9-3-2009

Switching Regression Estimates of EIS for Stockholders and Non-Stockholders

Sheng Guo
Depatment of Economics, Florida International University, sguo@fiu.edu

SWITCHING REGRESSION ESTIMATES OF EIS FOR STOCKHOLDERS AND NON-STOCKHOLDERS

SHENG GUO
DEPARTMENT OF ECONOMICS
FLORIDA INTERNATIONAL UNIVERSITY

Abstract. This paper analyzes a panel data set of Panel Study of Income Dynamics (PSID) households and demonstrates that the estimate of EIS (Elasticity of Intertemporal Substitution) for stockholders and non-stockholders is large and different between them, based upon the consumption-based capital asset pricing model (CAPM). However, recognizing possible laxities in defining and measuring stockholding status, and hence allowing for possible misclassification error therein, I use the switching regression framework to show the evidence that there is a significant portion of stockholders misclassified as non-stockholders. The correction for this misclassification error results in closer gap of EIS between these two groups. Estimates after the correction are in line with those found in repeated cross-section Consumer Expenditure Survey (CEX) samples, whereas estimates without the correction are not. This illustrates the importance of accounting for misclassification error in such contexts. To some extent this result along with others of this research validates the use of repeated cross-section data in quatitative estimation of CAPM.

1. Introduction

EIS, defined as the elasticity of the ratio of consumption between two periods with respect to the relative price of consumption between the same two periods, is one of the most important economic concepts to study individuals' intertemporal consumption choices. Earlier literature explores economywide aggregate variables in estimation (Hansen and Singleton 1982, Hansen and Singleton 1983, Hall 1988). Hall (1988) argues that if approapriately estimated, the EIS should be very small, close to zero, or even in the negative territory. Recent literature utilizes micro data sets and attempts to identify the subset of individuals that are supposed to respond to changes of real interest rates (Vissing-Jørgensen 2002a, Attanasio, Banks, and Tanner 2002). These studies estimate EIS from consumption data often

Date: July 14, 2008.
Correspondence address: 11200 SW 8th Street, UP DM316, Miami, FL 33199. Email: Sheng.Guo@fiu.edu.

using repeated cross-section sample : variables of interest (such as consumption, family size change) are averaged across observations for each wave available to obtain a long time series of these variables. They find large estimates for stockholders or bondholders but not so for non-stakeholders, with estimates for the latter essentially complying with Hall's (1988) claims.

However, when measuring a person's stock market participation, this approach encounters its own problem: due to cross-section nature of data, no observations's market participation status can be observed over all of the years. If an observation appears more than once in consecutive interviews, her consumption growth can be defined and measured, so is her stockholding status from these adjacent periods; simple cross-section averages can be computed on such defined stockholders and non-stockholders just the same as those in Vissing-Jørgensen (2002a). Any systematic misclassification existing between these two groups will bias their cross-section averages, and this bias will not vanish no matter how large the sample size is. For example, if true stockholders tend to underreport their holdings or hide their stockholding status and hence are more likely to be subject to misclassification into non-stockholders, the bias in estimation due to this misclassification will never be corrected no matter how large the sample size of stockholders is. If, worse than that, one observation only appears once in the data, researchers would have to resort to computing sample averages of variables for the observations with propensity scores of stockholdings exceeding some predesignated critical value (Attanasio, Banks, and Tanner 2002). However, still, whether this propensity score is affected by misclassification error of stockholders versus non-stockholders is unable to be addressed, and if there is any, the estimates will still be contaminated. The availability of panel data enables me to evaluate the potential problem of misclassification and compare with those studies' results whether there will be any bias in estimates. Consequently, the second step of the estimation procedure employed below are not readily applied to cross-section averages. Moreover, in contrast to above studies, the availability of panel data where an observation's stockholding status is observed all over the years enables me to evaluate how severely the misclassification will affect estimates, or even how misclassifications differ according to different definitions. To my best knowledge, no other papers have taken on the issue of misclassification errors when estimating EIS seperately for stockholders and non-stockholders.

A call for evaluating the quality of stockholder status data, before categorizing observations according to its values, is not something peculiar. Research on aspects of other large-scale data sources has discovered significant misreporting error that is not affordable to ignore. Hausman, Abrevaya, and Scott-Morton (1998) find that job changers are more likely to misreport their status, i.e., to report that they have not changed their jobs. Poterba and Summers (1995) illustrates that as high as 10% of the truly unemployed were misclassified as not in the labor force in the Current Population Survey (CPS). In light of these results, it is hardly convincing that stockholding status data are not subject to any incidence of misclassification errors. In theory, stockholders should be those who optimize their intertemporal consumption at the interior solution. Any kinds of systematic measurement errors in stockholding status will lead to misclassification of stockholders versus non-stockholders in accordance to their theoretical definitions, and thus give rise to inconsistency in estimates.

