The Superior Measure of PSID Consumption: an Update

Sheng Guo *

Department of Economics

Florida International University

May 10, 2010

Abstract

This paper provides an update of Skinner's (1987) predicted consumption measure for households in the Panel Study of Income Dynamics from 1980 to 2003. More than 80 percent of the variance in total non-durable consumption is explained by three expenditure components.

JEL Code: D10 D12 C23 E21

Key Words: Consumption; Expenditure; PSID; CEX; Household behavior

*Correspondence address: 11200 SW 8th Street, DM 320B, Miami, FL 33199. Email: Sheng.Guo@fiu.edu; Phone: 1-305-348-2735; Fax: 1-305-348-1524. I am grateful to Jonathan Skinner and an anonymous referee for helpful suggestions and comments. All remaining errors are my own.

1 Introduction

More than twenty years have passed since Skinner (1987) popularized the empirical procedure of predicting total consumption from a subset of individual consumption components in the Panel Study of Income Dynamics (PSID) data. PSID is a longitudinal study of a representative sample of U.S. households starting from 1968 with 4,800 families. Individuals from these families were interviewed annually (until 1996) and then biennially (after 1996). Children from these families, when they grow up, are also included in the study. PSID collects a variety of socio-economic information from surveyed families, including income, employment, wealth, health, and family composition. However, one of the notable limitations of PSID is that its data on consumption is limited, often confined to food expenditure, although in later years it has been expanded to utilities and transportation items. In light of the high quality data on income and other variables available in PSID, this is regrettable to researchers of household behavior.

Skinner's (1987) procedure exploits the relative comprehensiveness of expenditure items in Consumer Expenditure Surveys (CEX). Posing it as an omitted-variable problem, Skinner proposes predicting the total consumption from a subset of consumption components that are available both in CEX and in PSID. Using CEX 1972-1973 data and validating the estimates on the CEX 1983 wave, Skinner demonstrates that the predictive power of a few components, such as food, rent or value of house, and utilities, is quite high, with R^2 more than 0.68. Recent work adopting variations of Skinner's procedure to deal with PSID consumption measurement in order to address various research questions include Fisher and Johnson (2006), Charles, Danziger, Li, and Schoeni (2006), Filer and Fisher (2005), Waldkirch, Ng, and Cox (2004), Browning, Crossley, and Weber (2003), Meyer and Sullivan (2003), Dynan (2000), and Mulligan (1997). Benito and Mumtaz (2009) apply this imputation procedure to the British Household Panel Survey (BHPS) and the Family Expenditure Survey (FES), the UK counterparts of PSID and CEX.

How well does this procedure fare after 1983? The aforemenetioned work estimate for specific years based upon their respective research objectives, but none examines the performance of the procedure itself over time systematically. Thanks to the availability of a long series of CEX samples, we are now in a better position to evaluate the performance of Skinner's consumption measure over twenty years. The conclusion we achieve from this exercise is heartening: for the period 1980-2003, more than 80 percent of the variance in total non-durable consumption is sufficiently explained by three expenditure components (food, utility, and transportation). Further, the estimated coefficients as well as predicative power are highly stable for this period.

2 Data and Estimation

The CEX sample used for our analysis is compiled by Anguiar and Hurst (2009), which in turn is constructed from various files of the CEX Extracts compiled by Ed Harris and John Sabelhaus of the Congressional Budget Office that are publicly available from the National Bureau of Economic Research website (http://www.nber.org/data/ces_cbo.html)[1]. The sample is restricted to households where the head is between the ages of 25 and 75, who report expenditures in all four quarters of the year of the survey[2], and report non-zero annual expenditures on six key consumption components (including food, utility, and transportation). Food expenditures include food eaten at home, eaten out and at work; utilities expenditures include electricity, gas, water, home fuel, and telephone expenditures; transportation expenditures include car services, gasoline, tolls paid, auto insurance, public transportation, and other miscellaneous items. These sub-components of the three categories match remarkably well with those in PSID, except that in some years some sub-components of a category or the whole category are missing in PSID. All numbers are discounted by

[1] As of April, 2010, the available NBER CEX Extracts has covered the period 1980-2003.

[2] Except in 2003, for which only data for the first two quarters are available.

respective CPIs to obtain real measures in constant dollars of year 2000.

