

Using Engel Curves to Estimate the Bias in the Australian CPI

Garry F. Barrett
School of Economics
University of New South Wales
Sydney, NSW, 2052
g.barrett@unsw.edu.au

and

Matthew Brzozowski
Department of Economics
York University
Toronto, Ontario, M3J 1P3
brzozowski@yorku.ca

This draft: August 2008.

Abstract

The Australian CPI is a Laspeyres index with fixed quantity weights based on the consumption patterns observed in an earlier period. Laspeyres-type price indices are subject to number of well-known biases which were highlighted in the Boskin Report (1996) for the US. This paper evaluates the performance of the Australian CPI as a true cost of living index. The analysis is based on ‘Engel’s Law’ which states that, other things equal, the budget share for food declines as total expenditure increases. Food Engel curves are estimated, based on the Working-Leser specification, using the series of ABS Household Expenditure Surveys spanning 1975/76-2003/04. The difference in the food share for households with the same level of CPI-deflated total expenditure provides a measure of bias in the CPI. The main findings are that the Australia CPI overstated changes in the general cost of living by approximately 28% between 1984 - 2003/04 (38% between 1975/76 -2003/04). However, there is substantial heterogeneity in the inflation rates experienced by different households and demographic groups. It was found that the CPI was an accurate measure of changes in the cost of living facing working families and two-adult families. The CPI was found to be a substantially less accurate measure of changes in the cost of living for single men and women and lone parent families.

1 INTRODUCTION

Assessing the performance of the economy over time relies on the accurate measurement of price changes. The measurement of changes in real GDP over time, productivity growth, changes in real wages and household income all depend on the a measure of changes in nominal prices. The most widely used price deflator is the consumer price index (CPI). The Australian CPI, compiled by the Australian Bureau of Statistics, is a Laspeyres-type index which has fixed quantity weights for different commodity groups based on expenditure patterns observed in an earlier period. Laspeyres price indices are subject to number of well-known biases (such as substitution, outlet, new good and quality bias) which were highlighted in the Boskin Report (1996) for the US.

This paper evaluates the performance of the Australian CPI as a true cost of living index. The analysis is based on ‘Engel’s Law’ which states that, other things equal, the share of a household (or family) budget devoted to food declines as total expenditure increases. Engel’s law has formed the basis of many studies of household welfare, where households are assumed to be equally well off if they devote the same share of their budget to food. Engel’s law is an empirical relationship which has been observed in data from many countries, and within countries over time. As Houthakker (1987) succinctly summarised, “of all the empirical regularities observed in economic data, Engel’s Law is probably the best established; indeed it holds not only in the cross-section data where it was first observed, but has often been confirmed in time-series analysis as well.”

Hamilton (2001) and Costa (2001) used Engel’s law to estimate the bias in the US CPI. Their analyses are based on the simple idea that if the CPI is an accurate measure of the cost of living then CPI-deflated Engel curves (food-share equations expressed as a function of real expenditure) estimated at different points in time should coincide. Alternatively, drift in the CPI-deflated Engel curves over time will reflect systematic bias in the measurement of the CPI (after controlling for changes in the relative price of food and for changes in the composition of the population).. This paper follows the Hamilton-Costa approach by estimating food Engel curves for Australia using the ABS

Household Expenditure Surveys which span the period 1975/76-2003/04. The Working-Leser specification for Engel curves, with extensions recommended by Blow (2003), are estimated and the accuracy of the Australian CPI as a cost of living index (COLI) evaluated.

The main findings are that the Australia CPI overstated average changes in the general cost of living by approximately 28% between 1984 - 2003/04 (38% between 1975/76 -2003/04). There is substantial heterogeneity in household-specific inflation rates, and it was found that the CPI was an accurate measure of the average change in the cost of living facing Australian working families and two-adult families. However, the CPI was found to be substantially less accurate in measuring changes in the cost of living for single men and women and lone parent families.

2 LITERATURE REVIEW

2.1 The Australian Consumer Price Index

The historical background of the Australian consumer price index (CPI) is outlined in ABS (2005a). The CPI was introduced in 1960, with the index calculated retrospectively back to 1948. The original aim of the CPI was to measure changes in retail prices of goods and services purchased by metropolitan employee households (ABS 2005a: 3). That is, at its inception, the primary purpose of the CPI was to provide a COLI for metropolitan wage-earnings households. This purpose of the CPI (and prior retail price index series dating back to the ‘A Series’ first compiled in 1912) reflected its role in the wage determination process in Australia.

The Australian CPI is a Laspeyres index with fixed quantity weighted for each commodity groups based on past observed expenditure patterns. The CPI is reviewed and re-weighted approximately every five years. The last substantial review of the CPI occurred with the release of the 13th series in September 1998. Coinciding with that release, primary objective of the CPI changed from providing a COLI for employee households to providing a general measure of price inflation facing the household sector (ABS 2005a).

This change in the objective of the CPI reflected its increasingly important role as an input into macroeconomic policy development, especially by the Reserve Bank of Australia in setting monetary policy (ABS 2005a: 45). The main consequence of this change on the compilation of the CPI was the exclusion of interest charges, and the inclusion of house purchase prices. Since its introduction the CPI has now been reviewed and re-weighted fifteen times. The latest CPI series was released in September 2005 and is based on expenditure patterns recorded in the HES 2003-04 (ABS 2005b: 7).

2.2 Studies of CPI Bias Based on the Engel Curve Method

US: Hamilton (2001), Costa (2001)

Canada: Beatty and Larsen (2005), Brzozowski (2006)

NZ: Gibson and Scobie (2002),

Russia: Gibson, Stillman and Le (2007)

Norway: Larsen (2007)

Brazil and Mexico: de Carvalho Filho and Chamon (2006, 2007)

3 METHODS

Spending on food is the focus of this literature for several reasons. First, food is non-durable and reported expenditure in even relatively short reporting periods is likely to correspond closely to consumption. Second, it has been established in many empirical demand studies that the elasticity of food expenditure with respect to total expenditure (or income) is significantly less than one. It is important that the elasticity of the commodity is not equal to unity, otherwise it would not be possible to distinguish between shifts along the Engel curve due to changes in real total expenditures from shifts of the Engel curve due to mis-measured prices (that is, mis-measured prices would appear as movement along a unit-income elasticity Engel curve). It is feasible that luxury goods, such as recreational expenditures, could also be used to measure CPI bias (see Costa 1999), although such goods are more likely to be subject to infrequency of purchase

problems, unlike food expenditure. Further, food has been found to be separable from other commodities in empirical studies of consumer demand systems in developed and developing countries, which further supports the singular focus on food expenditure.

