPI actually mirrors changes in the costs-of-living.

Understanding the extent to which the CPI captures or fails to capture changes in the cost of living is a crucial public policy question. The government indexes a number of programs to the CPI, such as the Canada Pension Plan, in an effort to maintain recipients' standard of living. Indeed, the CPI affects every Canadian directly; as of 2004, federal tax brackets have been indexed to the CPI.

We build on work by Hamilton [2] and Costa [3] by using Engel curves to estimate bias in the CPI as a COLI. The intuition here is straightforward and borrows from the literature on estimating household equivalence scales (in particular from Yatchew et al.

---

[1] For a complete discussion see Diewert [1].

[4]). The idea of using food expenditures as an indicator of welfare has a long history in economics. Engel's original notion is that households are assumed to be equally well off if and only if they dedicate the same share of their budget to food.

Rather than focusing on the differences in food expenditures between household types, we study the differences in food expenditures between time periods for the same household types. Controlling for changes in the price of food relative to the prices of all other goods, we compare demographically similar households with the same level of CPI deflated total expenditures at different points in time and then compare the share of total expenditures dedicated to food. For these households, differences in food's share of total expenditures are attributed to the CPI's inability to measure changes in the true cost of living.

As pointed out by Hausman [5], this approach accounts for two sources of bias: outlet and substitution bias. Outlet bias occurs when the when prices are not measured by the statistical agency where consumers are actually making their purchases (See White [6] for a discussion in the Canadian context). Substitution bias occurs when a fixed CPI basket fails to reflect the consumer's ability to substitute in response to changes in relative prices. However, this methodology neglects two other important sources of bias, new product introduction and quality change. Estimates of bias obtained in this manner, can therefore be thought of as a lower bound on the bias in the CPI.

Focusing on food expenditures offers a number of advantages: It leverages the empirical regularity known as Engel's law, which states that *ceteris paribus*, the budget share for food declines with total expenditure. Food prices are relatively easy to measure. In addition, contrary to many durable goods that present a host of measurement problems, food is perishable and therefore food expenditure in each period should closely track food consumption.

This research differs substantially in econometric approach from Hamilton and Costa by estimating a semi-parametric model that imposes far less structure on the estimation problem, allows a more direct estimation of potential bias and as a result is easier to interpret. In addition, we also differ from previous work by focusing specifically on Canadian data. To the best of our knowledge this is the first study to apply the Engel curve approach to Canadian data and one of the few to look at bias in Canadian cost of living measures.

Model

We begin by writing the food share for household *i* in period s, $w_{i,s}$ as a function of the household's total expenditure $y_{i,s}$ in period *s* deflated to period *t* using $p_{t,s}$ a true cost of living index, where $p_{t,t} = 1$. This yields a standard Engel curve of the form:

$$w_{i,s} = g\left(y_{i,s}/p_{t,s}\right) + \varepsilon_{i,s}.$$

Following Hamilton and Costa' approach, we observe $w_{i,s}$ and $y_{i,s}$over several time periods and in several geographic locations across Canada and attempt to infer the unobservable component $p_{t,s}$. However, rather than specifying a parametric form for the Engel curve, we employ a semi-parametric approach.

Rewriting $f = g \circ \exp$, a convolution of the food share with the exponential function, we obtain:

$$w_{i,s} = f\left(\ln(y_{i,s}) - \ln(p_{t,s})\right) + \varepsilon_{i,s}.$$

This form is similar to those proposed by Blundell et al. [7] which was subsequently employed by Pendakur [8] and Yatchew et al. [4] to estimate household equivalence scales.

As previously noted, we do not observe the true cost of living index, $p_{t,s}$, rather we observe the CPI at time t, $P_{t,s}$, which we model as a true cost of living index measured with error $\Delta_{t,s}$:

$$\ln(P_{t,s}) = \ln(p_{t,s}) + \ln(\Delta_{t,s}).$$

Rewriting in terms of observable components yields:

$$w_{i,s} = f\left(\ln(y_{i,s}) - \ln(P_{t,s}) + \ln(\Delta_{t,s})\right) + \varepsilon_{i,s}.$$

Finally denote total expenditure in period s deflated by the CPI to base period t for household i as $Y_{i,s,t} = \ln(y_{i,s}/P_{t,s})$ and $\ln(\Delta_{t,s}) = \delta_{t,s}$. For each household, this allows us to write Engel curves of the form:

$$w_{i,s} = f\left(Y_{i,s,t} + \delta_{t,s}\right) + \varepsilon_{i,s}.$$

Data

The data used in this paper are drawn from the public-use microdata files of the Survey of Family Expenditures for the years 1978, 1982, 1984, 1986, 1990, 1992, 1996 and the successor survey, the Survey of Household Spending for 1997, 1998, 1999 and 2000.