We include food, utility, and transportation progressively as independent variables in our specifications. PSID survey questions regarding these expenditure components show up unevenly for the period 1980-2003 (Table 1). For instance, food expenditure is unavailable in the 1988 and 1989 surveys, while questions related to transportation expenditure had not been included in the survey before 1999. The dependent variable is the expenditure on non-durable goods and services. In contrast to Skinner (1987), we do not add the rent or imputed rent value of housing into the regressions, on the grounds that the consumption of housing may be of substantially durable nature and warrants a separate treatment. We chose to examine the logarithm transformation of consumption measure in our regressions. This is because the log transformation of consumption is routinely employed in the literature. In addition, had we sought to transform the predicted measure into logarithms after the linear prediction, the original linear measurement error associated with this prediction exercise would be converted into non-linear errors that would be difficult to eliminate.

We estimate the coefficients for each year of 1980-2003 of the CEX sample. This assumes the prediction coefficients can vary over time, and time effects will not be completely captured by the year dummies in an otherwise pooled-sample regression. Table 2 and Table 3 present the estimated results: with food alone as the predicting variable, the R^2 is already above 0.65 in all years; adding utility and transportation yields R^2 more than 0.80. The estimated coefficients and R^2 are fairly stable over these years: the standard deviation for the food coefficient in the one-variable prediction conducted for all the years is 0.020 and when all three consumption components are included as predicting variables, the standard deviation for R^2 is 0.007.

Comparing the estimates across years, we can formally reject, at significance level 0.001, the hypothesis that fixed effects of individual years are jointly zero; however, for the sub-period 1990-1999, year fixed effects are not significant at significance level 0.1 (estimated coefficients are displayed at the bottom of Table 3). Despite this, any attempt to deploy the estimates for the period 1990-1999 for predicting the consumption for future years can be decisively rejected, which is consistent with the conclusion of Skinner's (1987) original quest of whether 1972-1973 coefficients can be used in predicting 1983 consumption. Also consistent, however, is the fact that this formal statistical rejection masks the negligible difference in magnitude: if we regress the predicted consumption for individual years 2000-2003 based on 1990-1999 coefficients, on the predicted consumption based on own coefficients of these years, the resulted R^2 is at least 0.9955.

3 Discussion

Explaining the cross-section variance in total consumption is one objective. More important is to obtain a predictive measure of total consumption as one of the inputs in subsequent analyses. For that purpose what are possible consequences of using the Skinner's predicted consumption instead of the true measure? For simplicity, suppose the true log consumption (c_i) includes two components $x_{1,i}$ and $x_{2,i}$ ($c_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i$) while $x_{2,i}$ is the omitted consumption component, and a constant term is included. ε_i is assumed to be independent of x's. The standard solution to this omitted-variable problem yields plim $\hat{\beta}_1 = \beta_1 + \beta_2 \frac{Cov(x_2, x_1)}{Var(x_1)}$ (Wooldridge 2002, Chapter 4). Asymptotically the difference between predicted and true consumption is $\hat{c}_i - c_i = \beta_2 \left(\frac{Cov(x_2, x_1)}{Var(x_1)} x_{1,i} - x_{2,i} \right) - \varepsilon_i$. It can be seen from this expression that if second-stage regressions involve no x_1 or x_2 as their independent variables, and no unobservables are correlated with x_1 or x_2, then using $\hat{c}_i$ instead of c_i as the dependent variable would generate no bias in second-stage estimates.

One favorable application of the prediction measure is to obtain a time series of cross-section averages of total consumption. With t subscript indicating the time dimension, the large sample cross-section average of predicted and true logarithm of total consumption are equal: $\frac{1}{N} \sum_{i=1}^{N} \hat{c}_{i,t} = \frac{1}{N} \sum_{i=1}^{N} c_{i,t}$ (with a constant term, $\frac{1}{N} \sum_{i=1}^{N} x_{1,it} = \frac{1}{N} \sum_{i=1}^{N} x_{2,it} = 0$); so are the averages of linear transformations of $\hat{c}_{i,t}$ and $c_{i,t}$. Consequently, the time series covariance structure of total consumption with other variables is still preserved, which implies in principle, studies exploiting the time variation of cross-section averages of total consumption from CEX (e.g., Vissing-Jørgensen (2002)) can be extended to PSID, with the benefit of additional abundant variables in PSID.

