Online Robustness Appendix to
"Are Household Surveys Like Tax Forms: Evidence from the Self Employed"

December 2010

Erik Hurst
University of Chicago

Geng Li
Board of Governors of the Federal Reserve System

Benjamin Pugsley
University of Chicago

Online Robustness Appendix

This document serves as the online robustness appendix to our paper "Are Household Surveys Like Tax Forms: Evidence from Income Underreporting of the Self-Employed". In this document we present details on five additional topics that are referenced within the main text. Specifically, we start by discussing our use of log-linear Engel curves for the main specification that relates income and expenditure within the paper. Second, we discuss the validity of our assumption that the slopes of the Engel curves and the relationship between the control variables in the Engel curves are similar between the self employed and wage and salary workers. Third, we discuss the procedure we used to test for systematic mismeasurement of expenditures by the self employed relative to wage and salary workers. Fourth, we describe the specifications where we tested for the extent to which our estimates of underreporting vary with either total household wealth or business wealth. Finally, we outline the procedure we used to test for the effects of income underreporting by the self employed on estimates of precautionary savings.

1. Log Linear Functional Form

In the main text we assume that, conditional on the covariates and time fixed effects, a log linear Engel curve is a reasonable approximation to the underlying structural relationship between income and expenditure. Using the Consumer Expenditure Survey (CE), we evaluate the validity of this assumption. To do so we pursue two strategies. First, we estimate a log linear Engel curve and inspect the residuals of an augmented component plus residual plot.[1] Second, we estimate

[1] A component plus residual plot plots the potentially nonlinear regressor x_1 in a multivariate regression against fitted values (omitting the effect of additional regressors). An augmented component plus residual plot first includes the regressor x_1 and its square and plots x_1 against the fitted values from that augmented estimation (again omitting the effect of additional regressors). Mallows (1986) shows that the augmented form is better able to detect nonlinear effects of x_1 since other regressors may covary with the nonlinear function of x_1.

alternative functional forms of the Engel curve and test for the significance of higher order terms. For both approaches we examine all combinations of income and expenditure measures and focus exclusively on the sample of wage and salary workers. Finally, where appropriate, we estimate a measure of underreporting under alternative functional form assumptions.

Inspecting the residuals shows little evidence of nonlinearity. For each income and expenditure measure we compute an augmented component plus residual plot in Stata.[2] Because income is measured with error, we instrument for the income measure with education dummies and an expanded instrument set that includes occupational dummies.[3] The augmented component plus residual plot graphs predicted income (from the first stage) against residuals from an estimation that includes the predicted income and its square along with the other regressors. Across both income measures and all expenditure measures, but particularly for nondurable and total expenditures, we see only weak evidence of any nonlinearity. Figure R1 shows the augmented component plus residual plots for each income and expenditure measure. Each plot also includes both a line and nonparametric Lowess curve for reference. Comparisons of the fitted line through the residuals and a Lowess nonparametric regression show only very minor deviations in the lower tail of the income distribution.

Next we explicitly estimate two higher order deviations from the baseline log linear specification and test the significance of the nonlinear terms. The first variant is a log quadratic Engel curve

$$\log c_{ikt} = \alpha + \beta_1 \log y_{ikt}^{p} + \beta_2 (\log y_{ikt}^{p})^2 + \varepsilon_{ikt} \,, \qquad \text{(R1)}$$

[2] The accompanying Stata do files implement all procedures described here.
[3] For just the wage and salary sample we can use variation in occupation as an additional instrument for income. Self employment is considered an occupation, which considerably limits this approach for the pooled sample.

where the log of permanent income, $\log y_{ikt}^{p}$, is assumed to have both first and second order effects on consumption. This form is similar to the quadratic Engel curves derived by Banks, Blundell, and Lewbel (1997) in the Quadratic Almost Ideal Demand System. The second variant uses the inverse of income in levels to capture nonlinearity as follows

$$\log c_{ikt} = \alpha + \beta_1 \log y_{ikt}^{p} + \beta_2 \frac{1}{(y_{ikt}^{p})} + \varepsilon_{ikt} \ . \qquad \text{(R2)}$$