Beyond the misclassification issue, my exploration of the PSID panel sample, which keeps track of same individuals over time, reveals that no systematic cross-section unobserved heterogeneity is evident in the sample, perhaps because differencing in logarithm of consumption has swept out much of the unobserved heterogeneity between individual observations. This piece of evidence is in support of the legitimacy of use of repeated cross-sections that have comprehensive information on consumption to construct long time-series for research on consumption-based asset pricing models, barring from sampling error.

The rest of this paper is organized as follows: Section 2 lays out the estimation CAPM model from the intertemporal rational choice framework; Section 3 gives an overview of the data and estimate the model without correction for misclassification; Section 4 motivates the possibility of misclassification linked to the current context, estimate the model with correction for misclassifications and discuss the results; Section 5 summarizes.

2. The Model

Consider the standard consumption-based asset pricing model with CRRA utility functions (hence the elasticity of intertemporal substitution (EIS) of consumption is not seperated from the risk aversion parameter) under complete market structure. This model, due to Hansen and Singleton (1982, 1984), underlies a large body of literature. Although this setup does not allow for the seperation between EIS and the coefficient of relative risk aversion (Hall 1988, Epstein and Zin 1989), it is fine given my goal of this chapter is to illuminate possible impacts of misclassification error on estimates.

For a household indexed by i at the time period t, augment the conventional utility function with observable control variables $W_{i,t}$, unobserved household-specific heterogeneity v_i and other uncounted disturbance effects and/or measurement error $\widetilde{u}_{i,t}$ that is orthogonal to v_i (Zeldes 1989, Attanasio and Low 2004)

$$U(C_{i,t}, W_{i,t}, v_i, u_{i,t}) = \frac{C_{i,t}^{1-\gamma}}{1-\gamma}\exp(\delta W_{i,t} + v_i + \widetilde{u}_{i,t}) \qquad (2.1)$$

This specification assumes the heterogeneity in consumption tastes can be decomposed into a common EIS, a common discount rate (ρ, see below) and other heterogeneous factors, namely, v_i and $\widetilde{u}_{i,t}$ across households.

The single most essential implication derived from this model is the stochastic Euler equation

$$E_t\left(\left(\frac{C_{i,t+1}}{C_{i,t}}\right)^{-\gamma}\exp(\delta\bigtriangleup W_{i,t+1}+\bigtriangleup\widetilde{u}_{i,t+1})\frac{1+R_{t+1}}{1+\rho}\right)=1 \tag{2.2}$$

where ρ is the discount rate, γ is the risk aversion parameter (the reciprocal of EIS), R_{t+1} is the real net rate of return of a particular asset between time t and $t+1$, and E_t is the expectations operator conditional on the agent's information set at time t. $C_{i,t}$ is the measured level of consumption as of time t. $\bigtriangleup W_{i,t+1} = W_{i,t+1} - W_{i,t}$, and likewise for $\bigtriangleup\widetilde{u}_{i,t+1}$. Note that household-specific effects, v_i, are swept out in the differencing.

In view of that data may not be available for every period of t, (2.2) needs to be modified so as to account for this possibility; especially, the rates of return from financial markets should be appropriately timed. For instance, for the time interval from t to $t+\tau$, the Euler equation is

$$E_t\left(\left(\frac{C_{i,t+\tau}}{C_{i,t}}\right)^{-\gamma}\exp(\delta\bigtriangleup W_{i,t+\tau}+\bigtriangleup\widetilde{u}_{i,t+\tau})\frac{1+R_{t,t+\tau}}{1+\rho}\right)=1 \tag{2.3}$$

in which $R_{t,t+\tau}$ refers to the rate of return from t to $t+\tau$. This is ensured by the law of no arbitrage: if the equation (2.3) fails to hold for any particular period from t to $t+\tau$, it implies the agent can increase her utility from reallocating consumptions from this period to other period, hence contradicts the conclusion of having achieved the optimal consumption.