The capability of tracking the same households and/or individuals over years in PSID is of substantial advantage compared to CEX or other data sets. In panel data estimation schemes where the first-order difference in consumption is employed to eliminate idiosyncratic effects, plugging in $\hat{c}_{i,t}$ instead of $c_{i,t}$ will suffice. This may seem somewhat surprising, but the key observation leading to this conclusion is that by construction, the first-order difference of $\hat{c}_{it}$ contains no individual fixed effect in the first place. Augment the expression of the true log consumption for individual i at year t with an unobserved individual fixed effect η_i: $c_{it} = \beta_{1,t}x_{1,it} + \beta_{2,t}x_{2,it} + \eta_i + \varepsilon_{it}$. Cross-section regressions of consumption for each year t only on $x_{1,it}$ gives us $\hat{\beta}_{1,t} = \beta_{1,t} + \beta_{2,t}\left(\frac{Cov(x_{2,t},x_{1,t})}{Var(x_{1,t})} + \frac{Cov(\eta_i,x_{1,t})}{Var(x_{1,t})}\right) \equiv \beta_{1,t} + \beta_{2,t}(\hat{\gamma}_{12,t} + \hat{\gamma}_{1\eta,t})$. $\hat{\gamma}_{1\eta,t}$ is determined by the population correlation between $x_{1,it}$ and η_i but by itself is a constant. Therefore

$$\triangle\hat{c}_{it} - \triangle c_{it} = \beta_{2,t}(\hat{\gamma}_{12,t} + \hat{\gamma}_{1\eta,t})x_{1,it} - \beta_{2,t-1}(\hat{\gamma}_{12,t-1} + \hat{\gamma}_{1\eta,t-1})x_{1,i(t-1)} - (\beta_{2,t}x_{2,it} - \beta_{2,t-1}x_{2,i(t-1)} + \triangle\varepsilon_{it}) \tag{3.1}$$

Once again, as long as the right-hand side variables of second-stage regressions do not include $x_{1,it}$, $x_{1,it-1}$, $x_{2,it}$, or $x_{2,i(t-1)}$, or any unobservables correlated with them, there should be no biases in the estimates for the second-stage.

However, problems will arise when the subsequent regression depends on the second or higher moments of the consumption variable. When $R^2 = 0.8$, $Var(\hat{c}_i) = 0.8Var(c_i)$; under this circumstance, when $\hat{c}_i$ is in place of c_i appearing as one of the independent variables on the right-hand side of the equation, the coefficient as well as its standard error will be biased upward, all else being equal.

References

Anguiar, M., and E. Hurst (2009): "Deconstructing Lifecycle Expenditure," Manuscript, The University of Chicago Booth School of Business.

Benito, A., and H. Mumtaz (2009): "Excess Sensitivity, Liquidity Constraints, and the Collateral Role of Housing," *Macroeconomic Dynamics*, 13(03), 305–326.

Browning, M., T. F. Crossley, and G. Weber (2003): "Asking Consumption Questions in General Purpose Surveys," *Economic Journal*, 113(491), F689–F705.

Charles, K. K., S. Danziger, G. Li, and R. F. Schoeni (2006): "Studying Consumption with the Panel Study of Income Dynamics: Comparisons with the Consumer Expenditure Survey and an Application to the Intergenerational Transmission of Well-being," Working Paper, Finance and Economics Discussion Series, Federal Reserve Board, Washington, D.C.

Dynan, K. E. (2000): "Habit Formation in Consumer Preferences: Evidence from Panel Data," *The American Economic Review*, 90(3), 391–406.

Filer, L. H. I., and J. D. Fisher (2005): "The Consumption Effects Associated with Filing for Personal Bankruptcy," *Southern Economic Journal*, 71(4), 837–854.

Fisher, J. D., and D. S. Johnson (2006): "Consumption Mobility in the United States: Evidence from Two Panel Data Sets," *The B.E. Journal of Economic Analysis and Policy*, 6(1), 1–36.

Meyer, B. D., and J. X. Sullivan (2003): "Measuring the Well-Being of the Poor Using Income and Consumption," *Journal of Human Resources*, 38, 1180–1220.

Mulligan, C. B. (1997): *Parental Priorities and Economic Inequality*. University of Chicago Press, Chicago.

Skinner, J. (1987): "A Superior Measure of Consumption from the Panel Study of Income Dynamics," *Economics Letters*, 23(2), 213 – 216.

Vissing-Jørgensen, A. (2002): "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, 110(4), 825–853.

Waldkirch, A., S. Ng, and D. Cox (2004): "Intergenerational Linkages in Consumption Behavior," *Journal of Human Resources*, 39(2), 355–381.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, Mass.

Tables

PSID Domain	Data available for the year (period of interest: 1980-2003)
Food	
eaten at home	1980-2003, except 1988-1989
delivered	1994-2003
away from home	1980-2003, except 1988-1989
Utilities	
general (not specified)	1980-1987
specified sub-components[(a)]	1999-2003
Transportation[(b)]	1999-2003

Notes: (a) electricity and water, gas or heating fuel, water and sewer, and other utility expenses; (b) car insurance, repair and maintenance, gasoline, taxi / cabs, parking and car pool, bus and train fares, other transportation expenses.

