Economics Letters 23 (1987) 213–216
North-Holland

A SUPERIOR MEASURE OF CONSUMPTION FROM THE PANEL STUDY OF INCOME DYNAMICS

Jonathan SKINNER *

University of Virginia, Charlottesville, VA 22901, USA
National Bureau of Economic Research, Cambridge, MA 02138, USA

Received 8 October 1986

This paper provides a simple technique for assigning a measure of total consumption to households in the Panel Study of Income Dynamics (PSID). Using regressions from the Consumer Expenditure Surveys of 1972–1973 and 1983, linear predictions of total consumption are estimated which are stable over time and explain up to 78 percent of the total variance in consumption.

1. Introduction

A number of empirical studies have used the Panel Study of Income Dynamics (PSID) to examine the behavior of consumption at the household level [Hall and Mishkin (1982), Altonji and Siow (1985), Hotz, Kydland and Sedlacek (1985), Zeldes (1985), Altonji (1986)]. Because the PSID does not report total consumption, these researchers have used instead food expenditures as a proxy for consumption. The degree to which empirical results using food consumption generalizes to total consumption is not clear. For example, Zeldes (1985) does not reject Euler equation restrictions for high-income family food consumption, while preliminary results from Flavin (1986), using a different panel survey, do reject the restrictions for *total* consumption.

Studies using the PSID need not limit themselves to food consumption. Other components of consumption reported in the survey include the market value of owned homes, rent, utilities, number of automobiles, and a breakdown of food eaten at home and away from home. A linear prediction of total consumption based on these measures provides a more accurate measure of total consumption than simply using food consumption. Following the procedure developed by Hamermesh (1984), the Consumer Expenditure Surveys (CEX) from 1972–1973 and 1983 are used to estimate the coefficients of this prediction. It is shown in the CEX regressions that food alone explains as little as 26 percent of the variance in consumption; adding the additional consumption measures increases the explanatory power to 78 percent. Predicted consumption estimated using 1972–1973 data appears to be stable over time. The regression coefficients reported below provide the researcher with an accurate and easily implemented measure of consumption for the PSID.

* I am grateful to James Berkovec, Maxim Engers, Marjorie Flavin and William R. Johnson for helpful comments.

2. Empirical estimates of consumption

Predicting total consumption from a limited number of its components is best viewed as an omitted variable problem. The 'true' model is that total consumption is the sum of its parts; the coefficient of each element is identically one. In a regression of total consumption on a subset of consumption components, the estimated coefficients will be functions of the covariance between the included and excluded variables. The 'true' coefficient on food eaten away from home is one (despite its high income elasticity), but the estimated coefficient can exceed one if it proxies for similar but omitted categories (e.g., expensive vacations).

Consider first the regressions using the 1972–1973 CEX. Variables available from the PSID which match with the CEX are food consumed at home, food away from home, the market value of the house, rent paid, number of automobiles (up to two), size of the family (up to seven) and age. Consumption is defined to be total consumption expenditures *minus* housefurnishings and equipment, purchases of automobiles, motorcycles and boats, and mortgage payments, *plus* the imputed rental value of the house, assumed to be 6 percent of its market value. [1] Observations were excluded if consumption was less than $2,000 or greater than $60,000, if food at home was 0, if age exceeded 64 or was less than 19, and if the market value of the home was not reported or exceeded $100,000 in 1972–1973 dollars.

The first question is whether food consumption is a good proxy for total consumption. Simply replacing total consumption by food consumption corresponds to the regression $C = \alpha C_{\mathrm{f}}$, where C is total consumption and C_{f} food consumption. Using the 1972–1973 CEX, the regression estimate for α is 3.456 (t-statistic of 291.7) with $\bar{R}^2 = 0.262$. (All $\bar{R}^2$ are defined with respect to deviations from the mean.) Since many studies use log transformations of C_{f}, an alternative model, $C = \alpha C_{\mathrm{f}}^{\beta}$, yields a log-linear regression equation $\ln(C) = 6.209 + 0.360 \ln(C_{\mathrm{f}})$, with t-statistics of 226.9 and 97.3, respectively, and $\bar{R}^2 = 0.395$. Since the estimated β is less than one, the variability of food consumption may overstate the true variability in total consumption.

Eq. (1) in table 1 includes food away from home, food at home, rent and house value. Suppressing the constant term imposes linear homogeneity, which frees the researcher from adjusting for inflation if the predictions are stable over time (a topic discussed below) and used in other years of the PSID. With these four measures, $\bar{R}^2$ rises to 0.724. Eq. (2) introduces a constant term and the number of automobiles, leading to $\bar{R}^2 = 0.759$.

Eq. (3) includes utility payments (which are unavailable from the PSID during 1973–1976), which increases $\bar{R}^2$ to 0.777. Finally, a homeownership dummy, family size, and the household head's age (not reported) had almost no effect on $\bar{R}^2$. The results differed little when the sample was separated into married families and single female household heads. The $\bar{R}^2$ dropped to 0.55 with a log-linear specification of eq. (3).

The consumption regressions for 1983 are calculated using recent CEX survey information from the third wave of 1983, which asked questions about consumption during the second quarter of 1983. These consumption data were matched with family information asked during the second quarter of 1983. Families were excluded if they consumed less than $2,000 or more than $100,000, or if the consumption categories were topcoded because they hit the upper limit of reporting (3 percent of the sample were topcoded); a total of 3433 observations remained. Consumption was constructed in the same way as for 1972–1973.

Eq. (4) reports coefficients for the simple four-component consumption regression which imposes

[1] Since the house value is also included as an independent variable, assuming a 10 percent return of housing services (for example) would simply increase the coefficient on house value by 4 percentage points and leave other coefficients unchanged.

linear homogeneity. Eq. (5) adds a constant term and the number of automobiles, while the addition of utility payments increases the $\bar{R}^2$ to 0.734 (eq. 6). Coefficients for family size, age of head and a homeowner dummy variable (unreported regression) are insignificant.