The methods used in the analysis build on the Working-Leser specification for the food budget share defined as a function of total expenditure - the ‘food Engel curve’. The food budget share ω_{ijt} for family i residing in region j at time t is expressed as

$$\omega_{ijt} = \phi + \gamma \ln \left(p_{jt}^f / p_{jt}^{nf} \right) + \beta \ln (Y_{ijt} / p_{jt}) + X'_{ijt} \theta + \mu_{ijt} \quad (1)$$

where p_{jt}^f is the true price of food, p_{jt}^{nf} is a true price of non-food and p_{jt} is the true price level (which is a weighted average of p_{jt}^f and p_{jt}^{nf}), y_{ijt} is nominal total expenditure¹ and X_{ijt} is the vector of other covariates. The term $\ln \left(p_{jt}^f / p_{jt}^{nf} \right)$ in equation (1) is the log of relative price of food and the term $\ln (Y_{ijt} / p_{jt})$ is the log of real total expenditure.

Following Hamilton (2001) any price level p_{jt} can be decomposed into the true price level and an error term :

$$\ln p_{jt} = \ln p_{j0} + \ln(1 + \Pi_{jt}) + \ln(1 + E_t) \quad (2)$$

where p_{j0} is the true price level at time 0, Π_{jt} is the cumulative percent increase in the CPI from year 0 to t and E_t is the cumulative percent measurement error. It is assumed that any bias is uniform across regions j . To simplify notation let $\pi_{jt} = \ln(1 + \Pi_{jt})$ and $\varepsilon_t = \ln(1 + E_t)$, substitute these into (2), then substitute the expression from (2) into (1) to obtain:

$$\begin{aligned} \omega_{ijt} = & \phi + \gamma \left(\pi_{jt}^f - \pi_{jt}^{nf} \right) + \beta (y_{ijt} - \pi_{jt}) + X'_{ijt} \theta + \\ & \gamma \left(\varepsilon_t^f - \varepsilon_t^{nf} \right) - \beta \varepsilon_t + \gamma \left(p_{j0}^f - p_{j0}^{nf} \right) - \beta p_{j0} + \mu_{ijt} \end{aligned} \quad (3)$$

The model can be estimated by

$$\omega_{ijt} = \varphi + \gamma \left(\pi_{jt}^f - \pi_{jt}^{nf} \right) + \beta (y_{ijt} - \pi_{jt}) + X'_{ijt} \theta + \sum_{t=1}^T \delta_t D_t + \mu_{ijt} \quad (4)$$

where D_t is a time dummy variable equal to one at time t and is zero otherwise. The coefficients on the time dummies reflect, ceteris paribus, the extent of the cumulative bias

¹Hamilton (2001) uses income rather than total expenditure.

in the CPI from the base period. The terms in the equation (3) denoted by subscript 0 are constants and are absorbed into the intercept term in (4).²

It follows that:

$$\delta_t = \gamma (\varepsilon_t^f - \varepsilon_t^{nf}) - \beta \varepsilon_t \quad (5)$$

It is further assumed that the bias in the price of food and non-food is equal ($\varepsilon_t^f - \varepsilon_t^{nf} = 0$) and hence

$$\varepsilon_t = \frac{-\delta_t}{\beta}. \quad (6)$$

The accumulated bias in each year (relative to the base year) can then be calculated as:

$$Bias_t = 1 - \exp\left(\frac{-\delta_t}{\beta}\right) \quad (7)$$

Equivalently, the correction factor which multiplies the measured CPI in period t to give the true CPI in period t is

$$Correction_t = 1 - Bias_t = \exp\left(\frac{-\delta_t}{\beta}\right) \quad (8)$$

Costa (2001) extends Hamilton's specification³ by adding a quadratic term in the log of real expenditure:

$$\omega_{ijt} = \phi + \gamma \ln(p_{jt}^f/p_{jt}^{nf}) + \beta_1 \ln(Y_{ijt}/p_{jt}) + \beta_2 \ln(Y_{ijt}/p_{jt})^2 + X'_{ijt}\theta + \mu_{ijt} \quad (9)$$

The inclusion of the quadratic log of real income term in the budget share equation allows for additional curvature in the Engel curve. Banks et al. (1997) demonstrated that Engel curves specified as quadratic in the logarithm of real income accurately approximate their non- parametric representation.

Substituting (2) into (9) and again assuming $\varepsilon_t^f = \varepsilon_t^{nf}$ gives

$$\begin{aligned} \omega_{ijt} = & \varphi + \gamma (\pi_{jt}^f - \pi_{jt}^{nf}) + \beta_1 (y_{ijt} - \pi_{jt}) + \beta_2 (y_{ijt} - \pi_{jt})^2 + X'_{ijt}\theta \\ & - \beta_1 \varepsilon_t - \beta_2 \varepsilon_t^2 - 2\beta_2 (y_{ijt} - \pi_{jt}) \varepsilon_t + \mu_{ijt} \end{aligned} \quad (10)$$

²Note that $\varphi = \phi + \gamma (p_{j0}^f - p_{j0}^{nf}) - \beta p_{j0}$

³Costa (2001) uses the more conventional approach to Engel curve estimation by examining the food share of total expenditures (rather than total income as used by Hamilton 2001).

In turn the model simplifies to:

$$\begin{aligned} \omega_{ijt} = & \varphi + \gamma \left(\pi_{jt}^f - \pi_{jt}^{nf} \right) + \beta_1 (y_{ijt} - \pi_{jt}) + \beta_2 (y_{ijt} - \pi_{jt})^2 + X'_{ijt} \theta \\ & + \sum_{t=1}^T \delta_{1t} D_t + \sum_{t=1}^T (\delta_{2t} (y_{ijt} - \pi_{jt}) D_t) + \mu_{ijt} \end{aligned} \quad (11)$$

The model in equation (11) is overidentified and cannot be estimated by ordinary least squares. A linear estimator cannot extract the bias component and as a result the nonlinear least squares estimator is used.

To estimate the quadratic model rewrite equation (10) as

$$\omega_{ijt} = \varphi + \gamma \left(\pi_{jt}^f - \pi_{jt}^{nf} \right) + \beta_1 (y_{ijt} - \pi_{jt} - \varepsilon_t) + \beta_2 (y_{ijt} - \pi_{jt} - \varepsilon_t)^2 + X'_{ijt} \theta + \mu_{ijt} \quad (12)$$

which can be approximated by

$$\begin{aligned} \omega_{ijt} = & \varphi + \gamma \left(\pi_{jt}^f - \pi_{jt}^{nf} \right) + \beta_1 \left(y_{ijt} - \pi_{jt} - \sum_{t=1}^T \lambda_t D_t \right) \\ & + \beta_2 \left(y_{ijt} - \pi_{jt} - \sum_{t=1}^T \lambda_t D_t \right)^2 + X'_{ijt} \theta + \mu_{ijt} \end{aligned} \quad (13)$$

The parameters of the model in (13) are estimated through iterative grid search. Solving for λ_t components enables the estimates of cumulative bias to be recovered.

4 DATA AND SAMPLE CONSTRUCTION

The analysis is based on the Australian Bureau of Statistics Household Expenditure Survey (HES) unit record files. The HES was conducted in 1975/76, 1984, 1988/89, 1993/94, 1998/99 and 2003/04 (referred to as HES75-HES03). Using the HES data for the analysis is advantageous given that a key objective of the HES program is to obtain information on household expenditure patterns in order to revise the commodity weights underlying the construction of the CPI series.