Table 1: Consumption components availability in PSID 1980-2003

Year	Obs.	Food		Utility		Transportation		(Constant)		R squared
		Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	
1980	550	0.8678	(0.0297)					2.2127	(0.2539)	0.6870
		0.7389	(0.0292)	0.1574	(0.0150)			2.1975	(0.2177)	0.7376
		0.5844	(0.0236)	0.1064	(0.0143)	0.2017	(0.0149)	2.3180	(0.1797)	0.8476
1981	1644	0.7969	(0.0159)					2.9576	(0.1392)	0.6781
		0.6945	(0.0168)	0.1718	(0.0183)			2.5396	(0.1417)	0.7270
		0.5417	(0.0132)	0.1184	(0.0130)	0.2360	(0.0114)	2.3701	(0.1004)	0.8462
1982	2164	0.8159	(0.0129)					2.8096	(0.1123)	0.6933
		0.7431	(0.0130)	0.1484	(0.0141)			2.3171	(0.1176)	0.7327
		0.5683	(0.0129)	0.1081	(0.0106)	0.2234	(0.0122)	2.3260	(0.0880)	0.8378
1983	2302	0.8046	(0.0126)					2.9117	(0.1098)	0.6843
		0.7135	(0.0130)	0.1749	(0.0156)			2.3657	(0.1231)	0.7324
		0.5543	(0.0124)	0.1289	(0.0127)	0.2228	(0.0112)	2.2928	(0.0959)	0.8403
1984	2628	0.8351	(0.0123)					2.6568	(0.1070)	0.6965
		0.7368	(0.0130)	0.1847	(0.0142)			2.1015	(0.1108)	0.7388
		0.5650	(0.0139)	0.1428	(0.0119)	0.2272	(0.0134)	2.0646	(0.0875)	0.8394
1985	1301	0.8166	(0.0159)					2.7993	(0.1395)	0.7187
		0.7376	(0.0171)	0.1394	(0.0187)			2.4246	(0.1405)	0.7518
		0.5771	(0.0167)	0.1028	(0.0159)	0.2212	(0.0154)	2.3008	(0.1153)	0.8390
1986	2623	0.8017	(0.0110)					2.9598	(0.0951)	0.6942
		0.7084	(0.0115)	0.1835	(0.0116)			2.3657	(0.0949)	0.7392
		0.5420	(0.0110)	0.1366	(0.0099)	0.2406	(0.0097)	2.1921	(0.0799)	0.8436
1987	2546	0.7991	(0.0180)					2.9584	(0.1566)	0.6637
		0.6924	(0.0173)	0.2084	(0.0172)			2.2915	(0.1443)	0.7246
		0.5302	(0.0143)	0.1459	(0.0129)	0.2387	(0.0101)	2.2301	(0.1071)	0.8407
1988	2578	0.8265	(0.0124)					2.7110	(0.1082)	0.6986
		0.7244	(0.0127)	0.1938	(0.0147)			2.1129	(0.1102)	0.7491
		0.5707	(0.0129)	0.1358	(0.0113)	0.2189	(0.0104)	2.1096	(0.0836)	0.8482
1989	2601	0.8325	(0.0118)					2.6535	(0.1031)	0.7058
		0.7196	(0.0125)	0.2238	(0.0119)			1.9106	(0.1019)	0.7510
		0.5643	(0.0115)	0.1655	(0.0100)	0.2340	(0.0101)	1.8036	(0.0810)	0.8557
1990	2598	0.8305	(0.0132)					2.6611	(0.1141)	0.6937
		0.7414	(0.0140)	0.1764	(0.0164)			2.0792	(0.1208)	0.7379
		0.5625	(0.0121)	0.1155	(0.0135)	0.2531	(0.0093)	2.0467	(0.0966)	0.8497
1991	2607	0.8181	(0.0137)					2.7841	(0.1183)	0.6600
		0.6950	(0.0165)	0.2376	(0.0278)			2.0132	(0.1610)	0.7188
		0.5358	(0.0127)	0.1602	(0.0220)	0.2437	(0.0101)	2.0112	(0.1232)	0.8383
1992	2608	0.8481	(0.0127)					2.5033	(0.1099)	0.6918
		0.7335	(0.0140)	0.2034	(0.0167)			1.9276	(0.1190)	0.7405
		0.5745	(0.0151)	0.1395	(0.0168)	0.2323	(0.0113)	1.9130	(0.1074)	0.8436
1993	2683	0.8441	(0.0128)					2.5649	(0.1101)	0.6821
		0.7221	(0.0140)	0.2363	(0.0178)			1.7886	(0.1151)	0.7454
		0.5603	(0.0119)	0.1517	(0.0137)	0.2465	(0.0099)	1.8415	(0.0901)	0.8559