Following the log-linearized approximation in Attanasio and Low (2004), equation (2.2) could be rewritten as (use $\sigma \equiv 1/\gamma$ to denote EIS)

$$\bigtriangleup\log C_{i,t+1} = \sigma\log(1+R_{t+1}) + \sigma\delta\bigtriangleup W_{i,t+1} + \bar{\alpha}_i + u_{i,t+1} \tag{2.4}$$

where the regression constant, $\bar{\alpha}_i$, includes the log of common discount factor, and the unconditional mean of second and higher moments of this household's consumption growth and rates of return. The residual $u_{i,t+1}$ includes $\bigtriangleup\widetilde{u}_{i,t+1}$, expectational errors, and deviation of second and higher moments of consumption growth and rates of return from the unconditional mean. A similar multi-period linearized equation will hold corresponding to equation (2.3).

Due to the inclusion of elements of second and higher moments of consumption growth and rates of return, $u_{i,t+1}$ will be correlated with the contemporaneous rate of return $\log(1+R_{t+1})$. Two-stage least squares can be used to address this concern : in the first stage, time-series stock returns are predicted from regressions of $\log(1 + R_{t+1})$ on the vector of instrumental variables (including righ-hand side variables themselves) $I_{i,t}$; this predicted stock return will then be plugged into place of (2.4) at the second stage. So long as some of variables in $\triangle W_{i,t+1}$ differ between households, the fitted value of R will differ between households even for the same period. Therefore I use $\log(1 + \widehat{R_{i,t+1}})$ to denote the fitted values of $\log(1 + R_{t+1})$ for household i. This step is somewhat a counterpart of obtaining cross-section averages in using repeated cross-section data.

If R refers to the rate of return in stock market, equation (2.2) (and hence (2.4)) only makes sense for stockholders, for they are the investors on the margin. Similarly, if R refers to bond returns, the same equation only makes sense for bondholders. The reasons why there is only limited participation in the stock market are out of the scope of this paper[1]. Following Vissing-Jørgensen (2002a), I will estimate equation (2.4) separately for stockholders and non-stockholders, but the parameters for the non-stockholder group are by no means eligible for structural interpretation, whereas the parameters for the stockholder group are. However, recognizing the definition, and hence the classification, of stockholders and non-stockholders are imperfect, a major contribution of this paper is to call for the switching regression estimation to account for possible misclassification, which will be detailed in Section 4.

3. The Data and Conventional Estimates

3.1. **Data Overview.** The PSID survey was conducted once a year from 1975 to 1996, and once every two years from 1997 to 2005. Financial variables such as rates of return are correspondingly timed in view of the two-year gaps from 1997 to 2005, as well as the three-year gap of 1987 to 1990 when food consumption data are not collected for 1988 and 1989. I restrict my sample to only those household heads who have headed the same household all over these years. The overall sample size composed by these households is 1214.

For consumpton measurements, each component of PSID consumption (food consumed at home, food eaten out, utilities and transportation expenses, *et cetera*) is deflated by the respective category-level CPI (2000Q1=100) before added up or used as predictors of the aggregate consumption measure. The first-quarter numbers of CPI series are used conforming to the timing of PSID surveys. The earliest and latest available year of this series are 1979 and 2003; therefore I use numbers of 1979 for years prior to 1979 and numbers of 2003 for the year 2004.

[1]Vissing-Jørgensen (2002b) and Attanasio and Paiella (2006) respectively propose transaction-cost based models to explain the limited market participation and estimate relevant parameters therein, under certain simplified assumptions.

CPI series and subsequent CEX data, which are used for aggregate consumption prediction later, are all from those used by Anguiar and Hurst (2008).

For stock market rates of return, I use NYSE value-weighted returns (including dividend distributions) deflated by CPI. The dividend-price ratio is calculated in the same way as in Fama and French (1988), and the bond horizon premium and bond default premium are defined in the same way as in Fama and French (1989) and Vissing-Jørgensen (2002a) and computed in Ibbotson Associates (2007). Stock market rates of return are correspondingly matched with respect to consumption growth according to (2.4), for instance, the rate of return in 1975, computed as the sum of distributed dividends and the capital appreciation from the end of year 1974 to the end of year 1975, is matched with the growth of consumption in 1974 to that in 1975.

For the definitions of stockholders and non-stockholders, I adopt two different measures. The first is to classify the observation (household head) as stockholders if his/her household owns positive stockholdings over all of the years between 1984 to 2005 when such information is ever available (denoted by the dummy variable D_3); the second is to classify the observation as stockholders if the household owns positive stockholdings in year 1984 (denoted by the dummy variable D_4), the first year when the survey began to ask for such information. Apparently the first definition is more stringent than the second, and the second is legitimate only under the assumption that once someone begins to hold stocks in the initial period, s/he is always "'in"' the stock market even if at certain points of time later s/he does not own any positive stockholdings[2].