This is similar to the rank 3 form that Leser (1963) finds to fit expenditure data well.[4] Again ,in both cases we observe an annual measure of income, y_{ikt} , rather than the permanent component y_{ikt}^{p} driving consumption. To account for the measurement error we instrument for our measure of income. Since nonlinear functions of permanent income measure appear in both specifications, we instrument both log income and its transformation with the expanded instrument set that includes occupation dummies.[5]

Using just the wage and salary sample, we estimate both specifications for both measures of annual income and all three measures of expenditures in the CE. Table R1 summarizes the results of these estimations. Each column represents either an OLS or 2SLS estimation of the modified Engel curve for both measures of household income. We refer to the first Banks, Blundell, and Lewbel (1997) based specification as “log quadratic” and the Leser (1963) based form as “inverse”. Although we present both the OLS and 2SLS for reference, we focus on the 2SLS results that account for the transient variation in income.

---

[4] Both Banks, Blundell, and Lewbell (1997) and Leser (1963) are interested in a demand system where the left hand side is the budget share of a particular expenditure category, which differs from our specification. We consider these higher rank systems strictly to determine whether or not nonlinearity is an important confounding factor in our estimation of underreporting.

[5] This additional instrument has a negligible effect on the estimated income elasticity in the benchmark log linear specification.

The estimations only weakly support the presence of nonlinearity. The data do not support the inverse income specification for any combination of income and expenditure. The log quadratic specification does appear to fit the data better, but with wide confidence intervals on the first order effects. These imply large confidence sets for the Engel curve, which easily contain the linear case. Considering these alternative specifications that explicitly integrate nonlinearity, the data fail to provide strong evidence against our assumed log linear specification in the main text.

Lastly, we show that even if we were to accept the point estimates of the log quadratic specification at face value, the implied bounds on our measure of underreporting are still relatively tight. We construct a measure of underreporting consistent with a log quadratic Engel curve by plugging in the determinants of annual income from equations (6) and (7) in the main text into equation (R1) to form

$$\log c_{ikt} = \alpha + \gamma_1 D_{it} + \gamma_2 D_{it} \times \log y_{ikt} + \beta_1 \log y_{ikt} + \beta_2 (\log y_{ikt})^2 + \xi_{ikt} \tag{R3}$$

where now $\gamma_1 = \beta_2 (\log \kappa)^2 - \beta_1 \log \kappa$ and $\gamma_2 = -2\beta_2 \log \kappa$. The parameter $\log \kappa$ is overidentified with knowledge of γ_1, γ_2, β_1, and β_2. It is both a root of the quadratic equation from the first equation or proportional to the coefficient on the interacted dummy.

We estimate equation (R3) using the CE, instrumenting for all functions of annual income using an expanded instrument set interacted with the self employment dummy.[6] Table R2 shows the implied estimate of underreporting, $1-\hat{\kappa}$, for various income and expenditure measures. Only the food expenditure based Engel curves are estimated with any precision.

[6] There is very little occupational variation in the self employment sample, since self employment is an occupation code, so we use industry instead. Industry is missing for almost half of the observations, so these 2SLS estimates are only suggestive. Also, the higher order terms technically make the self employment dummy endogenous since the cross product of transient income and the self employment dummy is introduced by the square of reported income and included in the error. We do not account for this in these estimations.

Because of the limitations of our data, these estimates are really only suggestive. However, they are compatible with our findings in the main text. Overall, we place the most confidence in our food expenditure based estimates, which are robust across surveys and alternative Engel curve specifications.

2. Comparing Wage and Salary and Self Employed Engel Curves

The model relies on the crucial assumption that the slope of the expenditure Engel curve is invariant to the class of the household between wage and salary and self employed. Moreover, as described in the main text, we also assume the coefficients on the vector of controls are equal across household type, leaving only the intercept to shift with the distinction between wage and salary and self employed. Through the income elasticity we convert this vertical difference in intercepts into a measure of underreporting. In this section, using the CE, we review additional evidence that supports this assumption. We re-estimate the model under increasingly relaxed constraints and confirm that evidence fails to convincingly reject our restrictions.