For the analysis the detailed expenditure information in the HES data (over 600 unique commodity groups are recorded in 2003/04) is aggregated into two broad group - food expenditure and total expenditure. Food expenditure includes spending on food and non-alcoholic beverages, either for consumption at home or outside the home (such

as at cafes or restaurants). The food category explicitly excludes alcoholic beverages and tobacco. Total household expenditure consists of spending on food, alcohol and tobacco, current housing costs, fuel and power, household furnishings and equipment, household services and operations, clothing and footwear, medical care, health and personal care, transportation, recreation and ‘miscellaneous’ goods and services.⁴ The food budget share is simply defined as the ratio of food expenditures to total household expenditures.

Several steps were taken in selecting the analysis sample from the raw HES micro-data files. First, records for multiple-family households were dropped. Most multiple family households are comprised of unrelated young adults, and the expenditure information which is obtained from interviewing one household member can be very inaccurate. To minimise reporting errors only single-family households are selected. Approximately seven percent of observations (2,832 out of a total of 39,498 observations) were excluded on this criteria⁵. Observations were also dropped if total expenditure or food expenditures were reported as negative.⁶ To minimise the influence of extreme observations, the sample was refined with the top and bottom three percent of observations trimmed from the distribution of total expenditures and food shares in each survey year.⁷ In order to make the geographic coverage of the HES samples as homogeneous as possible, observations from the Australian Capital Territory (1975/76, 1984, 1993/94, 1998/99 and 2003/04) and the Northern Territory (1993/94, 1998/99 and 2003/04) were dropped from the sample.⁸ The sequence of exclusions resulted in a final analysis sample consisting of

⁴The HES expenditure categories not included in the composite total expenditure bundle are income tax payments, mortgage principal repayments, other capital housing costs, and superannuation and Life assurance expenditures (HES expenditure groups 14-17, respectively). These items are excluded because they are direct taxes or forms of savings.

⁵The incidence of multiple-family households within a single cross-section ranged from a high of 15% in HES75 survey to a low of 4.6 % in HES03.

⁶These selections resulted in a further 26 and 77 observations, respectively, being dropped.

⁷That is, each cross-sectional distribution of total expenditure and food shares was trimmed at the 3rd and 97th quantile.

⁸The Northern Territory (NT) was not covered by the 1975/76 or the 1984 surveys. In the 2003/04 survey the Australian Capital Territory (ACT) and NT were not separately identified. As a result observations for households recorded as residing in the NT or ACT were excluded. Since the 1988/89 survey does not contain an identifier for state or territory of residence, observations from the ACT or NT could not be excluded from this cross-section. A total of 2423 observations were excluded by this geographic restriction.

30190 observations.

As demonstrated by Blow (2003), it is important to condition on household characteristics when estimating Engel curves. If household characteristics are not adequately controlled for differences in estimated food Engel curves over time may reflect changes in the composition of the underlying population, rather than mismeasurement of prices. The covariates used in the analysis includes characteristics of the household reference person (whether female, married, immigrant status, whether employed full-time, employed part-time or self-employed, whether aged 65 years or older), family characteristics (indicators for presence of dependent children, presence of students⁹, whether it is a lone-parent family, and household size¹⁰) and a set of indicator variables for state of residence.

The HES microdata files were augmented with price information from the ABS CPI series. The CPI series for three commodity groups - total expenditures, food and non-food groups - were matched to the HES unit records by survey period (the third quarter of the survey period for each HES was used¹¹) and state of residence.¹² The CPI series published in ABS (2007: Table 13) were used in the analysis - see the Appendix for the complete list of individual CPI series used. The reference period adopted for the CPI bias calculations is 1984, and the reference state is New South Wales (NSW). The CPI series are rescaled with NSW in 1984 adopted as the base group.¹³

Descriptive statistics for the analysis sample are presented in Table 1. The individual HES contain record weights¹⁴ which are used in calculating the sample means and in the estimation. The first column of statistics in Table 1 contains the mean value of the variables for the pooled sample. Additional columns contain the sample means by

⁹Full-time students aged 15-24 years (except for HES75 and HES84 where it is students aged 15 years and older).

¹⁰Household size is top-coded at six in all years.

¹¹The analysis was also conducted using the annual average CPI for the four quarters comprising each HES survey period. The results were invariant to this choice.

¹²Except for the HES88 which included no regional identifiers. The national CPI series was used for the HES88 sample records.

¹³That is, we divided the CPI series for total expenditure, food and non-food items by the respective 1984 value for New South Wales.

¹⁴The record weights represent the inverse probability of selection into the survey.

survey year. For the pooled sample the average food budget share is 0.214. Reading across the row, the average food budget share declined over time with each survey. Based on ‘Engel’s law,’ the decline in food shares suggests a progressive improvement in the average well-being of Australian families over the 1975/76-2003/04 period. The relative price of food (to non-food) increased on average from 1975/76 to 1984, decreased to 1993/94 before increasing from 1993/94 to 2003/04. Average real total household expenditure (as deflated by the CPI) progressively increased across adjacent surveys except for a slight decline from 1984-1988/89. Over the full sample period average real household expenditure is measured to have grown by approximately 21 percent.

Other features of the sample include the increase in the incidence of families with full-time students, the decline in the incidence of families with children and a decline in average household size over time. At the same time, there was a rise in the incidence of single adult and lone parent families, and a corresponding decline in two-adult or ‘couple’ families. The sample characteristics reflect the demographic trends in Australian society.

Figure 1 presents partially linear model estimates of food Engel curve by survey year. The partially linear model is a semiparametric estimator (see Yatchew 1998, 2003) where the food budget share is specified as a general function of CPI-deflated total expenditure and a linear function of covariates (including relative price of food and family characteristics). The partially linear model does not restrict the Engel curves to have a linear or quadratic function of real expenditures. The figures provide an indication of whether the parametric linear (or quadratic) specifications of the Engel curves (in the logarithm of real expenditures) are supported by the data, as well as whether there is a shift in the location of the curves over time (revealing systematic mismeasurement of the CPI). To reiterate, if the official CPI correctly measures changes in the true cost of living then, given stable preferences and controlling for changes in relative price of food, the Engel curves estimated from different time periods will coincide. The amount of drift in the Engel curves over time is then hypothesised to represent bias in the measurement of the CPI. Figure 1 shows that the Engel curves based on the sequence of HES84-HES03 are generally parallel, while the 1975 Engel curve has a steeper slope. The estimated

Engel curves with the later HES data are lower and to the left of the earlier surveys. This is interpreted as the official CPI over-correcting for changes in nominal prices. The distance between the 1984 Engel curve and that based on other survey years is a measure of cumulative bias in the official CPI for that year (relative to the 1984 base). The figure suggests that the CPI over-estimated changes in the cost of living - and the econometric models estimated next quantify the amount of that bias. The shape of the partially linear model estimates of the Engel curves suggest that the linear in log-real expenditure appear to be a reasonable approximation to the semiparametric estimates.