Variables included in the probability prediction equation of stockholding status are all at their 1984 level. The wealth variable refers to the level of 1984 family wealth not including home equity. Some of the households' 1984 wealth is negative; to get a meaningful logarithm transformation, I add a constant number to each family's wealth value and then convert them into the logarithm scale. This will not change the ranking of wealth between households and using the transformed variable is legitimate for probability prediction purposes.

Table 1 displays the summary statistics of variables of interest for stockholders and non-stockholders defined by D_3. Compared to non-stockholders, stockholders are more wealthy, more educated, have smaller family size and consume more.

3.2. **Conventional Estimation Results.** I start with the estimation of a random-effect version which models $\bar{\alpha}_i$ as the persistent, unobserved hetereogeneity across households. Aggregate econmic shocks may hit upon all households for the same

[2]As predicted by a fixed transaction-cost in participation hypothesis, Vissing-Jørgensen (2002b) confirms that PSID households who participated in 1984 stock market are 31.8 percentage more likely to participate in 1989, controlling for a number of variables of observed heterogeneity.

year, therefore I cluster standard errors of $u_{i,t+1}$ by years. Yet estimation results yield no support for a cross-section distribution of $\bar{\alpha}_i$: its variance is essentially zero, regardless of whether instrument variables are used or not. A fixed-effect version does not work as well. This may be seen as the evidence in favor of the repeated cross-section average approach employed in the literature. Without controlling for unobserved fixed effects, the regressions are equivalent to pooling various years of a same household as different households' observations. Table 2 shows that estimated coefficients from GMM estimation are similar to those from OLS and IV estimation with slightly smaller standard errors. In almost every estimation method, the standard error increases after clustering adjustment whereas the managnitude of the estimated coefficient undergoes no change compared with no adjustment at all, in other words, statistical significance is lost after adjusting for clustered year effects. This is acceptable since I only have around 20 years of time-series dimension for the data.

For the first-stage estimation, one-period (one year prior to 1997, and two years after 1997) lagged dividend-price ratios, lagged bond horizon premia and bond default premia are used as the instrument variable. For example, to instrument for the stock real return from the end of year 1974 to the end of year 1975, the lagged dividend-price ratio refers to the sum of dividends for the year 1974 divided by the portofolio value at the end of 1974, so are the timings of agged bond horizon premia and bond default premia. This is consistent with the literature (Fama and French 1988, Fama and French 1989). The correlation between these instruments and logarithm of NYSE real returns is very strong, for instance, an OLS regression of logarithm of lagged dividend-price ratios on logarithm of NYSE real returns yields an coefficient 0.2278 with stardard error 0.0816 and ajusted R-squared 0.2135.

Table 2 presents the conventional estimates of EIS for stockholders and non-stockholders, as defined by the dummy variable D's. The top panel uses annual changes of family food consumption as the dependent variable. The bottom panel uses the similar measurement of aggregate consumption of nondurable goods and services, based on the methodology in Skinner (1987): predicting the aggregate consumption by its individual components available in PSID where the weight before each component is taken from the regression of these components from CEX samples. My regressors do not include value of house value or rent that is of durable consumption feature, following the preassumption that only non-durable goods and services are relevant consumptions considered here. The other difference from Skinner (1987) is that I use logrithm scale in regressions. While Skinner only conducts this kind of regression for the 1972–1973 CEX data, I extend this analysis from 1980 to 2002, with components chosen according to what are available in PSID. For years before 1980 and after 2002, weights are taken from those of year 1980 and of year 2002. Table 6 and 7 present CEX consumption prediction coefficients estimated for all the years 1980–2003. The overall fit of this sort of consumption prediction is at leat 0.6515 and is as large as 0.8607. The estimates of EIS by using food consumption are overall larger than by using aggregate consumption, perhaps partly because aggregate consumption measures incorporate some of those expenses that are not easily adjusted in the short run (such as utilities, housing supplements) or related to habit consumption (such as alcohol or tobacco).

Even if only focusing on EIS estimates from aggregate consumption and the conventional estimation methods, I find the magnitude of at least 0.7 for stockholders, which is much greater than Vissing-Jørgensen's (2002a) findings of 0.3 for stockholders, although she does find similar magnitude for bondholders. My results differ significantly from hers, especially for 2SLS and GMM estimates, which very much agree with each other. At the first sight this may be probably because of the following reasons: first, my panel data is insulated from attenuation bias caused by changes on the extensive margin (more or less stockholders) and/or sampling error for each period of Vissing-Jørgensen's (2002a) study due to aggregation; second, my selected sample is not representative of the overall population, but rather, the middle to old age households of the population. This can be seen from that the average age of household heads in 1984 of my sample is about 45. Surprisingly, the next section shows that correction for the misclassification error sweeps away much of the discrepancy.