For each survey, we selected single adult and two adult households in urban (100,000+) areas where the respondent was between the ages of 18 and 64. These groups were the most consistently defined homogenous household types across all survey years. This yields the following sample sizes.

Table 1 Sample Size by Survey Year

Year	Single Adult Household	Two Adult Household
1978	**844**	**1100**
1982	**1280**	**1398**
1984	**840**	**744**
1986	**1332**	**1253**
1990	**668**	**709**
1992	**1072**	**1066**
1996	**1093**	**1098**
1997	**1296**	**1324**
1998	**1096**	**1071**
1999	**1257**	**1278**
2000	**1124**	**1124**
TOTAL	11902	12165

Rather than using the ad-hoc approach of deleting data points that seemed suspect, we used the following winsorization technique. For each survey, households with food and total expenditures below the 5th percentile or above the 95th percentile were recoded to be equal to value of the percentile they exceed. This approach has the advantage of leaving the median unchanged and preventing a number suspect data points from potentially influencing the results.

Figure 1 plots Engel curves for the survey years 1982-1999 against the reference year 1978. We can see that in each case the curve lies entirely below the reference curve. In addition the gap between the curves appears to be growing over time. This is consistent with the CPI failing to capture increases in wealth over time.

Table 2 Summary Statistics: Mean and Standard Error in Parentheses

Year	Single Adult Household		Two Adult Household	
	Food Share	Expenditure	Food Share	Expenditure
1978	**0.1606927 (0.06807605)**	**27650.17 (13528.64)**	**0.1530938 (0.05387821)**	**46776.89 (18005.71)**
1982	**0.1496959 (0.07320609)**	**26958.27 (14066.84)**	**0.1457039 (0.05441806)**	**45754.68 (19313.73)**
1984	**0.1545531 (0.06979923)**	**26976.02 (14008.28)**	**0.1386778 (0.05390399)**	**49133.94 (20738.40)**
1986	**0.1480444 (0.06807681)**	**27545.09 (14605.58)**	**0.1369416 (0.05174109)**	**49027.06 (19317.48)**
1990	**0.1301437 (0.05691208)**	**30764.74 (15708.53)**	**0.1240790 (0.04600280)**	**54015.42 (23230.62)**
1992	**0.1315975 (0.06162983)**	**28571.82 (15525.45)**	**0.1190408 (0.04766365)**	**51896.30 (23247.65)**
1996	**0.1314125 (0.06615837)**	**27521.73 (15863.04)**	**0.1200917 (0.04706552)**	**50020.69 (22317.17)**
1997	**0.1261803 (0.06211599)**	**28124.82 (16013.82)**	**0.1145800 (0.04824142)**	**50127.65 (22135.80)**
1998	**0.1242601 (0.05913459)**	**28174.88 (16467.31)**	**0.1122877 (0.04931827)**	**53535.72 (24879.12)**
1999	**0.1241522 (0.06119984)**	**28721.73 (16043.98)**	**0.1115259 (0.04573802)**	**53547.77 (24284.62)**
2000	**0.1268693 (0.06358135)**	**29520.63 (17670.82)**	**0.1063792 (0.04624768)**	**56431.09 (25347.59)**

Over the sample period, the budget share of food declines from 16 percent to 12 percent for single adult households and from 15 to 10 percent for two adult households. This is consistent both with a society becoming better off and with an increase in real expenditure over the period 1978 to 2000.

The CPI data were extracted from CANSIM II[2]. Provincial level CPI data are available for all years in the survey. However, the expenditure surveys have different geographic aggregation levels.