The analysis proceeds by estimating CPI bias for the general population. Alternative specifications for the food Engel curves are considered (linear and quadratic in log of real total expenditure) and the sensitivity of the bias estimates to the treatment of conditioning variables is then examined. The accuracy of the CPI as a COLI for a range of specific demographic groups is then considered. The distinct demographic groups examined are non-elderly single men and women, adult couples with no children, adult couples with children, lone mother families, working families and seniors.

5 EMPIRICAL RESULTS

5.1 Bias Estimates for the Full Population

Table 2 presents the estimated CPI correction factors¹⁵ (and asymptotic standard error constructed using the delta method) based on the full sample. The first column of results are for the linear specification of the food Engel curve with controls for the log of relative prices and the log of real total expenditures only. For this simplified model, the estimated correction factor for 1975 is 1.21. The interpretation of the correction factor is that, relative to the 1984 base period, the official CPI for 1975 multiplied by a factor of 1.21 will produce the true COLI for 1975. Equivalently, the CPI level in 1975 was under-estimated (and the official inflation rate between 1975 and 1984 over-estimated) by 21%. The estimated correction factor for 2003 is 0.66 which in turn indicates cumulative

¹⁵Correction factor_t = 1 - Bias_t = exp(-δ_t/β).

positive bias in the 2003 official CPI of 34% (relative to the 1984 base).

The second set of results are for the linear specification with the full set of covariates. Conditioning on the full set of individual and family characteristics substantially reduced the estimated bias in the CPI. The set of covariates are jointly highly statistically significant indicating that the bias estimates in model (1) in part reflected shifts over time in the Engel curves due to changes in the composition of the population. Based on this superior specification, the 1975 CPI is found to be underestimated by 10% relative to the 1984 base year. The CPI for 2003 is found to be over-estimated by 28% relative to 1984 (an annual average bias of 0.89% over this period), which in turn implies inflation between 1984 and 2003 was over-estimated. For the full 1975/76-2003/04 observation period, the CPI is found to overestimate changes in the true cost of living by 38% (or an annual average bias of 0.9%).

Several tests were performed for each model. The first test was an F-test of the null hypothesis that the coefficients on all of the time dummy variables were jointly insignificant (against the alternative that the null is false). A second F-test performed for the null that the coefficients on the subset of time dummy variable for 1988-2003 are jointly insignificant.¹⁶ The F-test statistics are reported in the row immediately below the estimated correction factors. Both null hypotheses were strongly rejected for linear model specifications at conventional levels of significance.

The quadratic models specifications were estimated and the implied CPI correction factors (and asymptotic standard errors) are presented in columns (3) and (4). Again the model with the full set of covariates produce lower estimates of CPI bias, underscoring the importance of controlling for changing population demographics. The quadratic specification produces a set of correction factors very similar to that found with the linear specification of the Engel curve. The cumulative bias from in the CPI from 1984 to 2003 (1975 to 2003) is estimated to be 20% (31%), with an annual average bias of 1.04% (0.97%).

¹⁶Since an intercept difference may not adequately capture the difference between the food Engel curve for 1975 and that for the later year, this second hypothesis (which exclude the 1975 estimates) is also considered.

5.2 Bias Estimates by Demographic Groups

The Engel curve methodology for assessing CPI bias was then implemented to consider the accuracy of the CPI for measuring changes in the true cost of living for a range of more narrowly defined demographic groups. Expenditure patterns vary across families, and each household essentially experiences their own unique inflation rate over time. It is informative to consider how the CPI performed as an indicators of changes in the cost of living for specific demographic groups.

Figures 2-7 presents the partially linear model plots of the Engel curves for non-elderly single men and women, couple families without children, couple families with children, lone mother families, working families (where an adult member is employed) and seniors (where the reference person is aged over 65 years), respectively. Table 3 presents the estimated correction factors based on the quadratic specifications with the full set of covariates. Several features are evident from the estimates. First, the CPI is found to be a particularly poor guide to cost of living for single men and women and for lone mothers. The cumulative bias over the period 1984-2003 (1975-2003) for single men and women is estimated to be 43% (88%) and for single mothers is 39% (148%). Second, the CPI bias estimated for seniors mirrors that found for the full sample. Third, the CPI is found to be a generally accurate measure of the cost of living for couple families with and without children, and especially for working families. In these models the point estimates of the cumulative bias are generally small, and individually statistically insignificant. Further, the set of 1984-2003 year dummy variables in the separate models estimated with these subgroups are jointly insignificant. The test results imply no significant bias in the CPI as a COLI over the 1984-2003 period for these groups. The point estimates of the correction factors for the working families sample are very close to 0 for the years 1975-1993 which coincides with the time period during which the explicit objective of the CPI was to measure the changes in the cost of living of metropolitan working families in Australia. When that objective of the CPI collection changed from 1998, the point estimates suggest more of divergence between the CPI and the cost of living for this

group. Even so, the divergence between the CPI and the estimated change in the cost of living up to 2003 was not statistically significant.

5.3 Sensitivity Analysis

The robustness of the empirical finding with respect to a series of model specification and sample definition choices was examined. The estimated CPI correction factors from the sensitivity analysis are presented in Table 4. Given that the linear and quadratic Engel curve models yielded very similar results, the sensitivity analysis was based on the more parsimonious linear Engel curve model. First, the model was re-estimated with the omission of the 1975 data. The Engel curve graphs suggested that a simple intercept difference may not adequately capture the difference between the Engel curves between 1975 and the other survey years. This may introduce inconsistency into the CPI bias estimation for survey years 1988-2003 through the estimated impact of the covariates. The estimated correction factors for the linear model with the 1975 data omitted are shown in column (1) of Table 4. The results indicate that the correction factors are marginally smaller for 1988, 1993 and 1998, though slightly larger for 2003. However, the point estimates are within one standard error of, and not statistically different from, the full sample estimates.

The model was then re-estimated with the 1975 data included but the 1988 survey data excluded. The HES88 data did not include regional identifiers so in this survey year there is no regional variation in the CPI series. Data from HES88 does not contribute to the identification of the relative price of food term, and there measurement error may be introduced into the estimation due to the aggregation across regions. As shown in the column (2) of Table 4, the exclusion of the 1988 data resulted in a slightly larger implied CPI bias estimate for 1975, slightly less for bias form 1993 and 1998, and marginally greater bias in 2003. However, again, these point estimates are not significantly different from the estimated based on the full set of HES surveys.

Another issue in the estimation of the models is the treatment of record weights for the microdata. In linear models, if the factors which determine the record weights are

controlled for in the estimation the estimates based on an unweighted regression will be unbiased and consistent (intuitively - this is equivalent to exogenous sample selection). The unweighted regression estimates are shown in column (3) of Table 4. Again, the point estimates are very similar to the base model estimates which use the inverse probability weights, and hence the conclusions regarding CPI bias measurement are robust to the treatment of the record weights.