4. Accounting for Misclassification Error

This section presents that it turns out quantitatively important whether the misclassification error has been corrected for and taken care of in subsequent estimations. Conditional on the true status of being stockholders, there exists a significant chance of misclassification compared to the true status being non-stockholders. The consequence is that the magnitude difference of EIS between true stockholders and non-stockholders are exaggerated. This can be clearly seen from Figure 1, where linear instrumental estimates after correction are 0.34 and 0.02 respectively, versus 1.02 and 0.12 before. This section explains the motivation and empirical procedure of the correction.

4.1. **The Misclassification of Stockholders versus Non-stockholders.**

The theoretical motivation for empirical distinctions between stockholders and non stockholders is that stockholders are believed to be the group of persons who are actively adjusting their marginal rates of substitution of consumption in alignment with the marginal rate of transformation in the economy. These are the type of persons who, when expecting the market return to be high, will defer their consumption to later periods. In constrast, since non-stockholders hold no stakes to exploit the stock market opportunity, they will have no incentives to do so. In other words, a stockholder's consumption bundle is the interior solution of the intertemporal optimization problem that s/he faces, which is not the case for non-stockholders, for non-stockholders' nonholding status of stocks suggests they have not utilized the stock market to adjust their intertemporal consumption allocation. Let $\Delta = 1$ if someone is a stockholder (as defined in the above by its intertemporal optimization implication), and $\Delta = 0$ if the person is not.

It is at much of the researcher's liberty when it comes to discern the stockholders and non-stockholders. Some authors try various thresholds of stockholdings to decide. For instance, Mankiw and Zeldes (1991) try three splitting strategies: thresholds of \$0, \$1,000 and \$10,000 in stockholdings respectively; while acknowledging a perfect separation is not possible, Vissing-Jørgensen (2002a) refers to households with positive responses to stock categories as stockholders and the rest as non-stockholders, and also tries splitting the whole sample into three layers based on these households' stockholding levels. Likewise, Fillat and Garduño (2005) vary the definition thresholds of asset holders on a finer grid. Remarkably, both Vissing-Jørgensen (2002a) and Fillat and Garduño (2005) fail to find the monotonicity relationship that is thought to exist when the definition is tightened.

The stockholding data of PSID households are only collected once every five years during the period of 1984 – 1999, and every other year during 1999 – 2005 when the survey itself became biannual after 1999. Up until 1994, the set of questions about stockholdings is concerned with stocks in all categories, including those in IRA's:

> (1994 Questionaire) G129. Do you (or anyone in your family living there) have any shares of stock in publicly held corporations, mutual funds, or investment trusts, including stocks in IRA's?

Starting from 1999, PSID survey has changed the question into not including stocks in IRA's:

> (1999 Questionaire) W15 (G129). Do you (or anyone in your family living there) have any shares of stock in publicly held corporations, mutual funds, or investment trusts, not including stocks in employer-based pensions or IRA's?

As discussed above, I focus on dummy variables derived from answers to this question in each available year.

Recall the first of the two dummy variables I have defined: $D_3 = 1$ if the respondent answered "Yes" to this question in all of the years, and $D_3 = 0$ if s/he answered "No" in all of the years. Apparently this definition of D_3 is already very tight. The aim of constructing this dummy is to consistently characterize whether someone is a stockholder or not during the periods that I will examine. I assume that for any particular household, the status of stockholder or non-stockholder will not switch from time to time, even if for some years data are not observed. Therefore the focus of this paper is on imperfect observed, time-invarying status of stockholding[3].

That stockholding status may be imperfect may arise due to genuine measurement errors, such as survey interviewers mistakenly check the wrong box, or respondents systematically misreport the holdings. Moreover, when it comes to the difference in risk preference between stockholders and non-stockholders, the misclassification may also come from the failure to distinguish IRA versus non-IRA stockholdings and the change of framed survey questions as noted above, for its asset pricing implications of long-term versus short-term. The examination of individuals' annual (or biannual) consumption change within the interpretation of CAPM indicates the focus on the short-term consumption behavior. But investing in an IRA account often calls for a long-horizon investing mindset for individual investors, because they cannot withdraw earnings from this sort of retirement accounts without incurring penalties, before achieving certain age criteria; in view of that our measure of consumption change is year by year, long-horizon stockholdings amount to non-stockholdings for the former does not necessarily imply contemporaneous consumption adjustments in response to stock market movements. Daniel and Marshall (1997) demonstrate that longer-horizon (thus low-frequency) implications of some asset pricing models may help close the equity premium puzzle that may arise otherwise; based upon their work, including IRA stockholders may blur the real cross-sectional difference between stockholders and non-stockholders. This is another reason that the value of the constructed variable D_3 may diverge from that of the true underlying variable Δ.

