

# Data Appendix

## Current Population Survey (CPS) Data

The CPS data are from the March Annual Demographic Files for survey years 1981 to 2000.<sup>1</sup> The BLS usually interviews around 60,000 households in any given Annual Demographic File. As stated in the text, the decision-making unit in this study is taken to be the household, and household income data is the primary CPS data employed in the study. The CPS defines a household in the following way:

A household consists of all the persons who occupy a house, an apartment, or other group of rooms, or a room, which constitutes a housing unit. A group of rooms or a single room is regarded as a housing unit when it is occupied as separate living quarters; that is, when the occupants do not live or eat with any other person in the structure, and when there is direct access from the outside or through a common hall.

Household income, broadly defined, is the sum of earned income, transfer income, asset income, and retirement income, minus state and federal taxes. Household earned income is constructed as the sum of the wage and salary income, self-employed income, and farm income of household members. Before aggregating across household members, the top-coding corrections recommend by Katz & Murphy (1992) are made for each of these three sub-components of earned income for each individual.

Prior to survey year 1990, household level transfer income is computed as the sum of several CPS income variables given at the family level,<sup>2</sup> summed over the families residing in the household. No correction for top-coding is made for any of these variables. Household level asset income and retirement income are treated in the same way. The list of family-level income variables used from survey years 1981-1989 is:

---

<sup>1</sup>The income data in a survey year refers to the previous calendar year, while the demographics information refers to the current calendar year, so the paper uses twenty years of income data, from 1980 to 1999.

<sup>2</sup>The CPS defines a family in the following way:

A family is a group of two persons or more (one of whom is the householder) residing together and related by birth, marriage, or adoption. All such persons (including related subfamily members) are considered as members of one family.

**CPS transfer, asset and retirement income variables  
for survey years 1981-1989**

- Transfer income
  - FINCUS: family income - money received from U.S. gov't. Includes social security and railroad retirement.
  - FINCSP: family income - supplemental security. Includes money received from U.S., state, and local gov't.
  - FINCPA: family income - public assistance and welfare. Includes aid to families with dependent children and other assistance.
  - FINCVP: family income - veterans payments etc. Includes veterans payments, unemployment compensation, and workers compensation.
  - FINCCS: family income - child support, etc. Includes alimony and child support, other regular contributions from persons not in household, and anything else.
  
- Asset income
  - FINCINT: family income - interest.
  - FINCDIV: family income - dividends, etc. Includes dividends, net rental income or royalties, estates or trusts.
  
- Retirement income
  - FINCRET: family income - retirement. Includes private pensions and annuities, military retirement, federal gov't employee pensions, and state or local gov't pensions.

Starting in the 1990 survey year, the CPS began reporting the various components of transfer, asset, and retirement income at the household level as well as the family level; the household level data is used from survey year 1990 onwards. Also in 1990, the CPS changed its classification system for transfer, asset and retirement income. The list of household-level income variables used from 1990 onwards is:

**CPS transfer, asset, and retirement income variables  
for survey years 1990-2000**

- Transfer income
  - HSSVAL: HHL D income - Social Security
  - HSURVAL: HHL D income - survivor income
  - HDISVAL: HHL D income - Disability income
  - HSSIVAL: HHL D income - Supplemental Security income
  - HPAWVAL: HHL D income - Public Assistance income
  - HUCVAL: HHL D income - Unemployment compensation
  - HWCVAL: HHL D income - Worker’s compensation
  - HVETVAL: HHL D income - Veteran Payments
  - HCSPVAL: HHL D income - child support
  - HALMVAL: HHL D income - alimony
  - HFINVAL: HHL D income - Financial Assistance income
  - HEDVAL: HHL D income - Education income
  - HOIVAL: HHL D income - Other income
  
- Asset income
  - HINTVAL: HHL D income - Interest income
  - HDIVVAL: HHL D income - dividend income
  - HRNTVAL: HHL D income - Rent income
  
- Retirement income
  - HRETVAL: HHL D income - Retirement income.

The paper uses the NBER’s TAXSIM program to estimate taxes paid. Given demographic and income information on a tax unit in any given year, the TAXSIM program computes that tax unit’s state and federal tax burden. Each household is treated as a tax unit, which may bias upwards the tax burden under the progressive U.S. income tax system, given that several returns may be filed separately within a household.