Table 2: Non-durable goods and services consumption prediction regressions from CEX samples of various years

Year	Obs.	Food		Utility		Transportation		(Constant)		R squared
		Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	
1993	2683	0.8441	(0.0128)					2.5649	(0.1101)	0.6821
		0.7221	(0.0140)	0.2363	(0.0178)			1.7886	(0.1151)	0.7454
		0.5603	(0.0119)	0.1517	(0.0137)	0.2465	(0.0099)	1.8415	(0.0901)	0.8559
1994	2506	0.8613	(0.0129)					2.4170	(0.1115)	0.6855
		0.7234	(0.0163)	0.2617	(0.0245)			1.5759	(0.1308)	0.7452
		0.5441	(0.0139)	0.1840	(0.0198)	0.2445	(0.0106)	1.7444	(0.1008)	0.8550
1995	1129	0.8499	(0.0187)					2.5122	(0.1610)	0.6917
		0.6814	(0.0191)	0.3214	(0.0202)			1.4646	(0.1460)	0.7562
		0.5371	(0.0176)	0.2192	(0.0186)	0.2346	(0.0151)	1.6077	(0.1198)	0.8578
1996	2189	0.8351	(0.0147)					2.6643	(0.1266)	0.6857
		0.7093	(0.0158)	0.2376	(0.0246)			1.8898	(0.1528)	0.7470
		0.5358	(0.0133)	0.1638	(0.0188)	0.2430	(0.0110)	1.9833	(0.1137)	0.8543
1997	2467	0.8363	(0.0136)					2.6484	(0.1169)	0.6636
		0.6693	(0.0126)	0.3392	(0.0140)			1.4338	(0.1101)	0.7521
		0.5112	(0.0115)	0.2290	(0.0110)	0.2421	(0.0105)	1.6885	(0.0887)	0.8607
1998	2300	0.8591	(0.0136)					2.4584	(0.1175)	0.6803
		0.7163	(0.0156)	0.2748	(0.0224)			1.5368	(0.1278)	0.7535
		0.5467	(0.0146)	0.1916	(0.0189)	0.2382	(0.0113)	1.7080	(0.1031)	0.8551
1999	3023	0.8596	(0.0125)					2.4475	(0.1073)	0.6774
		0.6848	(0.0134)	0.3500	(0.0183)			1.2057	(0.1136)	0.7618
		0.5209	(0.0123)	0.2465	(0.0177)	0.2421	(0.0115)	1.4693	(0.0952)	0.8568
2000	3146	0.8328	(0.0127)					2.6638	(0.1093)	0.6563
		0.6710	(0.0131)	0.3322	(0.0192)			1.4421	(0.1254)	0.7498
		0.5185	(0.0130)	0.2316	(0.0190)	0.2286	(0.0099)	1.6964	(0.1113)	0.8498
2001	3320	0.8302	(0.0115)					2.7183	(0.0988)	0.6560
		0.6643	(0.0115)	0.3535	(0.0158)			1.3483	(0.1070)	0.7471
		0.5127	(0.0112)	0.2414	(0.0138)	0.2393	(0.0093)	1.5965	(0.0862)	0.8476
2002	3641	0.8246	(0.0120)					2.7767	(0.1034)	0.6515
		0.6508	(0.0131)	0.3592	(0.0220)			1.4194	(0.1281)	0.7456
		0.5012	(0.0108)	0.2408	(0.0177)	0.2420	(0.0095)	1.6841	(0.1008)	0.8464
2003	1898	0.8457	(0.0168)					2.5865	(0.1433)	0.6764
		0.6657	(0.0165)	0.3377	(0.0232)			1.4470	(0.1642)	0.7607
		0.5078	(0.0140)	0.2456	(0.0170)	0.2414	(0.0120)	1.5796	(0.1264)	0.8475
1990–1999		0.8440	(0.0043)					2.5673	(0.0372)	0.6802
(pooled)		0.7137	(0.0050)	0.2514	(0.0073)			1.7390	(0.0435)	0.7426
		0.5458	(0.0044)	0.1707	(0.0061)	0.2437	(0.0035)	1.8373	(0.0349)	0.8509

Note: CEX sample for each year (see Aguiar and Hurst (2009) for a detailed description of the sample construction); the dependent variable is log of non-durable goods and services expenditure; regressors are log of food expenses, of utility expenses, and of transportation expenses respectively. Robust standard errors are reported in parentheses.

Table 3: Non-durable goods and services consumption prediction regressions from CEX samples of various years (continued)