According to the above discussion, we have that conditional on the observed status D_3

$$0 < \Pr(\Delta \mid D_3) < 1 \tag{4.1}$$

which spells out the misclassification that may exist in the sample. (4.1) implies $0 < \Pr(\Delta = 0 \mid D_3 = 0) < 1$, which indicates that even if some of the households claimed to hold zero stocks all over the years, they may indeed misrepresented their status in the survey, perhaps just to avoid the hassle of having to check current value of their portofolios. Another (rare) possibility is that some of these households happened to have liquidized or nearly liquidized their stockholdings around the time of interviews that ask questions about stockholdings, and this sort of on-and-off investing profile still qualifies them as the inverstors on the margin, i.e., with the interior solution at the corner location of budget sets.

[3]The case of imperfectly observed, time-varying status can be extended based upon the approach outlined in this paper.

The proxy of D_3 for Δ means that those who claimed they have positive stockholdings are more likely to have positive non-IRA holdings than those who claimed otherwise. This can be written as:

$$\Pr(\Delta = 1 \mid D_3 = 1) > \Pr(\Delta = 1 \mid D_3 = 0) \tag{4.2}$$

Notice that it immediately implies

$$\Pr(\Delta = 0 \mid D_3 = 0) > \Pr(\Delta = 0 \mid D_3 = 1) \tag{4.3}$$

Our estimation scheme of switching regressions involves two steps. In the first step, the goal is to obtain the consistent propensity score of stock ownership given switching variables and D, the measurement of true Δ with random error. The Hausman, Abrevaya, and Scott-Morton's (1998) version of Monotonicity Condition, required for identification for this step, is conditional on true status Δ rather than on measured status D_3

$$\Pr(D_3 = 0 \mid \Delta = 1) < \Pr(D_3 = 0 \mid \Delta = 0) \tag{4.4}$$

thus is different from (4.2). In fact, (4.4) is stronger than (4.2).To see this, notice that by the Bayesian Law (4.4) amounts to

$$\frac{\Pr(\Delta = 1 \mid D_3 = 0)\Pr(D_3 = 0)}{\Pr(\Delta = 1)} < \frac{\Pr(\Delta = 0 \mid D_4 = 0)\Pr(D_4 = 0)}{\Pr(\Delta = 0)}$$

which is equivalent to

$$\Pr(\Delta = 0 \mid D_3 = 0) > \Pr(\Delta = 0) \tag{4.5}$$

Symmetrically one can get

$$\Pr(\Delta = 1 \mid D_3 = 1) > \Pr(\Delta = 1) \tag{4.6}$$

Apparently combining (4.5) and (4.6) leads to (4.2) but not vice versa.

Now define α_0 and α_1 as

$$\alpha_0 = \Pr(\Delta = 1 \mid D_3 = 0) \qquad \alpha_1 = \Pr(\Delta = 0 \mid D_3 = 1)$$

which hereby maintains the assumption of misclassification error independent of individual characteristics (Poterba and Summers 1995, Hausman, Abrevaya, and Scott-Morton 1998). We are to specify a vector of variables, Z_i, that will shift the propensity of owning stocks but not the misclassification. Under normal distributions of prediction disturbance term in prediction of stockholding status, the likelihood function is the same as that in Hausman, Abrevaya, and Scott-Morton (1998) [4]

$$\begin{aligned}\mathbf{L} = \sum_{i=1}^{N} \Big\{ & D_{3,i} \log(\alpha_0 + (1-\alpha_0-\alpha_1)\Phi(Z_i'\gamma)) \\ & +(1-D_{3,i})\log(1-\alpha_0-(1-\alpha_0-\alpha_1)\Phi(Z_i'\gamma))\Big\}\end{aligned} \tag{4.7}$$