The TAXSIM program requires some data that is fairly straightforward to pull from the CPS files, such as the state of residence of the household, the marital status of the householder,<sup>3</sup> the number of children and elderly in the household, the wage and salary income of the householder and spouse, the household’s dividend income (FINCDIV and HDIVVAL), its pension income (FINCRET and HRETVAL), and its gross social security income (FINCUS and HSSVAL, HSURVAL, and HDISVAL). The construction of some other variables for TAXSIM requires more decisions. For a variable described as “Other property income, including interest, self-employment, may be negative. Also alimony, fellowships, and other taxable income,” we include the household’s farm and self-employment income, its interest income (FINCINT and HINTVAL), and some of its transfer income (FINCCS pre-1988 and HCSPVAL, HALMVAL, HFINVAL, and HEDVAL post-1988). In this category, we also include the wage and salary income of members of the household other than the householder and spouse of householder.<sup>4</sup> For a variable described as “Other non-taxable transfer income such as welfare, municipal bond interest, child support that would affect eligibility for state property tax rebates,” we include, before 1988, FINCSP, FINCPA, and FINCVP, and after 1988, HSSIIVAL, HPAWVAL, HUCVAL, HWCVAL, and HVETVAL.<sup>5</sup>

Finally, TAXSIM includes some fields where one can input information to compute rebates and deductions, information such as medical expenses, charitable contributions, child care expenses, rent paid, and property taxes paid. Since this information is unavailable from the CPS, we set these fields to zero.<sup>6</sup>

The primary sample of households we consider excludes:

1. households without a male head aged 23 to 59 (without a male householder or a male spouse of householder of that age), and

---

<sup>3</sup>Heads of households are called “householders” by the CPS over this time period

<sup>4</sup>Finally, since dividend income can be negative in the CPS, and the TAXSIM program requires it to be strictly positive, negative dividend income is subtracted here and set to zero in the dividend income field.

<sup>5</sup>TAXSIM provides a separate field for unemployment compensation, but prior to 1988 unemployment compensation is not separated from other transfer income in the CPS variable FINCVP. For consistency, the separate field is left at a value of zero for the whole sample and unemployment compensation is always included in the “other non-taxable income” field.

<sup>6</sup>Such information is available for households sampled in the CEX, through. One test of the likely impact of excluding the information on deductions is to run the CEX data through TAXSIM with and without the data on deductions, and compare the different average tax payments for the synthetic cohort data. The difference between the two sets of average tax payments was minor, less than an order of magnitude of the size of the average tax payment for most synthetic cohort cells.

2. households residing in group quarters.

Synthetic cohorts are constructed by the male head of household’s education and five-year birth cohort. The birth year variable is designed to run from March to February, reflecting the fact that the survey is taken in March. If any member of a five-year birth cohort violates the sample selection restriction on age in a given year, that cohort is excluded from the sample for that year. For example, a five-year birth cohort may contain heads aged 55, 56, 57, 58 and 59 at year  $t$ , in which case it would be excluded from the year  $t$  sample. The four education categories are: (i) less than 12 years of schooling, (ii) 12 years of schooling, (iii) more than 12 but less than 16 years of schooling, and (iv) 16 or more years of schooling.

### **Consumer Expenditure Survey (CEX) Data**

The CEX data are taken from the 1980 to 1999 Interview Survey files. The unit of analysis in the CEX is the Consumer Unit (CU), defined as a group of individuals who live together and are either “related by blood, marriage, adoption, or other legal arrangement” or pool expenditures on 2 out of the following 3 expenditure categories: food, housing, or other living expenses. We equate CUs with households throughout this study, and use the terms interchangeably. Each is interviewed up to a maximum of 5 times on a quarterly basis, although no data from the first interview is published on the Interview Survey files. Households continuously rotate in and out of the survey, and about 5,000 households are in the process of being interviewed at any time.

This study primarily employs CEX data on household expenditures and the demographic characteristics of household members. The expenditures data (extracted from the interview survey MTAB files) is monthly, and normally covers each of the three months prior to the month of the interview.<sup>7</sup> The demographics data (from the interview survey FMLY files) are collected at each interview, and hence are quarterly and current at the time of the interview. These data are converted to a monthly frequency by assigning the data values for a particular interview to the month of the interview and the preceding two months, or the preceding five months if a household skipped the preceding interview or if the interview is the first one for that household. These data are then merged with the data on household expenditures.

---

<sup>7</sup>In some cases, especially for the last interview, consumption data is also available for the month of the interview in addition to the three preceding months, and occasionally an interview collects information on consumption for four or five preceding months.