Table 3. Geographic Aggregation in Expenditure Surveys

Survey Years	Geographic Aggregation
1978, 1982, 1984	**Atlantic Provinces, Quebec, Ontario, Prairie Provinces (Manitoba, Saskatchewan, Alberta), British Columbia.**
1986, 1990	**Atlantic Provinces, Quebec, Ontario, Prairie Provinces (Manitoba, Saskatchewan), Alberta, British Columbia.**
1992, 1996, 1997, 1998, 1999, 2000	**Newfoundland, PEI, Nova Scotia, New Brunswick, Quebec, Ontario, Manitoba, Saskatchewan, Alberta, British Columbia.**

For survey years in which provinces are aggregated into regions, regional price indices were constructed as population weighted averages of their component provinces. Each household was matched to the most geographically disaggregated price index available.

Figure 2 plots the relative change in total CPI, Food and Non Food CPI. For most of the sample period the rate of increase in food prices has been below the rate of increase for the CPI as a whole.

Table 4 Aggregate Price Indices by Survey Year

---

[2] CANSIM II table 3260001

Year	**CPI**	**FOOD CPI**	**NONFOOD CPI**
1978	**43.60**	**46.75**	**42.93**
1982	**65.30**	**70**	**64.283**
1984	**72.10**	**76.6**	**71.1583**
1986	**78.1**	**82.76**	**77.1**
1990	**93.26**	**95.79**	**92.75**
1992	**99.98**	**99.96**	**100**
1996	**105.85**	**105.92**	**105.9**
1997	**107.57**	**107.55**	**107.625**
1998	**108.63**	**109.3**	**108.566667**
1999	**110.51**	**110.72**	**110.53**
2000	**113.53**	**112.24**	**113.88**

Estimation Approach

To reemphasize the intuition of our approach, if we were to observe a COLI $p_{t,s}$, then conditional on relative changes in the prices of foods P_f and nonfoods P_{nf} we would expect,

$$E\left[w_{i,t}(y_{i,t}) \mid P_f / P_{nf}\right] = E\left[w_{s,t}(y_{i,s}/p_{t,s}) \mid P_f / P_{nf}\right],$$

when $y_{i,t} = y_{i,s}/p_{t,s}$. That is, conditional upon relative prices, on average, the expenditure share of food should be the same for two households with the same level of total expenditures.

Note that P_f and P_{nf} are price indices and are presumably also measured with error. If we assume that the measurement error between food and nonfood is the same, then the bias will net out. If we assume food is measured with relatively less error (as seems plausible), then our estimates will provide a lower bound for the bias in the CPI.

Estimation requires two further assumptions. First we assume that the function $f(\cdot)$ is constant over the period. Second we assume that food and non-food are additively separable in the consumer's utility function.

These assumptions permit us to write the following estimating equation in matrix form:

$$W = f(Y + Z\delta) + \gamma \ln(P_f / P_{nf}) + \varepsilon_i,$$

where, for identification purposes, the coefficient on Y (an n vector of household total expenditures) is normalized to be equal to one. W is the n vector of food expenditure shares, Z is an n by q matrix of dummy variables, one column for each year beyond the reference year, δ is the q by 1 vector of bias parameters to be estimated, P_f / P_{nf} is the n vector of the ratio of food to non-food price indices and γ is the parameter on relative prices.

This model is estimated using a single-index penalized linear spline (p-spline) technique developed by Yu and Ruppert [9]. This approach offers a parsimonious means of estimating the model described above. It eliminates the need to execute a computationally expensive grid search of a q dimensional space for various choices of a smoothing parameter as is necessary in methods which rely upon Robinson's double residual approach [10].

The p-spline model was proposed by Ruppert and Carroll [11] and is exposited in Ruppert, Wand and Carroll [12]. This estimation technique uses a truncated power function basis of degree p for the component u that is to be modeled nonparametrically:

$$B(u) = \left(1, u, \cdots, u^p, (u-\kappa_1)_+^p, \cdots, (u-\kappa_K)_+^p\right),$$

with K knot points denoted κ_i. The function $(u-\kappa_j)_+^p$ equals $(u-\kappa_j)^p$ if $u > \kappa_j$ and zero otherwise. For tractability we choose $p = 2$.