Suppose after the correction we obtain consistent estimates of propensity score $p(Z_i) \equiv \Pr(\Delta_i = 1 \mid Z_i) = \Phi(Z_i'\hat{\gamma})$, the population regression of $\triangle \log C_{i,t+1}$ is the propensity-score weighted two underlying ones

$$\begin{aligned}\mathbf{E}(\triangle \log C_{i,t+1}) &= p(Z_i)\mathbf{E}(\triangle \log C_{i,t+1} \mid \Delta_i = 1) + (1-p(Z_i))\mathbf{E}(\triangle \log C_{i,t+1} \mid \Delta_i = 0) \\ &= p(Z_i)\left[\sigma_1 \log(1+R_{t+1}) + \sigma_1\delta_1 \triangle W_{i,t+1}\right] \\ &+ (1-p(Z_i))\left[\sigma_0 \log(1+R_{t+1}) + \sigma_0\delta_0 \triangle W_{i,t+1}\right]\end{aligned}$$

In presence of variations of $p(Z_i)$ across observations and that R_{t+1} would be instrumented by individual characteristics, σ_1 and σ_0 can be separately identified. However, anticipating possible population moments between $p(Z_i)$ and other regressors in (2.4), no general results are available regarding how estimates of coefficients from this propensity-score weighted regressions change compared to before.

In particular, the instrumental variable estimation will be likewise propensity weighted: recall that in the conventional estimation the first-stage is to obtain the fitted values of $\log(1 + R_t)$ from regressions of $\log(1 + R_t)$ on a host of instrumental variables $I_{i,t}$ for each group classified by the dummy variable D. Now the only difference is that both $\log(1 + R_t)$ and $I_{i,t}$ will be correspondingly weighted by consistent propensity-scores, as D is not a perfect measure of Δ. This can be seen from

$$\begin{aligned}\mathbf{E}(\triangle \log C_{i,t+1}) = p_i & \left[\sigma_1 \log(1+R_{t+1} \mid_{\Delta_i=1}) + \sigma_1\delta_1 \triangle W_{i,t+1} \mid_{\Delta_i=1} + \mathbf{E}(u_{i,t+1} \mid \Delta_i = 1)\right] \\ + (1-p_i) & \left[\sigma_0 \log(1+R_{t+1} \mid_{\Delta_i=0}) + \sigma_0\delta_0 \triangle W_{i,t+1} \mid_{\Delta_i=0} + \mathbf{E}(u_{i,t+1} \mid \Delta_i = 0)\right]\end{aligned}$$

where $|_{\Delta_i}$ denotes conditioning on the true type Δ_i. Although $\log(1 + R_{t+1} \mid_{\Delta_i=1}) = \log(1 + R_{t+1} \mid_{\Delta_i=0})$ in above expression because of the complete market assumption, technically $Cov(\log(1 + R_{t+1} \mid_{\Delta_i=1}), u_{i,t+1} \mid \Delta_i = 1) \neq Cov(\log(1 + R_{t+1} \mid_{\Delta_i=0}), u_{i,t+1} \mid \Delta_i = 0)$, for the correlation of consumption growth and predicted rates of return are supposedly different for stockholders versus non-stockholders. For a two-stage implementation, naturally the first-stage is to obtain the predicted value of

[4]Note that correction for misclassification error does not require parametric distributional assumptions such as normal distributions (Lewbel 2000), although I find probit setup is adequate for this study.

the whole term $p_i \log\widehat{(1 + R_{i,t+1})}$ from regressions of $p_i \log(1 + R_{t+1})$ on the vector of instrumental variables $p_i I_{i,t+1}$ (including $p_i \vartriangle W_{i,t+1}$), and likewise for obtaining $(1 - p_i)\,\widehat{\log(1} + R_{i,t+1})$. That this is consistent with no misclassification case can be verified from pushing p_i to the limit of one or zero.

4.2. **The Two-step Estimation Results of EIS.** The subsequent estimation follows the two-step procedure outlined in last subsection. I apply it to both mismeasurements of Δ. The dummy variable D_3 equals to one if the household owns any positive stockholdings for *all* the years, and equals to zero if it does not own any holdings for *all* the years; D_4 equals to one if the family owns any positive stockholdings in year 1984, the first year when the information about stockholdings became available. Note that by their sheer definitions the set of D_4 includes the set of D_3.

Numerous studies have found that the wealthy and educated are more likely to invest in stock markets, for instance, Mankiw and Zeldes (1991) find that people with higher income tend to have greater stockownership, and that higher proportion of stockholders is visible for households with college degree (but not for more advanced degrees). Therefore I choose to use age quadratic, wealth, education quadratic, and wealth-education interactions in predicting the propensity score of stockownership, with the underlying assumption that these variables will not affect the misclassification probability. The Pseudo R^2 of probit regression for predicting D_3 is 0.4752, which is better than the number (0.1203) reported in Attanasio, Banks, and Tanner (2002), although Pseudo R^2 is by no means the definitive measure of goodness of fit. The Pseudo R^2 for predicting D_4 is 0.1302.