A household's non-durables and services consumption is the sum of its expenditures on 16 sub-categories constructed following Orazio Attanasio's classification system.<sup>8</sup> These sub-categories are:

- food consumed in the home
- food consumed out of the home
- alcohol
- tobacco
- housekeeping services
- home maintenance
- fuel oil, coal, bottled gas, wood, kerosene and other fuels
- electricity and natural or utility gas
- public utilities
- telephone services
- fuel for transportation
- transportation equipment maintenance and repair
- vehicle rental and misc. transportation expenses
- public transportation
- personal care services
- non-durable entertainment expenses.

In 1982 and 1988, there were significant changes in the CEX survey questions covering food consumed at home. From 1980 to 1981, the surveyors asked the household how often it shopped for groceries, and asked what was the usual amount spent per shopping outing. From 1982 to 1987, the surveyors asked for the household's usual monthly

---

<sup>8</sup>The raw CEX data reports consumption by UCC (universal classification code); there are several hundred of these categories. A list of the UCC codes that comprise each of the sub-categories is available from the author on request.

expense on groceries, and from 1988 onwards they asked for the usual weekly expense. In addition, BLS statisticians have indicated that there were changes in how the data was processed at these break-points. The result was a large spike downwards in the food consumed at home data in 1982, and a large spike upwards in 1988 - see the aggregate data in figure B.1.

To correct these evident data problems, this paper assumes the effect of the survey changes was to scale up or down food at home expenditures of all households by the same amount. This would be the case if some survey regimes cause all households to mis-estimate their frequency of shopping by some fraction, say, 10 percent. These fractions are estimated by regressing household log real food at home expenditures on two dummy variables, one covering the 1980-81 time period, the other 1982-1987, and the log of an explanatory expenditure variable - real non-durable goods and services less food at home.<sup>9</sup> Real food at home expenditures are deflated by the CPI food at home deflator. The deflator for the explanatory expenditure variable is constructed as a geometric weighted average of the CPI deflators of its 15 sub-categories, with nominal household specific consumption shares as weights (i.e. a Stone price index with household-specific weights). The list CPI deflators that are matched to each CEX expenditure sub-category is:<sup>10</sup>

CEX category	CPI categories
food (home)	food at home (SAF11)
food (away from home)	food away from home (SEFV)
alcohol	alcoholic beverages (SAF116)
tobacco	tobacco and smoking productsp (SEGA)
housekeeping services	household furnishings and operations (SAH3)
home maintenance	housekeeping supplies (SEHN)
fuel oil, coal, etc.	fuel oil and other fuels (SEHE)

---

<sup>9</sup>criticized as a device for estimating Engel curves, due to its inability to accomodate the budget constraint (see Deaton and Muellbauer (1980)), the specification yields reasonable corrections. A more complicated alternative, not pursued in this paper, would be to estimate corrections in the context of a fully specified demand system.

<sup>10</sup>The CPI data were downloaded using the selective access program at [www.bls.gov](http://www.bls.gov) (as of October 2001 the download program name had been changed - it was called "create customized tables (multiple screens)", at [www.bls.gov/cpi/home.htm#data](http://www.bls.gov/cpi/home.htm#data)). The codes next to each CPI category are the post-1998 revision CPI codes. Where multiple CPI deflators were used for a single category, the price deflator was constructed as an unweighted arithmetic average of the multiple deflators.

electricity and gas	gas (piped) and electricity (SEHF)
public utilities	water and sewerage mainenance (SEHG01) and garbage and trash collection (SEHG02)
telephone services	telephone services, local charges (SEED01) and interstate toll calls (SS27051) and intrastate toll calls (SS27061)
transport fuel	motor fuel (SETB)
transp. maintenance	motor vehicle parts and equipment (SETC) and motor vehicle maintenance and repair (SETD)
misc. transp. expenses	private transportation (SAT1)
public transportation	public transportation (SETG)
personal care services	personal care products (SEGB) and personal care services (SEGC)
entertainment expenses	admissions (SERF02) and fees for lessons or instructions (SERFO3)

Unfortunately, this specification does not capture well a major feature of the data: the steady decline over time in the share of food at home in non-durable goods and services expenditures. The first panel of figure B.2 shows the NIPA log ratio of real food at home to the rest of real non-durable goods and services;<sup>11</sup> the second panel plots the ratio using the uncorrected CEX data, which is aggregated by taking the sum across households of the log expenditure variables. To address this phenomenon in the regression correction, we include as explanatory variables a time trend and a time trend interacted with our explanatory expenditure variable. We also experimented with including in the regression polynomials in the time trends and explanatory expenditure variable; this made little difference to the correction.