In order to prevent over fitting, the influence of the extended basis function $(u-\kappa_1)_+^p, \cdots, (u-\kappa_K)_+^p$ is constrained by the use of a penalty function. For the purposes of this paper we adopt a simple quadratic penalty function[3]. Following Yu and Ruppert's recommendation, we set K equal to 5 and space the knot points evenly over the range of the single index component. The results presented below are robust to increasing the number of knots at the cost of increased computational time.

We write the mean function of the estimation problem as:

$$m\left(Y, Z, \ln\left(P_f / P_{nf}\right); \delta, \gamma, \theta\right) = \theta' B(Y - \delta Z) + \gamma \ln\left(P_f / P_{nf}\right),$$

where θ is a K+3 vector of parameters on the elements of the power function basis.

We then minimize the penalized criterion function:

$$Q(\delta, \gamma, \theta) = n^{-1} \sum_{i=1}^{n} \left\{ w_i - m\left(Z, \ln\left(P_f / P_{nf}\right); \delta, \gamma, \theta\right)\right\}^2 - \lambda \sum_{j=3}^{K+2} \theta_j^2,$$

using nonlinear least squares where the roughness penalty λ, is chosen to minimize some goodness of fit criterion.

[3] Yu and Ruppert describe a large variety penalty functions.

To estimate the model we rely upon the algorithm provided by Yu and Ruppert:[4]

1. Obtain initial parameter estimates from linear model $w_i = \delta_0 Y + Z\delta + \gamma \ln\left(P_f / P_{nf}\right) + \varepsilon_i$ using ordinary least squares. Normalize the resulting $\hat{\delta}$ such that the first element (δ_0) is equal to one. Construct the single index $\hat{u} = Y + Z\hat{\delta}$.
2. Now minimize

$$Q(\theta,\gamma) = n^{-1}\sum_{i=1}^{n}\left\{w_i - \sum_{j=0}^{K+2}\theta_j B_j(\hat{u}_i) - \gamma \ln\left(P_f / P_{nf}\right)\right\}^2 - \lambda \sum_{j=3}^{K+2}\theta_j^2,$$

   in order to obtain initial estimates of $\hat{\gamma}$ and $\hat{\theta}$.
3. The function is then jointly minimized with respect to both δ,γ,θ for a given smoothness parameter λ, starting at the values calculated in steps 1 and 2.
4. We search over a 40-point grid equally spaced over $\log_{10}(-3)\ldots\log_{10}(3)$ and choose the value of λ that minimizes the Generalized Cross-Validation (GCV) Criterion. The GCV approximates the leave-one-out Cross-Validation criterion but requires far less computation.

Yu and Ruppert show that the resulting parameter estimates are strongly consistent and asymptotically normal. Because household survey data are notoriously noisy (see Deaton

[4] Yan Yu kindly provided Matlab code which we then ported to R [13].

[14] for a comprehensive discussion), we supplement the standard asymptotic results with bootstrapped confidence intervals for the parameters of interest[5].

We follow Horowitz's [16] dictum to use an asymptotically pivotal statistic to estimate the probability distribution of the bias estimate: the percentile t bootstrap confidence interval. For each bootstrap sample we calculate the t-statistic, $t = (\hat{\delta}_{s,t}^{B} - \hat{\delta}_{s,t}) \big/ s_{\delta_{s,t}}^{B}$, where $s_{\delta_{i,j}}^{B}$ is the standard error and the superscript B indicates that this is the bootstrapped estimate.

The resulting estimates are sorted and the 0.025[th] and 0.975[th] quantiles are denoted $q_{.025}$ and $q_{.975}$ respectively. This yields 95% percentile-t confidence intervals of the form:

$$\left(\hat{\delta}_{s,t} - q_{.975} \cdot s_{\hat{\delta}_{t,s}} \ldots \hat{\delta}_{s,t} + q_{.025} \cdot s_{\hat{\delta}_{t,s}}\right).$$