Using the CE, for all measures of income and expenditures, we re-estimate using 2SLS the Engel curve described by equation (8), reproduced here,

$$\ln c_{ikt} = \alpha + \beta \log y_{ikt} + \gamma D_{it} + \Psi ' X_{ikt} + \mu_t + \xi_{ikt} \quad ,$$

using the pooled sample while relaxing the restrictions of equal effects of controls, time effects, and slope. We relax these restrictions by including interaction terms on the appropriate regressors. Table R3 summarizes the re-estimation results allowing interactions on the controls, the time effects, the slope, and combinations of the three. For each estimation, we compute a Wald test of the joint restriction that the coefficient on all interaction terms is equal to zero. We show the p-values for each test statistic. The results are inconsistent across measures of income and expenditures. For food expenditures especially, we fail to find strong evidence against pooling the sample. For the other expenditure measures the difference in Engel curve slopes is statistically significant at even 1 percent levels. This is consistent with figures 2 and 3 that show where the estimated slopes of the separately estimated Engel curves differ to the naked eye, although the difference is still quantitatively small.

Even with potentially varying slopes we still view our underreporting estimates as robust across measures. To check the sensitivity of our main results, for each set of possible interactions, we compute the $1-\hat{\kappa}$ estimator. Holding the slopes fixed, the estimates in Table R3 are relatively close to our main results.

If the slopes do vary, it undermines our identification. Now the intercepts can vary for two reasons: for underreporting and for the same underlying determinants that also change the slope. If we allow the slope and intercept to depend on class, then

$$\log c_{ikt} = \alpha_W + (\alpha_S - \alpha_W) D_{it} + \beta_W \log y_{ikt}^p + (\beta_S - \beta_W) \log y_{ikt}^p \times D_{it} + \Theta' X_{ikt} + \varepsilon_{ikt}.$$

Without knowing the difference in intercepts from this second factor, $\alpha_S - \alpha_W$, we cannot separately identify κ. In table R3, we make the additional restriction that $\alpha_S = \alpha_W$ and estimate κ. If $\alpha_S > \alpha_W$, which is likely to be the case given an estimated κ close to 1, this estimate would decrease and vice versa.

Overall, after allowing coefficients to vary by class in varying combinations, we do not find strong evidence against our constant slope and pooling assumption in the paper. Even if the slopes we to vary, as they do somewhat for the broader measures of expenditures, our main results are consistent with self employed expenditures that exceed wage and salary expenditures for low levels of *accurately* reported income, i.e., when $\alpha_S > \alpha_W$.

3. No Systematic Expenditure Underreporting

We assume in the paper that while income may be underreported by the self employed, expenditures are reported accurately, at least relative to wage and salaried households. We address the possibility in the paper that the self employed may misclassify household expenditures as business expenditures or vice versa. It is also possible, that to be consistent across reports, the self-employed underreport both income *and* expenditures by uniformly misreporting their spending on all categories If expenditures were also underreported it would be impossible to identify the component of income underreporting separately, and our results in the paper would be biased towards less income underreporting. We test for systematic expenditure underreporting in some depth and find little evidence of it.

To do so, we estimate the following relationship based on Deaton and Muellbauer's (1982) Almost Ideal Demand System such that:

$$\omega_{jikt} = \alpha + \mu_t + \beta_j \log C^*_{ikt} + \Theta' X_{ikt} + \varepsilon_{ikt}, \qquad \text{(R4)}$$

where ω_{jikt} is the share of expenditures allocated to category j in period t by consumer i whose employment status can be $k=W$ or $k=S$. C^*_{ikt} represents the actual total expenditures of household i in period t and X_{ikt} is our maintained vector of controls. β_j, in this case, represents how the share of spending on a given category changes as total expenditures increase. β_j less than 0 reflect categories that are relative necessities while β_j greater than 0 reflect categories that are relative luxuries.

