

# “Imputing Consumption in the PSID Using Food Demand Estimates From CEX”

by Blundell, Preston and Pistaferri (2006, Working Paper)  
presented by Taisuke Nakata

January 19, 2010

- **Panel Study of Income Dynamics (PSID)**: Panel data on income as well as consumption on select items (food, rent, etc.).
- **Consumer EXpenditure Survey (CEX)**: Repeated cross-sectional data on consumption and income.
- Nice to have panel data on both income and total consumption.
- Some research use food consumption, instead of total consumption (Hall and Mishkin (1982), Altonji and Siow (1987), etc.)

- Propose a method to impute total consumption in PSID.
- Investigate the asymptotic properties of sample mean and variance of imputed consumption.
- Compare sample mean and variance of imputed PSID-consumption with those of actual CEX-consumption.

- Skinner (1987) use OLS to predict total consumption with food and other items.
- Ziliak (1998) impute consumption as the difference between income and imputed savings.

- Exogenous Variables:  $D_{i,t}$ ,  $Y_{i,t}$  and  $p_t$ .
- Endogenous Variables:  $C_{i,t}$ ,  $F_{i,t}$ .
- (Implicit) Model of Consumption: Mapping from  $\{D_{i,s}, Y_{i,s}, p_s\}_{s=0}^t$  to  $C_{i,t}$  and  $F_{i,t}$ .
- Dataset X: Draws  $N_X$  people every period independently. Record  $D_{i,t,X}$ ,  $F_{i,t,X}$ , and  $C_{i,t,X}$ .
- Dataset P: Draws  $N_P$  people at time 0 and follows them. Record  $D_{i,t,P}$ ,  $F_{i,t,P}$ , and  $Y_{i,t,P}$ .

$$A_{i,t,k} = A_{i,t} + \text{measurement error for any variable } A \text{ and } k = X, P$$

# Imputation: Food Demand Equation

$$C_{i,t} = C_t(\{D_{i,s}, Y_{i,s}, p_s\}_{s=0}^t)$$

$$F_{i,t} = F_t(\{D_{i,s}, Y_{i,s}, p_s\}_{s=0}^t)$$

$\Rightarrow$

$$F_{i,t} = D_{i,t}\beta + p_t\theta + \gamma(D_{i,t})C_{i,t} + u_{i,t}$$

where  $Cov(C_{i,t}, u_{i,t}) \neq 0$

Imputation Method:

Step 1: Estimate  $\beta$ ,  $\theta$ , and  $\gamma$  with  $F_{i,t,X}$ ,  $D_{i,t,X}$  and  $Z_{i,t,X}$  (instruments for  $C_{i,t,X}$ ).

Step 2:  $\hat{C}_{i,t,X} = \frac{1}{\hat{\gamma}}[F_{i,t,P} - D_{i,t,P}\hat{\beta} - p_t\hat{\theta}]$

-Sample Mean of  $\hat{C}_{i,t,P}$  is a consistent estimator of  $E[C_{i,t}]$ .

$$\begin{aligned}\text{plim}M(\hat{C}_{i,t,P}) &= \text{plim}M(C_{i,t,X}) + \frac{1}{\gamma}[\text{plim}M(F_{i,t,P}) - \text{plim}M(F_{i,t,X})] \\ &= E[C_{i,t}]\end{aligned}$$

where

- $M(a) = \frac{1}{N} \sum_{i=1}^N a_i$  for any variable  $a$ ,
- probability limit is taken with respect to  $N_X$  and  $N_P$ .