As discussed previously, it is important in Engel curve estimation to include a rich set of covariates to capture population heterogeneity. Otherwise, changes in the population (and reflected in the sample) over time will be absorbed into coefficients on the time dummy variables, and the CPI bias estimates measure true bias and time trends in the unmeasured underlying characteristics. Although a rich set of covariates were included in the estimation; richer information is available in the HES microdata. To check sensitivity of the estimation along this dimension, a set of 12 dummy variables for the age of the household reference person (age grouped into five-year bands ranging from 15-19 years, through to age 75-79 years). This set of variables will allow for discrete variation in the Engel curves over stages of the lifecycle, and control for the ageing of the Australian population over the sample period. Another variable added to the model was a dummy variable indicating whether the household owned the home (as opposed to renting). Home ownership status increased over the 1975-2003 period, and if home owners have different food expenditure patterns compared to renters, other things equal, then the measurement framework may attribute this to CPI bias. The estimated correction factors based on the richer covariate set are shown in column(4). It is evident that there was a slight increase in the estimated correction factors for 1975 and 2003, and slightly smaller correction factors for 1988-1988, relative to the base model. Again, the difference in estimates are neither large in magnitude nor statistically significant.

Overall, the sensitivity analysis has shown that the estimated correction factors, equivalently the estimates of Australian CPI bias, for the full population sample are insensitive to a wide range of model specification and sample definition issues.

6 CONCLUSION

The results of the analysis show that the CPI over-estimated changes in the cost of living experienced by the Australian population over the 1975/76 - 2003/04 period by approximately 1% per year on average. Sensitivity analysis showed that these estimates are robust to a range of sample definition, model specification and estimation issues. Given the heterogeneity in the cost of living across households, the CPI proved to be an especially inaccurate measure for single men and women and lone mother families. However, the Australian CPI has been successful in accurately tracking changes in the cost of living facing working families, and couple families, in Australia.

References

- [1] Australian Bureau of Statistics (2005a) *Australian Consumer Price Index: Concepts, Sources and Methods*, ABS Catalogue No. 6461.0
- [2] Australian Bureau of Statistics (2005b) *A Guide to the Consumer Price Index: 15th Series*, ABS Catalogue No. 6440.0
- [3] Australian Bureau of Statistics (2006) *Household Expenditure Survey 1975/1976-2003/2004*.
- [4] Australian Bureau of Statistics (2007) *Consumer Price Index Australia*, ABS Catalogue No. 6401.0
- [5] Banks, J., Blundell, R. and Lewbel, A. (1997), “Quadratic Engel Curves and Consumer Demand”, *Review of Economics and Statistics*, vol 79, pp. 769-88.
- [6] Beatty, T. and Larsen, E. (2005) “Using Engel curves to estimate bias in the Canadian CPI as a cost of living index” *Canadian Journal of Economics* v.38(2), pp.482–499.
- [7] Blow, L. (2003) “Explaining trends in UK household spending” *Institute for Fiscal Studies Working Paper 03/06*
- [8] Boskin et al (1996), “Toward a more accurate measure of the cost of living”, *Final Report to the Senate Finance Commission from the Advisory Commission to Study the Consumer Price Index*.
- [9] Brzozowski, M. (2006) “Does One Size Fit All? The CPI and Canadian Seniors,” *Canadian Public Policy*, v.32(4), pp.387-411
- [10] Costa, D. (2001), “Estimating real income in the US from 1888 to 1994: correcting CPI bias using Engel curves”, *Journal of Political Economy*, vol 109, pp. 1288-1310.

- [11] Diewert, W. E. (2001) “The Consumer Price Index and index number purpose,” *Journal of Economic and Social Measurement*, vol, 27, no.3-4, pp. 167-248.
- [12] Gibson, J., Stillman S. and Le, T. (2007), “CPI bias and real living standards in Russia during the transition” *Journal of Development Economics*, forthcoming.
- [13] Gibson, J., Scobie, G. (2002), “Are we growing faster than we think? An estimate of ‘CPI bias’ for New Zealand” mimeograph, University of Waikato.
- [14] Hamilton, B., (2001) “Using Engel’s law to estimate CPI bias” *American Economic Review*, vol 91, pp 619-30.
- [15] Hausman, J. (2003) “Sources of bias and solutions to bias in the Consumer Price Index” *Journal of Economic Perspectives*, v.17(1), pp.23–44.
- [16] Houthakker, H. (1987) “Engel’s law” in: Eatwell, J., Milgate, M., and Newman, P. (eds.), *The New Palgrave Dictionary of Economics*, vol. 2. McMillan, London, pp. 143–144.
- [17] Yatchew, A. (1998) “Nonparametric Regression Techniques in Economics” *Journal of Economic Literature*, vol 36(2), pp 669-721.
- [18] Yatchew, A. (2003) *Semiparametric Regression for the Applied Econometrician*, Cambridge, Cambridge University Press.

Table 1. Summary Statistics

Variable	Pooled Sample	1975	1984	1988	1993	1998	2003
Food Share	0.214	0.231	0.226	0.222	0.209	0.207	0.195
Ln(rel price food)	-0.013	-0.053	-0.003	-0.031	-0.046	0.014	0.045
Ln(expenditure)	5.661	5.579	5.643	5.618	5.649	5.694	5.768
<i>Demographics</i>							
Employed	0.592	0.635	0.546	0.561	0.556	0.588	0.673
Senior	0.256	0.230	0.264	0.265	0.259	0.249	0.263
Student	0.128	0.079	0.104	0.128	0.136	0.151	0.149
Female	0.301	0.176	0.190	0.210	0.390	0.385	0.387
Married	0.694	0.735	0.730	0.710	0.685	0.670	0.653
Immigrant	0.281	0.263	0.279	0.292	0.291	0.277	0.274
<i>Family Type</i>							
Single	0.207	0.188	0.171	0.191	0.210	0.229	0.239
Couple	0.694	0.763	0.732	0.707	0.680	0.661	0.653
Lone Parent	0.063	0.042	0.061	0.062	0.069	0.070	0.070
Other	0.035	0.006	0.037	0.040	0.042	0.040	0.038
Children Present	0.412	0.496	0.455	0.438	0.374	0.382	0.363
Household Size	2.705	2.887	2.878	2.799	2.626	2.610	2.535
<i>Location*</i>							
NSW	0.340	0.353	0.350		0.342	0.327	0.331
Vic	0.261	0.282	0.272		0.257	0.257	0.248
Qld	0.182	0.152	0.163		0.184	0.195	0.204
SA	0.089	0.092	0.092		0.089	0.087	0.086
WA	0.100	0.092	0.094		0.100	0.106	0.103
TAS	0.028	0.029	0.029		0.028	0.028	0.028
Sample Proportion	1.000	0.132	0.117	0.199	0.203	0.172	0.177
Observations	30190	3989	3543	6006	6118	5197	5337