We modify (R4) to allow for systematic misreporting of total expenditure by the self-employed (by a constant fraction across all categories) such that: $C_{ikt} = \kappa_k{}^C C^*_{ikt} \exp(\nu_{ikt})$ where C

equals reported total expenditures, κ_W^C equals 1 for wage and salary workers ($k = W$), and κ_S^C different from 1 for the self-employed ($k = S$). Assuming the parameters of (9) are otherwise similar between the self-employed and wage and salary workers we can identify κ_S^C from:

$$\omega_{jikt} = \alpha + \mu_t + \beta_j \log C_{ikt}^* + \gamma D_S + \Theta' X_{ikt} + \varepsilon_{ikt} \qquad \text{(R5)}$$

where $\kappa_S^C = \exp(-\gamma / \beta_j)$. Underreporting all expenditures by a constant fraction κ_C of course leaves the share ω_{jikt} unchanged. The identification relies on the decreasing and increasing budget shares of various necessity and luxury good categories. For example, we expect high income households to allocate a smaller share of their budget to food at home, which has an income elasticity significantly less than 0. We treat the budget shares of the wage and salary household as a benchmark. If conditional on controls and total expenditures we observe a lower than expected budget share on food at home, we attribute this difference to underreporting total expenditures. This is similar to the idea underlying the procedure Aguiar and Bils (2010) uses to identify differential measurement error in expenditure reporting between the rich and poor using the CE data. Using the food at home example, when we estimate equation (R5) using the food at home category we fail to reject the claim that $\kappa_C = 1$ at a 5 percent level.

Table R4 shows the complete results of estimating (R5) across an exhaustive set of nondurable spending categories.[7] We estimate (R5) separately for each category and do not impose an adding up constraint across categories. We are primarily interested in inferring whether the budget shares for the self employed are systematically too large (small) for their reported level of *nondurable expenditures*. For each category we compute an estimate of κ^C.

[7] Our measure of nondurable spending is the same as in the paper and excludes housing, health, and education expenses.

Table R4 reports OLS estimates of the parameters; instrumenting for measurement error in nondurable expenditures with total income has very little change on the results. The estimates point to little evidence of differential expenditure underreporting aside from the issues already discussed in the nondurable transportation category.

There is some evidence however, Banks, et al. (1997) and more recently Lewbel and Pendakur (2009) for example, that budget share Engel curves for certain expenditure categories like entertainment or clothing are highly nonlinear. We modify (R5) to include higher order terms that, when estimated, also fails to uncover a systematic difference in expenditure reporting for these nonlinear categories. The lower part of Table R4 shows the parameter and underreporting estimates and for a log quadratic specification of (R5). In almost all cases it is difficult to rule out $\kappa^C = 1$.

Using our procedure, and considering the results across a large number of expenditure categories, we find very little evidence that the self-employed systematically underreport their expenditures relative to wage and salary workers. The combined results of our two robustness exercises make us confident that any misreporting of expenditures is not biasing our results in any meaningful way.

4. Does Income Underreporting Change with Business or Total Wealth?

It is possible that the self-employed are business owners save in the form of retained earnings in business. In such a scenario the self-employed withdraw from their business only the income enough to finance their expenditure, inducing a higher ratio between reported expenditure and reported income. We argue that taking this factor into account does not materially change our results. Specifically, we examine the underreporting behavior of the self-employed with zero reported business wealth relative to the underreporting behavior of the self-employed with non-zero business wealth for if the self-employed are reinvesting a substantial amount of their income into the business, it should show up in the value of their business. During our PSID sample years, business value data were collected in 1984 and 1994.[8] Nearly one third of the self-employed in our PSID sample have zero business wealth. We merge our PSID sample of 1984 and 1994 with the business value data of the corresponding year and estimate the following model:

$$\ln c_{ikt} = \alpha + \beta \log y_{ikt} + \gamma D_{it} + \lambda D_{it}^{0} + \Psi' X_{ikt} + \mu_t + \xi_{ikt} ,$$

where D^0 is the self-employed dummy interacting with a dummy of having zero business wealth. We find that the estimated λ coefficient is a very small *positive* number (0.01) relative to the estimated γ coefficient (0.11) and has a large standard error (0.04). Adding to the sample the business value 1989 data merged with our PSID sample of 1990 yields essentially very similar results.