The standard bootstrap is not consistent in the presence of heteroskedasticity of unknown form. Following Hardle [17] we employ the "wild" bootstrap and construct the percentile-t confidence intervals for the parameters of interest. Each bootstrap sample is constructed in the following manner. For each estimated residual $\hat{\varepsilon}_i$, we draw from a two-point distribution that takes on the value $\hat{\varepsilon}_i\left(1 - \sqrt{5}\right)\big/2$ with probability $\left(5 + \sqrt{5}\right)\big/10$ and $\hat{\varepsilon}_i\left(1 + \sqrt{5}\right)\big/2$ with probability $\left(5 - \sqrt{5}\right)\big/10$. The result of this draw is added to the

---

[5] Chapter 8 in Yatchew [15] provides an excellent overview of bootstrapping in nonparameteric and semiparametric models.

observed values w_i to construct a bootstrap sample. In this manner, we constructed 1000 bootstrap samples.

Results

For each household type we report the number of effective parameters (df_{FIT},), the value of the generalized cross validation criterion (GCV) for the chosen roughness penalty ($\hat{\lambda}$).

Table 5 Goodness of Fit Measures

Statistic	Single Adult Household	Two Adult Household
	Value	Value
df_{FIT}	**8.582797**	**8.586746**
GCV	**0.002584577**	**0.001499446**
$\hat{\lambda}$	**0.03455107**	**0.004124626**

The number of effective parameters df_{FIT}, is a measure proposed by Hastie and Tibshirani [18] which corresponds to the trace of the smoothing matrix. We see that the partial linear penalized spline approach reduces the dimension of the fit relative to estimating each coefficient separately.

For both samples, the bias is significantly different from zero in every year. We report the coefficient, standard error and percentile-t confidence intervals. The confidence intervals for the $\hat{\delta}$ terms are slightly skewed to the left and the confidence interval about relative prices is slightly skewed to the right.

Table 6 Single Adult Household

Variable	Coefficient	Standard Error	Percentile-t 95% C.I.
$\hat{\delta}_{78,82}$	**0.15789026**	**0.02940806**	**0.10666160... 0.2072154**
$\hat{\delta}_{78,84}$	**0.08525843**	**0.03239604**	**0.03828955... 0.1355529**
$\hat{\delta}_{78,86}$	**0.16152968**	**0.03023276**	**0.11400900... 0.2094380**
$\hat{\delta}_{78,90}$	**0.25483786**	**0.04132643**	**0.19724605... 0.3060570**
$\hat{\delta}_{78,92}$	**0.32438679**	**0.03812655**	**0.26605445... 0.3764835**
$\hat{\delta}_{78,96}$	**0.38208223**	**0.03738303**	**0.32545572... 0.4342219**
$\hat{\delta}_{78,97}$	**0.43063562**	**0.03778899**	**0.37655926... 0.4806150**
$\hat{\delta}_{78,98}$	**0.49660940**	**0.03815946**	**0.44215029... 0.5415409**
$\hat{\delta}_{78,99}$	**0.44920178**	**0.03807702**	**0.39238240... 0.4986614**
$\hat{\delta}_{78,00}$	**0.38356242**	**0.03889406**	**0.32676732... 0.4348292**
$\hat{\gamma}$	**0.09754696**	**0.01610468**	**0.07685293... 0.1188890**

Table 7 Two Adult Household

Variable	Coefficient	Standard Error	Percentile-t 95% CI
$\hat{\delta}_{78,82}$	**0.14851876**	**0.01863084**	**0.11848747... 0.1773427**
$\hat{\delta}_{78,84}$	**0.12282413**	**0.02327989**	**0.08528389... 0.1575135**
$\hat{\delta}_{78,86}$	**0.14030446**	**0.02052676**	**0.10710307... 0.1688713**
$\hat{\delta}_{78,90}$	**0.20979804**	**0.02792397**	**0.16695810... 0.2450384**
$\hat{\delta}_{78,92}$	**0.31107418**	**0.02711950**	**0.27021232... 0.3451201**
$\hat{\delta}_{78,96}$	**0.33990214**	**0.02684407**	**0.30166924... 0.3745945**
$\hat{\delta}_{78,97}$	**0.41065593**	**0.02617763**	**0.36509011... 0.4421265**
$\hat{\delta}_{78,98}$	**0.39710517**	**0.02695372**	**0.35510718... 0.4353965**
$\hat{\delta}_{78,99}$	**0.39253526**	**0.02703719**	**0.35358153... 0.4262884**
$\hat{\delta}_{78,00}$	**0.40010374**	**0.02986965**	**0.35432970... 0.4354250**
$\hat{\gamma}$	**0.08732175**	**0.01235918**	**0.07228192... 0.1066685**

The parameter estimates are all significantly different from zero at the 0.01 percent confidence level. The point estimates of the bias are larger for single adult households than for two adult households for most years.