Note: * The pooled sample means for location exclude observations from 1988

Table 2. Estimated Correction Factors, Full Sample

Year	Linear Model		Quadratic Model	
1975	1.21*	1.100	1.200	1.110
	(.11)	(.08)	(.37)	(.22)
1984	1.000	1.000	1.000	1.000
	-	-	-	-
1988	1.020	0.960	1.010	0.980
	(.07)	(.06)	(.31)	(.19)
1993	0.85**	0.930	0.840	0.950
	(.06)	(.06)	(.26)	(.19)
1998	0.72**	0.84**	0.73*	0.840
	(.05)	(.04)	(.22)	(.16)
2003	0.66**	0.82**	0.67*	0.800
	(.06)	(.06)	(.21)	(.16)
<i>Covariates</i>	<i>No</i>	<i>Yes</i>	<i>No</i>	<i>Yes</i>
<i>F-Test I</i>	15.04 ^{††}	5.97 ^{††}	16.3 ^{††}	6.46 ^{††}
<i>F-Test II</i>	13.71 ^{††}	4.33 ^{††}	11.93 ^{††}	1.76
<i>Cumulative Bias Estimates (relative to 1984)</i>				
1975	21%	10%	20%	11%
1988	-2%	4%	-1%	2%
1993	15%	7%	16%	5%
1998	28%	16%	27%	16%
2003	34%	18%	33%	20%
<i>Estimates of Average Annual Bias</i>				
1975-2003	1.58%	0.89%	1.53%	0.97%
1975-1998	1.75%	1.01%	1.69%	1.04%
1984-2003	1.51%	0.85%	1.47%	0.94%
1984-1998	1.72%	1.03%	1.66%	1.03%

Notes:

Standard Errors reported in parentheses.

* Denotes individually statistically different from 1.0 at the 10% significance level and

** denotes individually statistically different from 1.0 at the 5% significance level.

†† Denotes joint significance at the 1% level.

F-Test I is the test statistic for the null that the coefficients on D_{1975} - D_{2003} are jointly equal.

F-Test II is the test statistic for the null that the coefficients on D_{1984} - D_{2003} are jointly equal.

Table 3. Estimated Correction Factors by Demographic Group

Year	Group					
	Singles	Couples no kids	Couples with kids	Lone Mothers	Working Families	Seniors
1975	1.45 (.69)	0.93 (.22)	1.03 (.16)	2.09 (1.08)	1.02 (.14)	1.11 (.22)
1984	1.00 -	1.00 -	1.00 -	1.00 -	1.00 -	1.00 -
1988	1.15 (.54)	0.93 (.22)	0.95 (.14)	1.30 (.66)	0.99 (.14)	0.98 (.19)
1993	1.01 (.48)	0.93 (.22)	0.98 (.15)	1.18 (.60)	1.00 (.14)	0.95 (.19)
1998	0.77 (.36)	1.01 (.23)	0.93 (.14)	0.82 (.41)	0.96 (.13)	0.84 (.16)
2003	0.57 (.29)	0.95 (.23)	0.90 (.14)	0.61 (.33)	0.90 (.13)	0.8 (.16)
<i>F-Test I</i>	6.93 ^{††}	0.81	2.26	7.54 ^{††}	0.88	6.46 ^{††}
<i>F-Test II</i>	3.15 [†]	1.00	0.83	2.39 [†]	0.74	1.76
<i>Cumulative Bias Estimates (relative to 1984)</i>						
1975	45%	-7%	3%	109%	2%	11%
1988	-15%	7%	5%	-30%	1%	2%
1993	-1%	7%	2%	-18%	0%	5%
1998	23%	-1%	7%	18%	4%	16%
2003	43%	5%	10%	39%	10%	20%
<i>Estimates of Average Annual Bias</i>						
1975-2003	2.28%	-0.07%	0.44%	3.30%	0.41%	0.97%
1975-1998	2.28%	-0.36%	0.42%	3.63%	0.25%	1.04%
1984-2003	1.85%	0.25%	0.49%	1.70%	0.49%	0.94%
1984-1998	1.44%	-0.07%	0.47%	1.15%	0.27%	1.03%

Notes:

Based on Quadratic specification with covariates

Standard Errors reported in parentheses.

* Denotes individually statistically different from 1.0 at the 10% significance level and

** denotes individually statistically different from 1.0 at the 5% significance level.

†† Denotes joint significance at the 1% level and † denotes joint significance at the 5% level.

F-Test I is the test statistic for the null that the coefficients on $D_{1975}-D_{2003}$ are jointly equal.

F-Test II is the test statistic for the null that the coefficients on $D_{1984}-D_{2003}$ are jointly equal.

Table 4. Estimated Correction Factors, Linear Engel Curves

	(1)	(2)	(3)	(4)	(5)	(6)
Specification	Base	Exclude 1988	Exclude 1975	Unweighted	Extra Covariates ^a	IV ^b
1975	1.100 (.08)	1.116		1.095	1.126	1.233
1984	1.000	1.000	1.00	1.00	1.00	1.000
	-	-	-	-	-	-
1988	0.960 (.06)		0.973	0.948	0.968	0.940
1993	0.930 (.06)	0.958	0.959	0.944	0.950	0.875
1998	0.84** (.04)	0.872	0.862	0.875	0.869	0.766
2003	0.82** (.06)	0.797	0.779	0.820	0.792	0.646

Notes:

Based on Linear specification with covariates

Standard Errors reported in parentheses.

* Denotes individually statistically different from 1.0 at the 10% significance level and

** denotes individually statistically different from 1.0 at the 5% significance level.

a. The additional covariates are a set of 12 indicator variables for the age of the household reference person and a dummy variable for home ownership status.

b. The log of real total expenditure is treated as endogenous and instrumented by the log of real total income.