We then examine whether the extent to which the self-employed underreport their income varies with their total wealth. We estimate the following model

$$\ln c_{ikt} = \alpha + \beta \log y_{ikt} + \gamma_1 D_{it}^{1} + \gamma_2 D_{it}^{2} + \gamma_3 D_{it}^{3} + \gamma_4 D_{it}^{4} + \Psi' X_{ikt} + \mu_t + \xi_{ikt} ,$$

where D^1, D^2, D^3, and D^4 are the self-employed dummy interacting with the low, second, third and high wealth quartile dummy, respectively. The F-test cannot reject the hypothesis that all γ coefficients are equal (F value = 0.51, p value = 0.68) and the results are robust to whether the PSID 1990 data are included after being merged with the 1989 wealth data.

[8] The business value data were collected in 1989 but food expenditure data were not.

5. Estimating the Effect of Income Mismeasurement By Self Employed on Precautionary Savings Estimates

To estimate the effects of income underreporting of the self employed on estimates of precautionary savings, we draw on the specification from Hurst et al. (2010). The goal of the Hurst et al. paper was to show how the estimates of precautionary savings fall to close to zero when the self employed are excluded from the analysis. The procedure used in the Hurst et al. paper was nearly identical to the procedure used by Carroll and Samwick (1997, 1998) to provide micro data estimates of the importance of precautionary savings for younger households.

The empirical strategy of estimating the size of precautionary balances using micro data is based on the following specification:

$$\ln(W_{it}) = \alpha_0 + \alpha_1 \sigma_{it}^{permy} + \alpha_2 \sigma_{it}^{transy} + \alpha_3 \ln(y_{it}) + Z_{it}\beta + u_{it} \qquad \text{(R7)}$$

where $\ln(W_{it})$ is the log of a measure of household i's wealth in period t, $\ln(y_{it})$ is the log of i's permanent income in t, and σ_{it}^{permy} and σ_{it}^{transy} are, respectively, measures of the variance of permanent shocks and transitory shocks to i's income. The Z vector includes additional controls designed to capture potential household differences in preferences and the hump-shaped profile of wealth over the life cycle.

According to the precautionary saving model, wealth is a function not only of permanent income, but also of uninsurable income risk faced by the household. Almost all empirical studies designed to estimate the size of precautionary balances using micro data proxy uninsurable risk with either the variance of income, the variance of consumption, or they exploit actual job loss or expectations of future job loss. For our paper, we follow Carroll and Samwick (1997, 1998) by using panel data to distinguish between the variance of permanent and transitory shocks to income.

To estimate (R7), we use data from the PSID. We examine accumulated household wealth in either 1984 or 1994. This broadens the analysis performed in Carroll and Samwick (1997, 1998), which only analyzed household wealth accumulation within the PSID using 1984 wealth data. The measure of wealth used is total net worth, which is defined as the sum of checking and savings accounts, bonds, stocks and mutual funds (including IRAs), home equity, other real estate, business equity, cars and other vehicles, and other assets, minus the value of all debts. Since we use logs, we exclude households who have negative or zero net worth in our sample, which amount to a little more than five percent of our sample.

Following equation (R7), we regress the log of household wealth in year *t* (either 1984 or 1994) on both permanent income and measures of the variance of income. We construct permanent income for each household by taking the seven-year average of non-capital income around the period for which we are measuring their wealth. Specifically, when explaining 1984 (1994) wealth holdings, we define permanent income as the average of non-capital income between the years of 1981 and 1987 (1991 and 1997). We use panel data from the PSID to compute the variances of permanent and transitory shocks to income. We follow the same procedure put forth by Carroll and Samwick (1997, 1998).[9]

Since both permanent income and the variances of permanent and transitory income are measured with error, we instrument for these variables using a large instrument set. As suggested by Carroll and Samwick (1997, 1998), we use occupation dummies and these dummies interacted with age and age squared, as well as industry dummies. In addition, we use the unemployment rate in the county of residence during the prior year, the variance in the county

[9] See the data appendix to Hurst et al. (2010) for a detailed summary of how the income variances were computed.

unemployment rate over the sample period, and a dummy for whether the head belongs to a union.