Discussion

Using the parameter estimates calculated above, we can infer the cumulative implied bias according to formula. Table 8 reports estimated cumulative biases and their standard errors obtained via the delta method.

Table 8 Implied Cumulative Biases and Standard Errors in Parentheses

Years	Implied Cumulative Bias: Single Adult Household	Implied Cumulative Bias: Two Adult Household
1978-1982	**0.171 (0.034437)**	**0.160 (0.0216139)**
1978-1984	**0.088 (0.0352792)**	**0.130 (0.0263222)**
1978-1986	**0.175 (0.0355328)**	**0.150 (0.0236186)**
1978-1990	**0.290 (0.0533215)**	**0.233 (0.0344422)**
1978-1992	**0.383 (0.052736)**	**0.364 (0.0370151)**
1978-1996	**0.465 (0.0547786)**	**0.404 (0.0377108)**
1978-1997	**0.538 (0.0581283)**	**0.507 (0.0394708)**
1978-1998	**0.643 (0.0627014)**	**0.487 (0.0394708)**
1978-1999	**0.567 (0.059669)**	**0.480 (0.0400348)**
1978-2000	**0.467 (0.0570772)**	**0.491 (0.0445649)**

If we assume the bias is constant over the sample period, we can calculate an average annual bias. This is equal to an average annual bias of 1.81 percent for single adult households and 0.785 percent for two adult households for 1978-2000. Both Hamilton and Costa find (using alternative estimation approach) an average bias of 1.6 percent between 1972 and 1994 in the United States.

These averages hide considerable year-to-year variation in the estimated average annual bias (see Table 9). In particular, for the most recent survey years the CPI seems to understate the true cost of living index.

Table 9 Average Annual Biases

Years	Average Annual Bias: Single Adult Household	Average Annual Bias: Two Adult Household
1978-1982	**0.04026**	**0.03783**
1982-1984	**-0.03566**	**-0.01276**
1984-1986	**0.03887**	**0.00878**
1986-1990	**0.0236**	**0.01753**
1990-1992	**0.03539**	**0.05194**
1992-1996	**0.01453**	**0.00723**
1996-1997	**0.04975**	**0.07332**
1997-1998	**0.0682**	**-0.01345**
1998-1999	**-0.0463**	**-0.00455**
1999-2000	**-0.06353**	**0.0076**

Conclusions

In this paper, we use Engel curves to estimate bias in the Canadian CPI as a true cost of living indicator. We find that the CPI overstated the increase in the cost of living by 46.7% for single adult households and 49.1% for two adult households over the period 1978-2000. Using household expenditure survey data for these years, we confirm findings from earlier research in the United States that the CPI overstates the true cost of living for the entire period. In other words, in terms of their expenditure on food, households are behaving as if they were wealthier than the CPI would suggest.

It is interesting to note that the estimated bias is negative from 1997 to 1999 for Two Adult Households and 1998 to 2000 for Single Adult Households. The causes of this decline (which suggests that the CPI is understating the cost of living increases for those years) bears further research.

Our results suggest that over the period 1978-1997, recipients of government programs indexed to the CPI, were being overcompensated relative to the increase in the cost of living. Since 1997 there is some evidence that recipients are being undercompensated. One possible explanation is the increase in the cost of housing in the final years of our sample period.

Our econometric approach is novel in that we impose only the minimal structure on the estimation and inference required to quantify the magnitude and the variability of the bias. The result is a model which is computationally efficient, straightforward to implement and easy to interpret.