The dummy variable for 1980-81 took on a value of -0.044, while the dummy for 1982-1987 took on a value of -0.178. Table B.3 shows the raw and corrected real food consumed at home data, both in logs and in log ratio form. The corrections seem reasonable.

The study makes use of the income data from the CEX, in tables 4 and 5. These data come from the interview survey FMLY files and the interview survey MEMB files. The

---

<sup>11</sup>The composition of NIPA non-durable goods and services is matched as closely as possible to the CEX composition described above; the price indices used to deflate the data NIPA data are the CPI deflators described above.



CEX asks households questions about their income from the previous twelve months at the second interview (the first for which data is available) and fifth interview; in this paper, only income data from the fifth interview is used. The income data is extrapolated to a monthly frequency by assigning the fifth interview data to the month of the interview and the previous twelve months, and is then merged with demographics data from the FMLY files extrapolated to monthly frequency as before.

Following the CPS construction, household income is constructed as the earned income of its members, plus other household income. The three components of earned income of household members are wage and salary income (SALARYX), self-employed income (NONFARMX), and farm income (FARMINCX), and are drawn from the CEX MEMB files. Before aggregating across household members, top-coding corrections were made similar to those made to the CPS data. Before tax household income is then constructed as the FMLY file variable FINCBTAX (the sum of earned income, asset income, and various forms of transfer and retirement income) minus earned income without top-coding adjustments plus earned income with top-coding adjustments.

Although the CEX asks questions on taxes paid by households, this paper follows the CPS construction and uses the NBER's TAXSIM program to estimate taxes paid for each household. The primary difference from the CPS computations is that data on medical expenses, charitable contributions, rent paid, and state, local and property taxes paid, are all available for CEX households, allowing us to account for the impact of deductions on the total tax burden. The fraction of CEX households who itemized in our computations approximately matched the fraction in the population (private correspondence with Daniel Feenberg). After-tax income for CEX households was computed as the adjusted FINCBTAX variable minus the TAXSIM estimate of taxes paid.

The primary sample of CEX households we consider in this paper excludes:

1. households without a male head aged 23 to 59 (without a male householder or a male spouse of householder of that age),
2. households living in rural areas,<sup>12</sup>
3. households providing an incomplete income response,<sup>13</sup>

---

<sup>12</sup>This selection makes the CEX samples comparable over time, since non-urban consumers were not sampled in 1982-3.

<sup>13</sup>The CEX considers data given by incomplete reporters to be of low quality.

4. households whose reported head either ages by more than one over the sample period of 4 quarters, changes sex, or changes education,<sup>14</sup>
5. monthly household observations with non-positive non-durables and services consumption, and
6. households residing in student housing.