Figure 5.1 Engel Curves - Population

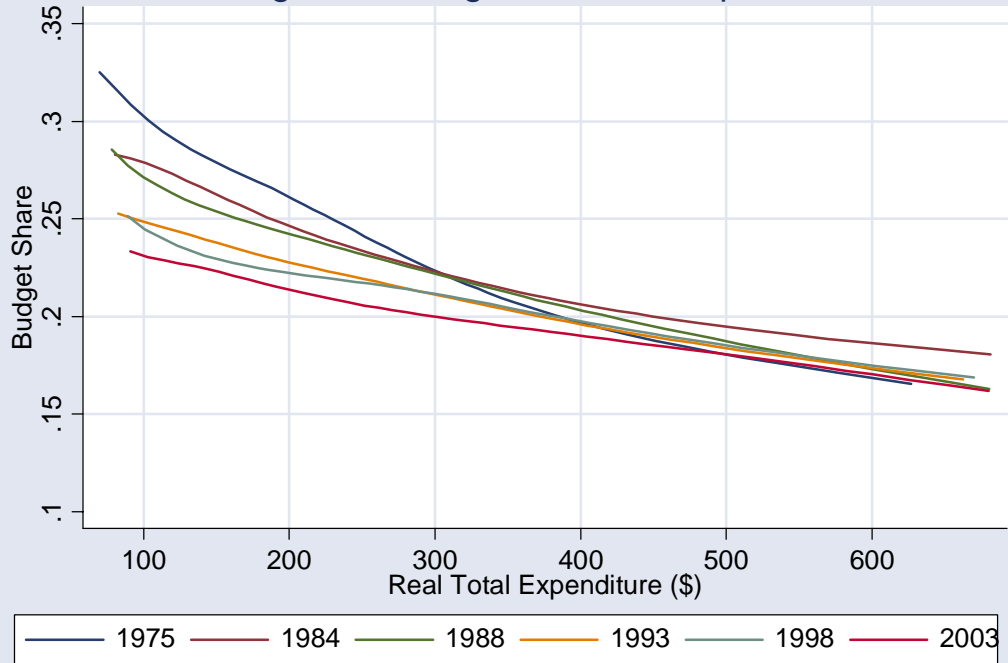


Figure 5.2 Engel Curves - Non-elderly Singles

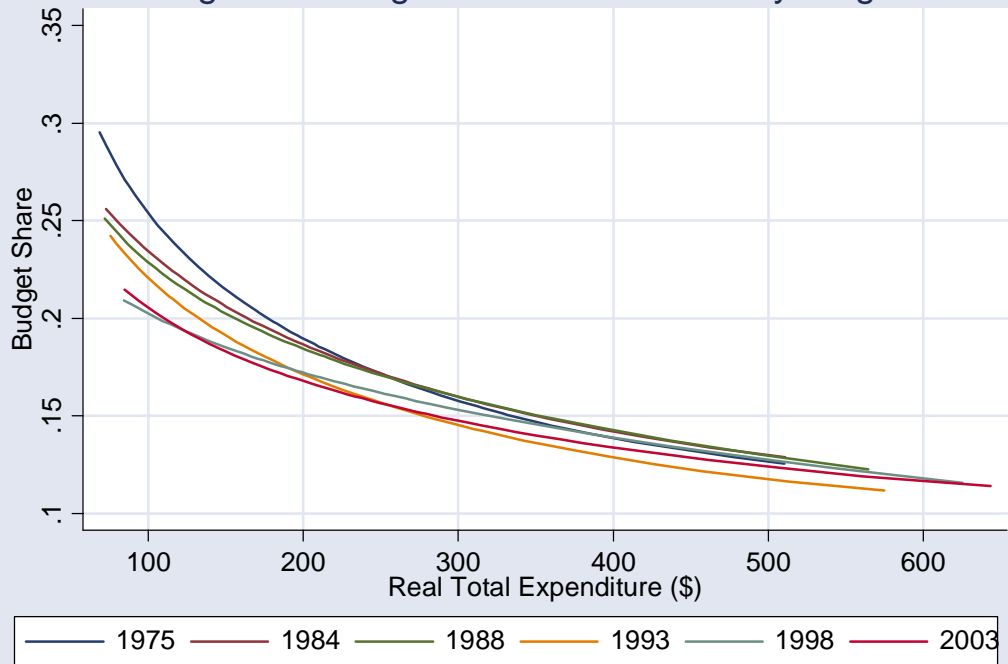


Figure 5.3 Engel Curves - Non-elderly Couples without Kids

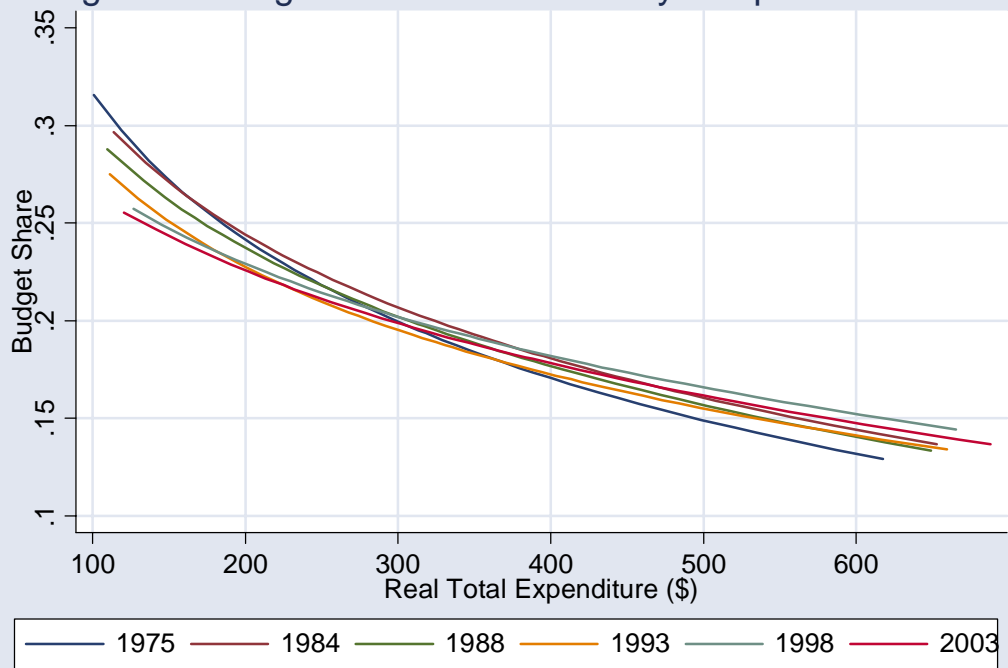


Figure 5.4 Engel Curves - Non-elderly Couples with Kids