References

1. W. E. Diewert, The Consumer Price Index and index number purpose, *Journal of Economic and Social Measurement* (2001), **27**, 167-248.
2. B. W. Hamilton, Using Engel's Law to Estimate CPI Bias, *American Economic Review* (2001), **91**, 619-630.
3. D. L. Costa, Estimating Real Income in the United States from 1888 to 1994: Correcting CPI Bias Using Engel Curves, *Journal of Political Economy* (2001), **109**, 1288-1310.
4. A. Yatchew, Y. Sun and C. Deri, Efficient Estimation of Semiparametric Equivalence Scales with Evidence From South Africa, *Journal of Business Economics and Statistics* (2003), **21**, 247-257.
5. J. Hausman, Sources of Bias and Solutions to Bias in the Consumer Price Index, *Journal of Economic Perspectives* (2003), **17**, 23-44.

6. A. G. White, Outlet types and the Canadian Consumer Price Index, *Canadian Journal of Economics* (2000), **33**, 488-505.
7. R. Blundell, A. Duncan and K. Pendakur, Semiparametric Estimation of Consumer Demand, *Journal of Applied Econometrics* (1998), **13**, 435-461.
8. K. Pendakur, Semiparametric Estimates and Tests of Base-Independent Equivalence Scales, *Journal Of Econometrics* (1999), **88**, 1-40.
9. Y. Yu and D. Ruppert, Penalized Spline Estimation for Partially Linear Single-Index Models, *Journal of the American Statistical Association* (2002), **97**, 1042-1054.
10. P. Robinson, Root-N-Consistent Semiparametric Regression, *Econometrica* (1988), **56**, 931-954.
11. D. Ruppert and R. Carroll, Penalized Regression Splines, Cornell University (1997).
12. D. Ruppert, R. Carroll and M. Wand, Semiparametric Regression, Cambridge University Press Cambridge (2003).
13. R. D. C. Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria (2003).
14. A. Deaton, The Analysis of Household Surveys: A Micro Economic Approach to Development Policy, The Johns Hopkins University for the World Bank Baltimore (1997).
15. A. Yatchew, Semiparametric Regression for the Applied Econometrician, Cambridge University Press Cambridge (2003).
16. J. Horowitz, The Bootstrap in Econometrics, *in* "Handbook of Econometrics" (J. J. Heckman and E. E. Leamer, Eds.), Elsevier Science (2001).
17. W. Hardle, Applied Nonparametric Regression, Cambridge University Press Cambridge (1990).
18. T. J. Hastie and R. Tibshirani, Generalized Additive Models, Chapman & Hall London (1990).