## Asymptotic Property: Sample Variance $\hat{C}_{i,t,P}$

-Sample Variance of  $\hat{C}_{i,t,P}$  is NOT a consistent estimator of  $\text{Var}(C_{i,t})$ , but its first difference is a consistent estimator of the first-difference in  $\text{Var}(C_{i,t})$ .

$$\begin{aligned}\text{plim } V(\hat{C}_{i,t,P}) &= \text{plim } V(C_{i,t,X}) + \frac{1}{\gamma^2} \text{plim}[V(F_{t,P}) - V(F_{t,X})] + \text{const.} \\ &= \text{Var}(C_{i,t}) + \text{const.}\end{aligned}$$

where

- $V(a) = \frac{1}{N} \sum_{i=1}^N (a_i - M(a))^2$  for any variable  $a$ .
- constant term depends on variance of measurement errors, etc.



$$F_{i,t} = D_{i,t}\beta + p_t\theta + \gamma(D_{i,t})C_{i,t} + u_{i,t}$$

- PSID and CEX from 1980-1986 and 1989-1992 (annual data).
- A sample of continuously married couple headed by a male aged between 30 and 65.
- $C_{i,t}$ : non-durable consumption.
- $F_{i,t}$ : food consumption at home and food consumption away from home.
- $Z_{i,t,X}$ : Cohort-education-year specific average of husband's earnings and cohort-education-year specific average of wife's earnings (and their interaction with time, education, kids, etc.)
- $p_t$ : price of food, alcohol, fuel, utilities, and transport.

# Application: Sample Mean of $\hat{C}_{i,t,P}$ and $C_{i,t,X}$

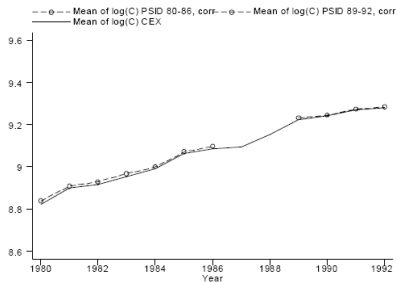


Figure 5: The mean of CEX consumption and the mean of “corrected” PSID imputed consumption.

# Application: Sample Variance of $\hat{C}_{i,t,P}$ and $C_{i,t,X}$

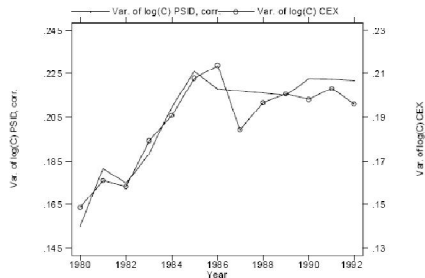


Figure 7: The variance of CEX consumption and the variance of PSID-imputed consumption (corrected for cross-dataset differences in food spending).

- PSID started collecting  $C_{i,t,P}$  in 2005.
- It would be useful to compare  $\hat{C}_{i,t,P}$  and  $C_{i,t,P}$ .

- Started in 1968.
- Survey conducted annually until 1996, and biannually thereafter.
- Unit of analysis is household.
- “Core sample” is representative of U.S. population, and “SEO sample” is representative of low-income families. “Latino sample” and “Immigrant sample” were added in 1990 and 1997.
- The interview format changed in 1993 from a pencil-and-paper telephone format to a computer-assisted telephone format.
- In addition to detailed information on earnings, it collects information on demographic variables, expenditure on select consumption items.
- Households report their taxable income (which includes transfers and financial income).
- The baseline analysis of BPP excludes income from financial assets and subtracts federal taxes on non-financial assets.
- Have asked expenditure on food and rent most of the times.
- Started asking expenditures on more items in 1999 (covering 70 percent of total expenditure).
- Started asking expenditures on almost all items in 2005 (covering 95 percent of total expenditure).
- Top-coding affects very few observations.

- Started in 1960. Consistently available from 1980. Monthly Observations.
- Unit of analysis is household.
- Rotating short-panel. But, repeated cross-section at annual frequency.
- Designed as a tool for periodic revision of the CPI.
- Consists of “Interview Survey (IS)” and “Diary Survey (DS)” (IS commonly used in economic research)
- IS consists of five interviews that are three months apart.
- In addition to detailed information on expenditure, CEX collects information on demographic variables, income, hours worked, taxes paid, etc.
- Top-coding is a serious concern.