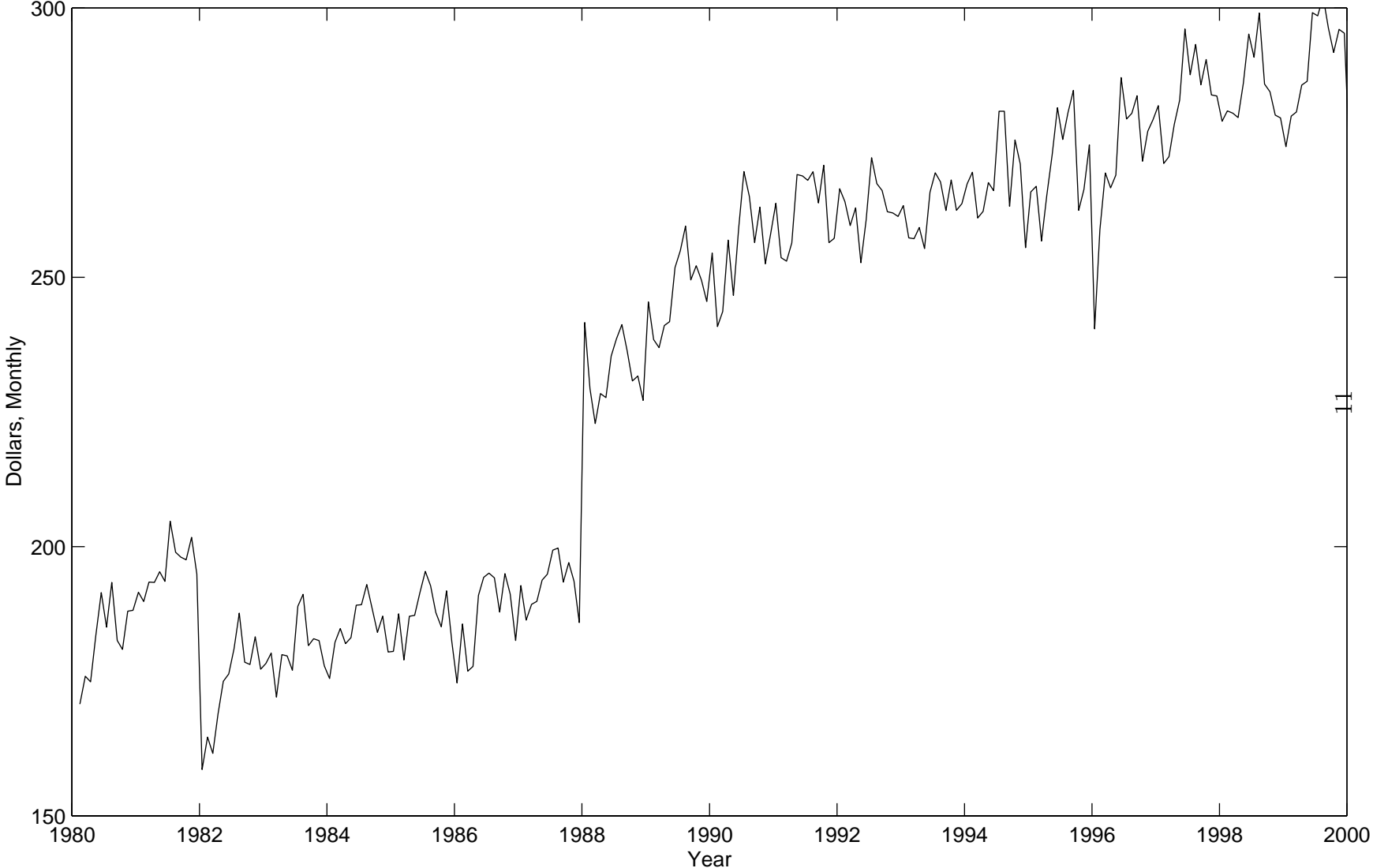
Synthetic cohorts are constructed in the same manner as with the CPS data.<sup>15</sup> The four education classifications are defined as: (1) less than high school graduate, (2) high school graduate but no subsequent schooling, (3) some college, and (4) 4-year college degree or more schooling.