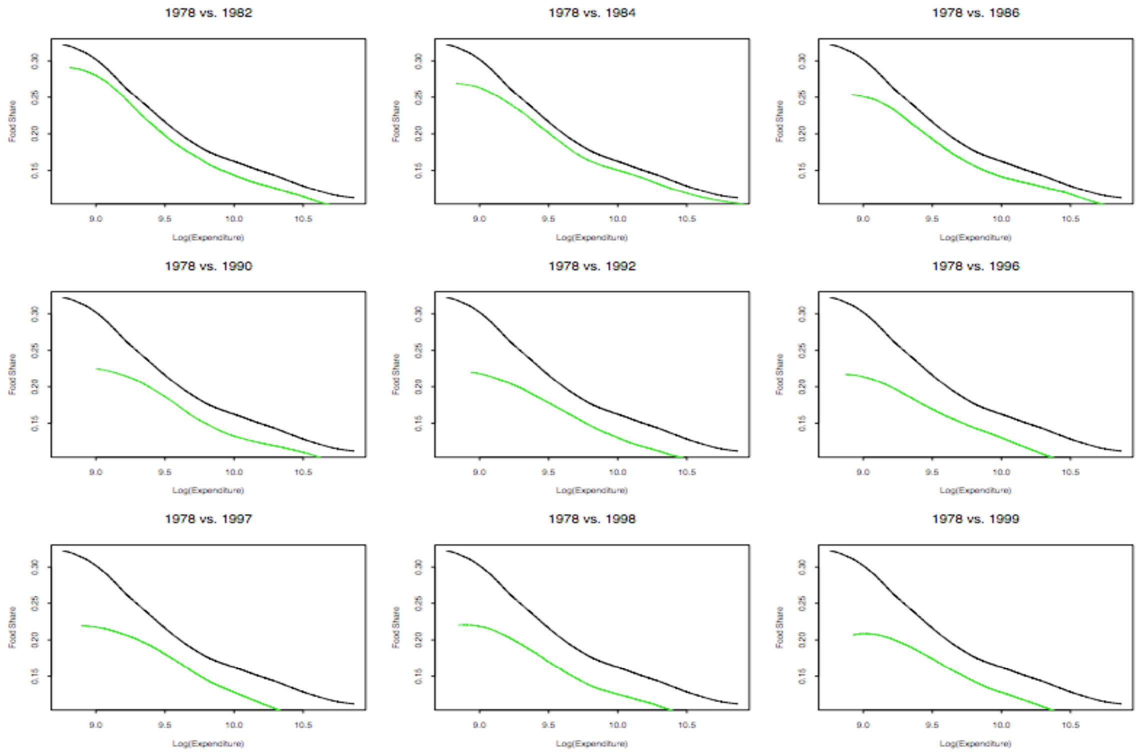

Figure 1 Single Adult Households: Engel Curves

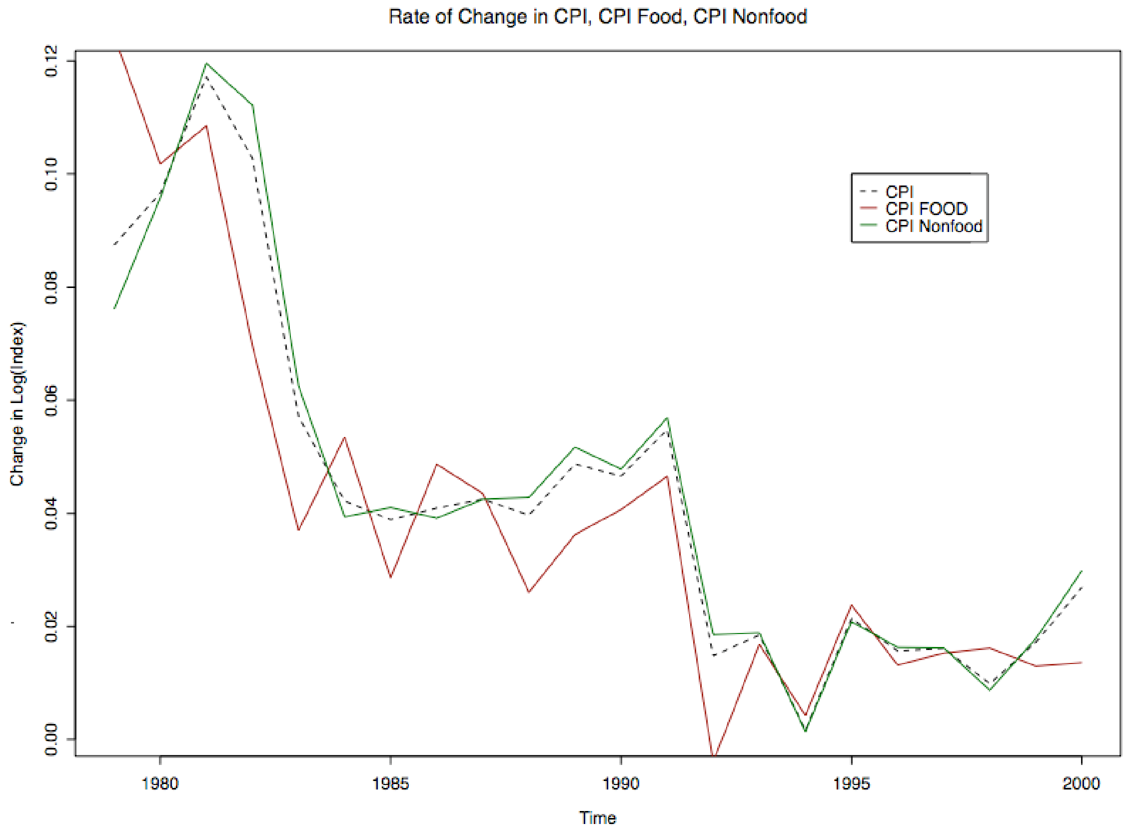

Figure 2 Relative Change in Price Indices