I adopt the 1972–1973 coefficients as the baseline measures because they were estimated using annual, rather than quarterly data on a sample four times the size of the 1983 sample. Can the 1972–1973 coefficients be used to predict consumption in other years, such as 1983? First, an F-test as to whether the set of coefficients in eq. (3) are equal to those in (6) is strongly rejected. However, since the primary interest is in the predictive power of the two sets of coefficients, an alternative test is whether the independent variables can 'explain' 1983 consumption conditional on the prediction which uses 1972–1973 coefficients. That is, let $\hat{C}_1$ be the prediction using eq. (3) from 1972–1973 data (with the auto and constant term adjusted by the CPI) for the 1983 sample. In a model $C = \hat{C}_1 + \gamma x$ estimated using 1983 data, where γ and x are vectors of coefficients and independent variables, respectively, the null hypothesis is $\Sigma\gamma\bar{x} = 0$, where $\bar{x}$ is the mean value of x. Comparing eq. (3) with eq. (6), the null hypothesis was rejected with a t-statistic of 3.7.

Despite the formal rejection, the magnitude of the difference is small. To see this, let $\hat{C}_2$ denote the predictions based on eq. (6). Regressing $\hat{C}_2$ on $\hat{C}_1$ using 1983 data yields a coefficient of 0.988 (with the constant suppressed) and $\bar{R}^2 = 0.994$. A similar story holds for the alternative consumption specifications; in each case, the $\bar{R}^2$ exceeded 0.99. These tests suggest that the 1972–1973 coefficient estimates can be used with confidence in other years.

Table 1
Consumption regressions, 1972–1973 and 1983. [a]

Equation Year	(1) 1972–1973	(2) 1972–1973	(3) 1972–1973	(4) 1983	(5) 1983	(6) 1983
Food (home)	1.930 (133.5)	1.552 (94.1)	1.418 (86.8)	2.250 (47.2)	1.785 (33.4)	1.536 (29.1)
Food (away)	2.928 (88.8)	2.639 (84.0)	2.604 (86.0)	3.401 (40.5)	3.097 (37.7)	3.026 (38.6)
Value of house	0.1374 (120.9)	0.1137 (95.7)	0.0988 (80.8)	0.1253 (51.6)	0.1044 (39.8)	0.0936 (36.3)
Rent	1.828 (117.3)	1.473 (81.5)	1.538 (87.9)	1.702 (33.9)	1.254 (21.5)	1.305 (23.4)
Utilities			2.257 (34.1)			1.953 (18.4)
Number of automobiles		716.2 (24.8)	624.6 (26.0)		1661.6 (12.5)	1338.7 (10.4)
Constant		582.5 (12.8)	110.1 (2.4)		1290.0 (5.7)	42.8 (0.2)
$\bar{R}^2$	0.724	0.759	0.777	0.680	0.708	0.734
S.E.E.	2043	1909	1836	5267	5034	4802
N	14,499	14,499	14,499	3431	3431	3431

[a] Dependent variable is consumption; absolute value of t-statistics in parentheses.

3. Discussion

This paper has presented a simple technique for assigning a fitted measure of total consumption for use in the Panel Study of Income Dynamics. Linear prediction weights were estimated using the Consumer Expenditure Surveys of 1972–1973 and 1983. Predicted total consumption using the 1972–1973 coefficients explain a large proportion of total consumption, and can be used in other years of the PSID.

There are at least three possible criticisms of this method. The first is that the optimal (implicit) weighting scheme for 'multiple indicators' of consumption should be estimated simultaneously with the researcher's model using only PSID data. However, most estimation procedures are already expensive, complex and sensitive to specification error. Adding additional parameters to the estimation process seems wasteful, especially since the consumption weights have little economic content.

The second is that food is a better measure of consumption than rent and housing services since food is more responsive to transitory income. On the other hand, other omitted factors could be even more volatile than food, and since the relative volatility of different elements of consumption is not known, a more complete measure of consumption can more faithfully replicate the behavior of total consumption.

Finally, cross-section regressions may provide little information about time-series properties of consumption. For example, if each family allocated *different* fixed proportions of total consumption to food, rent, housing, cars, etcetera, then proportional changes in household food consumption would proxy perfectly for proportional changes in total consumption, despite the poor explanatory power of food consumption in cross-section regressions. However, the predicted consumption measure presented in this paper would do no worse than the food measure, since each category (food, rent, utilities) would jump by the same proportion; given homotheticity, so too would total consumption. It remains to be seen whether using this new measure of consumption will lead to results which differ from previous studied which used only food consumption.

References

Altonji, Joseph G., 1986, Intertemporal substitution in labor supply: Evidence from micro data, Journal of Political Economy 94, S176–S215.

Altonji, Joseph G. and Aloysius Siow, 1985, Testing the response of consumption to income changes with (noisy) panel data, Mimeo.

Flavin, Marjorie, 1986, work in progress.

Hall, Robert E. and Frederic S. Mishkin, 1982, The sensitivity of consumption to transitory income: Estimates of panel data on households, Econometrica 50, 461–481.

Hamermesh, Daniel, 1984, Consumption during retirement: The missing link in the life cycle, Review of Economics and Statistics 66, 1–7.

Hotz, V. Joseph, Finn E. Kydland and Guilherme L. Sedlacek, 1985, Intertemporal preferences and labor supply, Working paper no. 85-18 (NORC, Chicago, IL).

Zeldes, Stephen, 1985, Consumption and liquidity constraints: An empirical investigation, Mimeo.