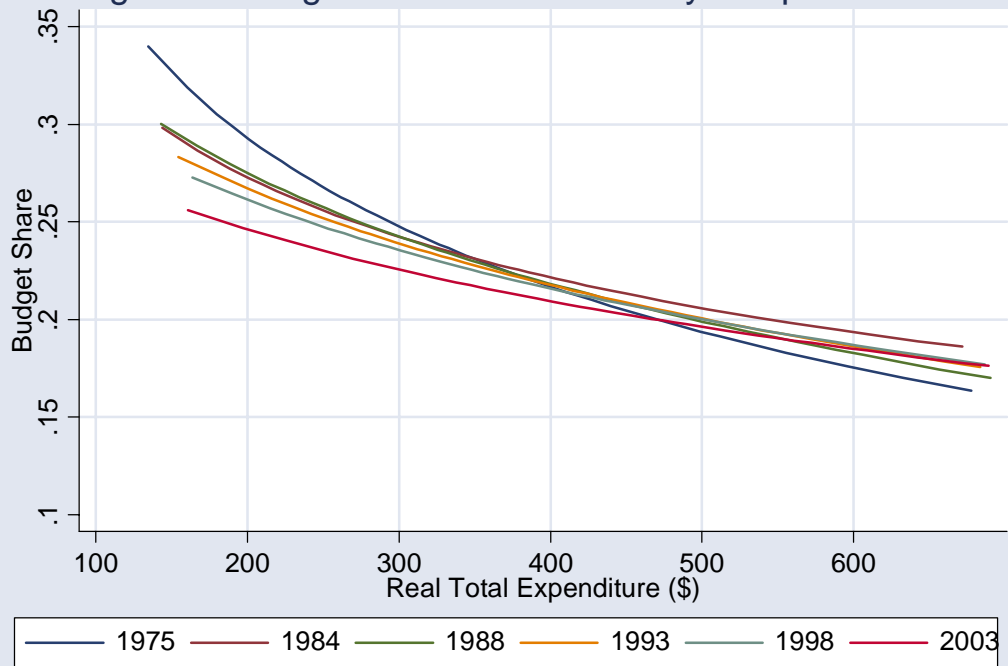


Figure 5.5 Engel Curves - Lone Mothers

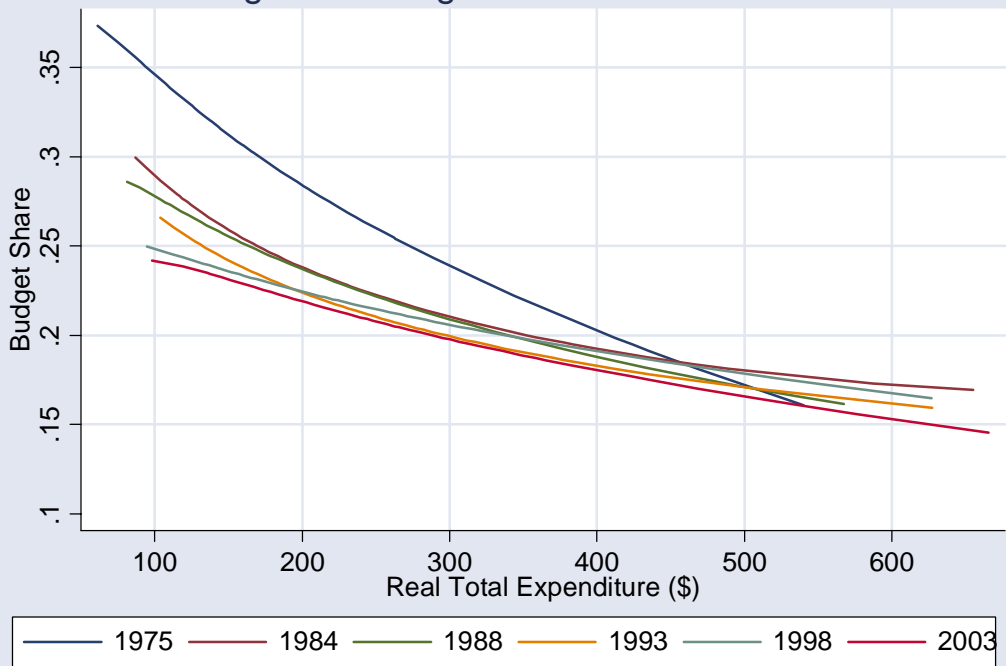


Figure 5.6 Engel Curves - Families with Kids

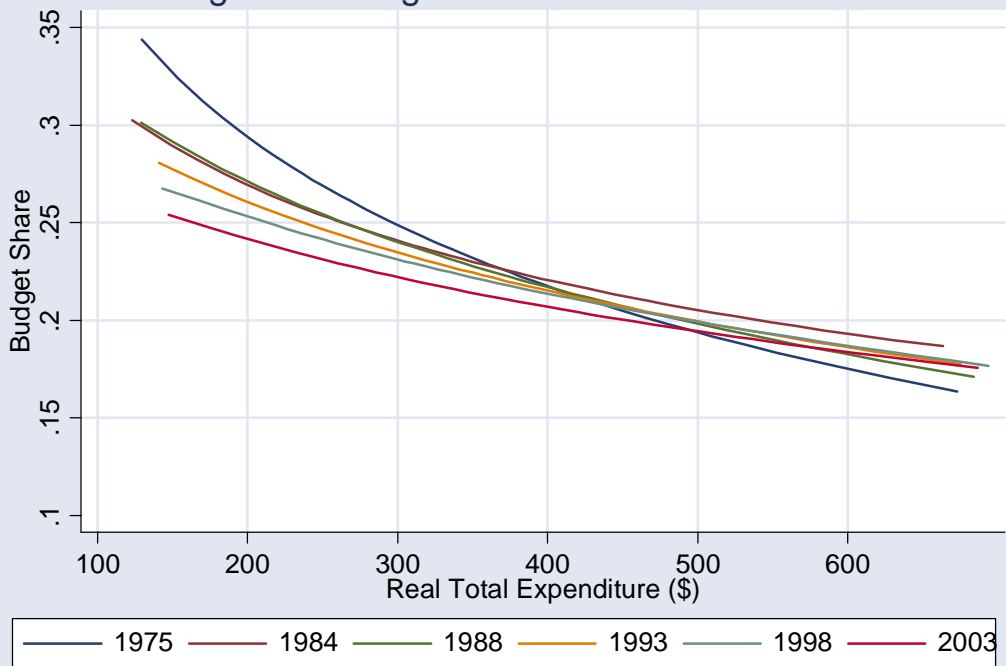


Figure 5.7 Engel Curves - Working Families

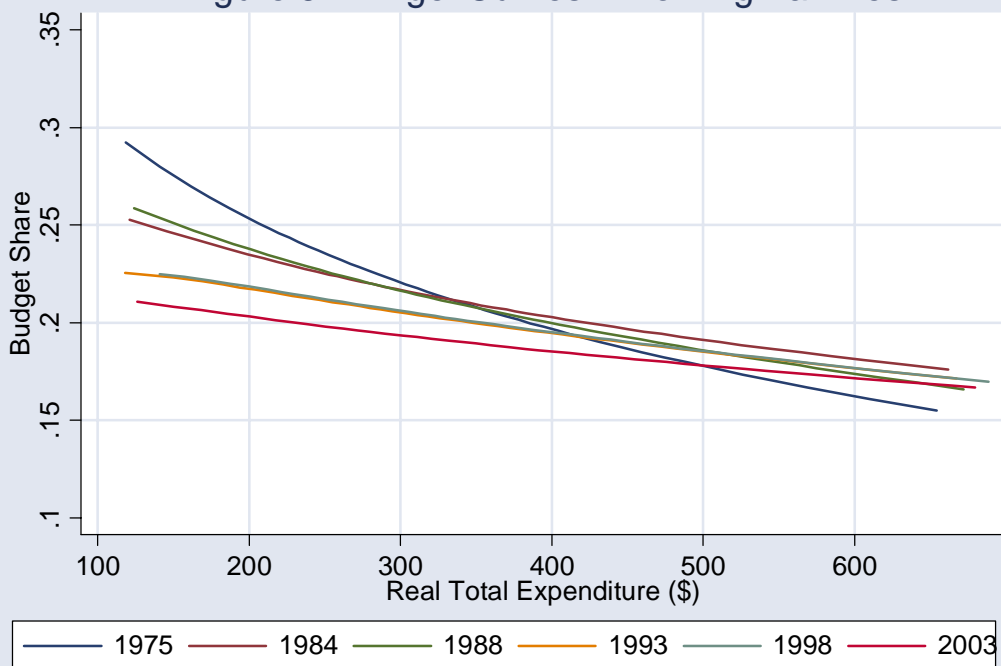


Figure 5.8 Engel Curves - Seniors