When estimating (R7), we also include additional controls (*Z*) to capture additional reasons why household wealth may differ across households. The *Z* vector includes the following demographics: age, age squared, race, gender, marital status, and educational attainment. In addition, we exploit the panel dimension of the PSID to control for past income and wealth shocks experienced by households. Specifically, we include year dummies, along with two dummies for whether the household head was unemployed during the year when the wealth data were collected and whether they were unemployed any time during the prior four years (1980–1983 or 1990–1993). Households that are more likely to face high income risk are also more likely to have been hit by past negative income shocks, and this may weaken the estimated relationship between wealth and risk. We also include dummies for past positive shocks, such as having received inheritances or other lump-sum payments. These were the same included when Hurst et al. (2010) estimated their version of (R7).

Lastly, similar to Carroll and Samwick, we restrict our sample to households whose head is between the ages of 26 and 50 in the year in which the wealth is measured. A detailed description of other restrictions we used in constructing our final sample is reported in the data appendix to Hurst et al. (2010). Our final sample includes 2,144 households.

The base results in the paper are identical to the ones reported in Hurst et al (2010). To assess the effects of the underreporting of income by the self employed on the precautionary savings estimates, we inflated the income measures of the self employed by 30 percent. Otherwise, the specification was identical to the base specification.

References

Aguiar, Mark and Mark Bils (2010). "Has Consumption Inequality Mirrored Income Inequality", *University of Rochester Working Paper.*

Banks, James, Richard Blundell, and Arthur Lewbel (1997). "Quadratic Engel Curves and Consumer Demand," *The Review of Economics and Statistics*, vol. 79(4), pp. 527-39.

Leser, Conrad Emanuel Victor (1963), "Forms of Engel Curve Functions," *Econometrica*, vol. 31, pp. 694-703.

Lewbel, Arthur and Krishna Pendakur (2009). "Tricks with Hicks: The EASI Demand System," *American Economic Review,* vol. 99(3), pp. 827-63.

Mallows, Colin L. (1986), "Augmented Partial Residuals," *Technometrics*, vol 28(4), pp 313-319.

Table R1: Estimated Nonlinear Engel Curves

	I. Expenditures = Food				II. Expenditures = Nondurable				III. Expenditures = Total			
	Log Quadratic		Inverse		A. Log Quadratic		B. Inverse		A. Log Quadratic		B. Inverse	
Income Variable	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
					Labor + Business Income							
$\log y$	-0.68	-1.87	0.270	0.42	-0.88	-0.85	0.38	0.61	-1.08	-1.10	0.44	0.76
	(0.12)	(0.57)	(0.01)	(0.01)	(0.15)	(0.48)	(0.01)	(0.01)	(0.15)	(0.55)	(0.01)	(0.02)
$(\log y)^2$	0.05	0.11			0.06	0.07			0.07	0.09		
	(0.01)	(0.03)			(0.01)	(0.02)			(0.01)	(0.03)		
y^{-1}			3.23	7.59			4.41	25.81			5.87	52.88
			(0.54)	(11.98)			(0.90)	(20.98)			(0.93)	(41.74)
					Total Household Income							
$\log y$	-0.82	-1.66	0.28	0.42	-1.04	-0.44	0.40	0.62	-1.26	-0.59	0.46	0.78
	(0.18)	(0.58)	(0.01)	(0.01)	(0.23)	(0.50)	(0.01)	(0.01)	(0.22)	(0.57)	(0.01)	(0.01)
$(\log y)^2$	0.05	0.10			0.07	0.05			0.08	0.06		
	(0.01)	(0.03)			(0.01)	(0.02)			(0.01)	(0.03)		
y^{-1}			2.92	6.84			3.84	18.18			5.36	46.54
			(0.32)	(12.41)			(0.44)	(16.71)			(0.54)	(39.82)

Notes: Robust standard errors in parentheses.

Table R2: Estimated Underreporting Under Log Quadratic Engel Curve

	I. Food Expenditures	II. Nondurable Expenditures	III. Total Expenditures
	Labor + Business Income		
$1-\hat{\kappa}$	0.357	-0.332	-1.158
	(0.138)	(0.444)	(0.742)
	Total Household Income		
$1-\hat{\kappa}$	0.252	-0.075	-0.611
	(0.098)	(0.172)	(0.340)

Notes: Robust standard errors in parentheses. 2SLS estimates of the parameters of the log quadratic Engel curve.