---

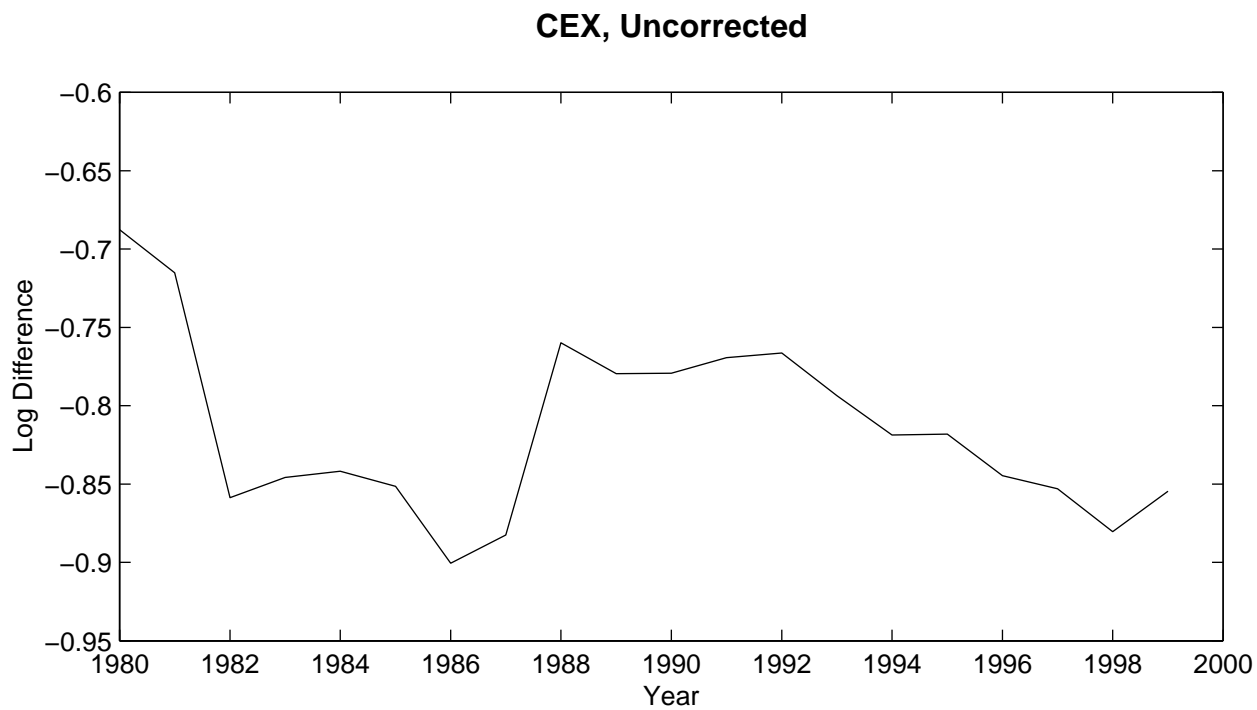
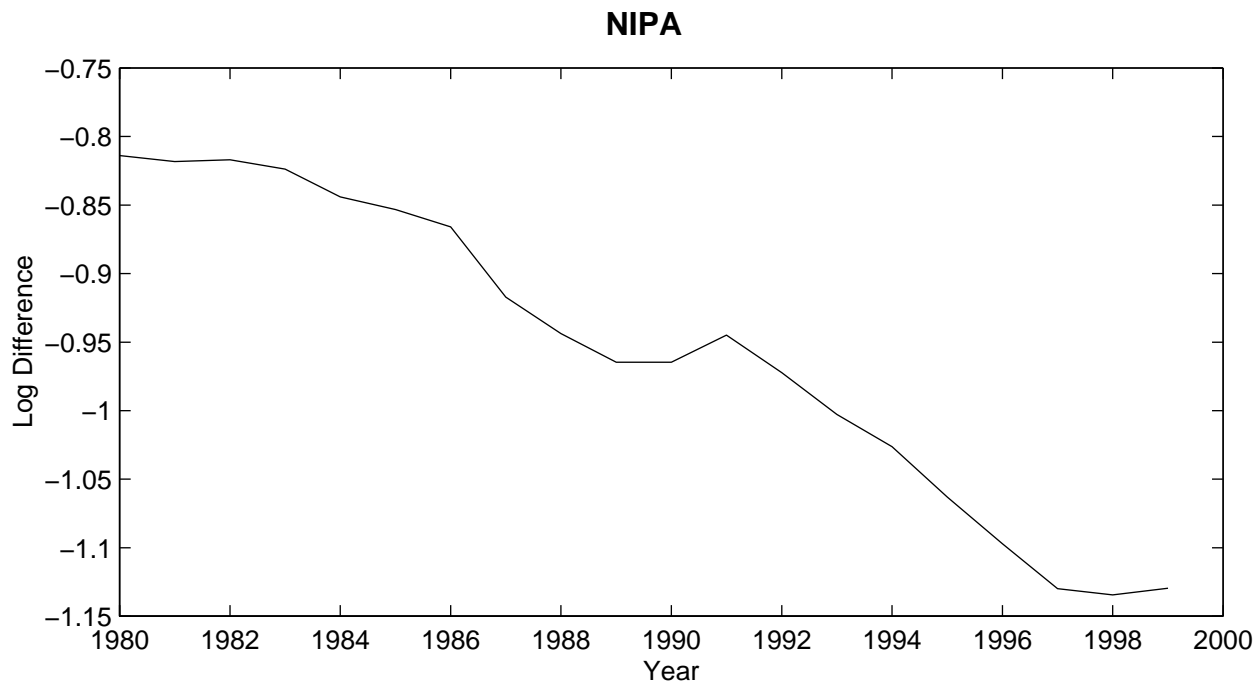
<sup>14</sup>These restrictions generally are meant to eliminate households whose reference person changes due to death or other circumstances, while the last restriction serves to eliminate households engaged in schooling.

<sup>15</sup>If the age of the head changes between quarterly interviews, we can usually pin down the birth year, but if not, we can often only pin down a set of possible birth months that may span two years. A reference person is assigned to birth year  $t$  rather than birth year  $t - 1$  if more than half the possible birth months of the head are in year  $t$ . For consistency with the CPS, birth years are defined to run from March to February.

**Figure B.1: CEX Food at Home Expenditures, Raw Data**



**Figure B.2: Log (Real Food at Home/Real Non-Durable Goods and Services)**



**Figure B.3: Corrected (Dashed) and Uncorrected (Solid) CEX Real Food at Home Data**