Table R3: Estimates of $1-\kappa$ under Additional Interactions

	I. Food Expenditures				II. Nondurable Expenditures				III. Total Expenditures			
Statistic	A	B	C	D	A	B	C	D	A	B	C	D
					Labor + Business Income							
p-value of restriction	0.05	0.04	0.04	0.46	0.15	< 0.01	< 0.01	0.01	0.69	0.12	< 0.01	< 0.01
$1-\hat{\kappa}$	0.48	0.46	0.79	0.70	0.32	0.37	0.93	0.92	0.26	0.33	0.99	0.99
	0.05	0.09	0.23	0.34	0.05	0.07	0.06	0.08	0.05	0.06	0.02	0.01
					Total Family Income							
p-value of restriction	0.15	0.08	0.08	0.58	0.2	< 0.01	< 0.01	0.01	0.57	0.15	< 0.01	< 0.01
$1-\hat{\kappa}$	0.46	0.43	0.73	0.62	0.30	0.33	0.91	0.90	0.23	0.30	0.98	0.98
	0.06	0.09	0.28	0.42	0.05	0.08	0.08	0.10	0.05	0.06	0.02	0.02
Interactions												
Slope	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Time Effects	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No
Controls	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No

Notes: Standard errors are in parentheses. Standard errors of the underreporting measure are asymptotic approximations using a robust estimate of the variance covariance matrix.. p-value of restriction represents the p-value of a test of the joint restriction that the coefficient on the interacted terms are all equal to zero.

Table R4: Estimated Log-Linear and Log-Quadratic Demand System and Relative Expenditure Underreporting

	Budget Share									
Estimated Parameter	Food Total	Food In	Food Away	Transport-ation	Entertain-ment	Clothing & Personal Care	Utilities	Alcohol & Tobacco	Domestic Services	Other Non-durable
Log Linear Demand System: Wage and Salary Households										
$\hat{\beta}$	-0.070	-0.107	0.037	0.015	0.026	0.026	-0.053	0.006	0.012	0.039
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log Linear Demand System: Relative Expenditure Underreporting										
$1-\hat{\kappa}_C$ (OLS)	-0.124	-0.007	0.182	-9.366	0.076	0.065	-0.178	-0.666	0.391	0.083
	(0.033)	(0.018)	(0.035)	(3.720)	(0.040)	(0.049)	(0.028)	(0.547)	(0.053)	(0.039)
$1-\hat{\kappa}_C$ (2SLS)	-0.133	-0.035	0.144	-5.630	-0.001	-0.004	-0.179	-0.074	0.191	0.096
	(0.027)	(0.015)	(0.034)	(2.429)	(0.028)	(0.036)	(0.033)	(0.084)	(0.040)	(0.041)
Log Quadratic Demand System: Wage and Salary Households										
$\hat{\beta}_1$	-0.220	-0.279	0.059	0.231	0.011	-0.094	0.134	0.129	-0.045	-0.146
	(0.063)	(0.084)	(0.043)	(0.069)	(0.021)	(0.028)	(0.033)	(0.034)	(0.017)	(0.039)
$\hat{\beta}_2$	0.008	0.009	-0.001	-0.011	0.001	0.006	-0.010	-0.006	0.003	0.009
	(0.003)	(0.004)	(0.002)	(0.004)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Log Quadratic Demand System: Relative Expenditure Underreporting										
$1-\hat{\kappa}_C$	-0.234	0.222	0.836	0.303	0.968	0.217	0.114	-0.003	0.539	-0.102
	(0.516)	(0.276)	(0.465)	(0.172)	(0.156)	(0.247)	(0.202)	(0.285)	(0.184)	(0.202)

Notes: Robust standard errors are in parentheses. Standard errors of underreporting measures are asymptotic approximations using a robust estimate of the variance covariance matrix.

Appendix Table A1: Mapping Consumption Categories into the Harris and Sabelhaus Classification

Our Category	Harris and Sabelhaus Category
Food At Home	Food Off Premise (23)
Food Away From Home	Food on Premise (24), Food Furnished Employees (25)
Alcohol and Tobacco	Tobacco Products (26), Alcohol Off Premise (27), Alcohol on Premise (28)
Clothing and Personal Care	Clothing and Shoes (29), Clothing Services (30), Jewelry and Watches (31), Toilet Articles and Preparation (32), and Barbershops, Beauty Parlors, and Health Clubs (33)
Utilities	Electricity (38), Gas (39), Water and Other Sanitary Services (40), Fuel Oil and Coal (41), Telephone and Telegraph (42)
Domestic Services	Domestic Services, Other Household Operations (43)
Transportation	Tires, Tubes, Accessories, and Other Parts (53), Repair, Greasing, Washing, Parking, Storage, and Rental (54), Gasoline and Oil (55), Bridge, Tunnel, Ferry, and Road Tolls (56), Auto Insurance (57), Mass Transit Systems (58), Taxicab, Railway, Bus, and Other Travel Expenses (59)
Entertainment	Recreation Services (64)
Other Non Durables	Airline Fares for Out of Town Trips (60), Pari-Mutuel Net Receipts (65), Charitable and Political Giving (69)
Housing Services	See Text For Details

Notes: See online documentation to the NBER CEX files for examples of specific expenditures included in the Harris and Sabelhaus classification at http://www.nber.org/data/ces_cbo.html.

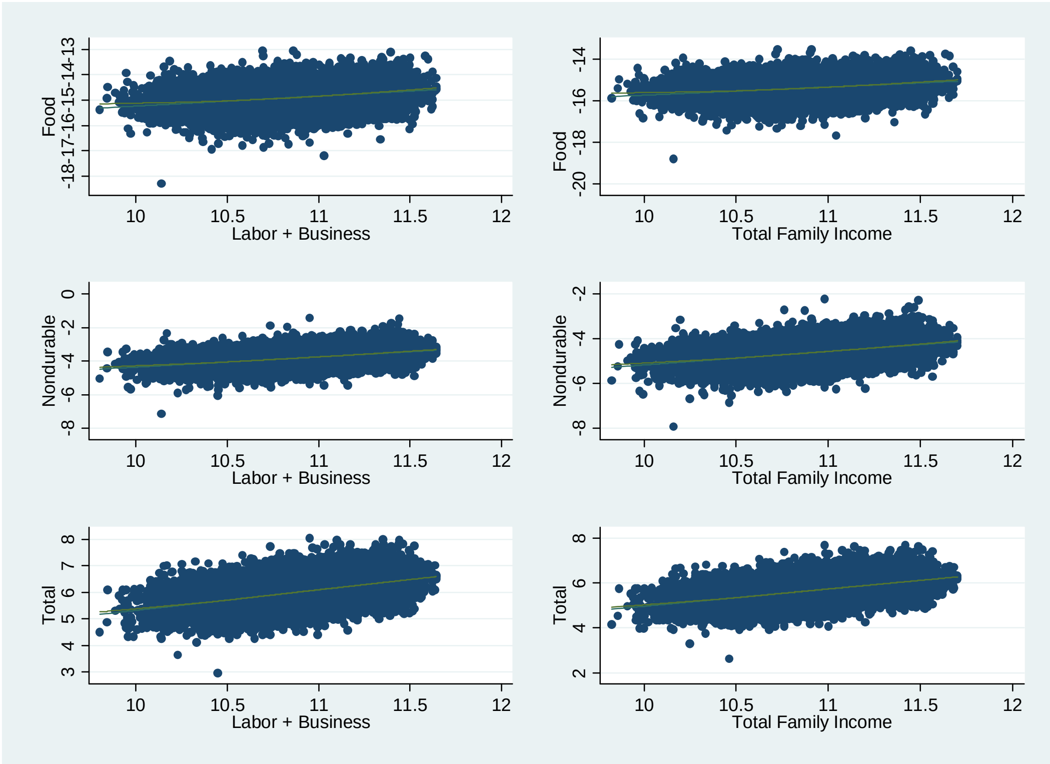

Figure R1: Augmented component plus residual plots of Engel curve for all income and expenditure measures