In the first step of probit estimation with misclassifications, I experiment with various sets of predicting variables and numerical optimization algorithms lest that the likelihood function may have non-concave regions, and they all yield very similar estimates for α_0 and α_1. Table 3 presents that under normal assumptions of predicting error, there is essentially no misclassification for those who report they own some stocks. On the other hand, it is both economically and statistically significant that a non-trival proportion of non-stockholders may be misclassified: their true state is one, yet are categorized as zero by D_3 or D_4. The error of misclassifying stockholder as non-stockholder is about 16% for the dummy variable D_3, and the same error is about 35% for the dummy variable D_4. This difference in misclassfiication error may be due to the less strict criterion by D_4. After correction for the misclassification of D_3, the education variable exhibits slightly concave shape of predicting stockownership, which is consistent with Mankiw and Zeldes's (1991) findings. Employing logit regressions generates very similar results.

Table 4 reports the consistent propensity-score weighted OLS estimates of EIS for the stockholding status (D_3) defined as positive holdings over all of the years (and 5 reports those for D_4, defined as positive holdings in 1984)[5]. I obtain EIS estimates of about 0.80 for stockholders and about 0.62 for non-stockholders. This pair of estimates only slightly closes the gap from 0.81 and 0.58 that we have obtained before. Intuitively, if more stockholders are misclassified as non-stockholders versus the otherwise, we may expect the true gap of estimates for these two groups after correction to be closer than estimates from without correction. But this intuition neglects the perhaps complicated correlation pattern between propensity scores and interacted regressors, which bars from a simple and general conclusion. Moreover, estimates reveal that this intuition works out more in the case of D_3 rather than D_4, perhaps because D_4 is itself a very weak definition of stockholders. Meanwhile, after correction, the estimates are more robust to which crude category variable, D_3 or D_4, is used.

The instrumental estimation results are more interesting. The instruments are supposed to correct for the endorgeneity of $\log(1 + R_{t+1})$ and $u_{i,t+1}$. In estimation directly using the dummy variable D_3, instrumental variable estimations open the gap of EIS estimates for stockholders versus non-stockholders, from 0.81 versus 0.58 to 1.02 versus 0.12 (see Figure 1). The inference from this result is that stockholders possess idiosyncratic shocks negatively correlated with real rates of return of the market while non-stockholders possess positively related ones.

However, in propensity-weighted estimation, instrumental variable estimations not only drive down the estimates, but also dramatically close the gap, from 0.80 versus 0.62 to 0.34 versus 0.02. Without adjustment for clustered year effects, the coefficient for corrected non-stockholders 0.02 is not statistically significant, but 0.34 for corrected stockholders is statistically significant. The same pattern occurs to the case of D_4, which yields almost the same magnitude of EIS for corrected stockholders. This is in line with the estimate around 0.3 obtained by Vissing-Jørgensen (2002a) for stockholders. Also, the estimate for non-stockholders, the majority of the population, fits into what Hall (1988) has claimed. The explanation for driven-down coefficients from propensity-score instruments may stem from the fact that those negatively correlated idiosyncratic shocks that are supposed to be associated with stockholders now are to be associated with non-stockholders, due to the misclassification issue. Likewise, the explanation for closing gap from instruments would be that as instruments detect more difference between two contaminated groups, the true difference ought to be less, because we know there

[5]Since a GMM estimation on top of the switching regression framework is unexplored in this research, in what follows, I only present pooled OLS and IV estimates with robust and cluster-ajusted standard errors.

is a great proportion of one group misclassified as another group. Again it is worthwhile to remind that *ex ante* we cannot tell how instrumental estimation will change the results, as the correlation of propensity score with other regressors may be complicated.

5. Conclusion

This paper examines the impacts of correction for misclassification errors between stockholders and non-stockholders on estimating EIS based upon consumption based asset pricing models. Estimates reveal that there is a significant proportion of stockholders are misclassified as non-stockholders in the PSID data, which leads to spuriously large estimates of EIS. I show that accounting for this incidence of misclassificaiton error by applying the switching regression framework tends to close the gap of the EIS estimates for these two groups. Controlling for the misclassification error yields results in line with those found in the literature; otherwise the estimates of EIS from PSID are much larger. These results suggest taking care of possible misclassifications a prerequisite step for structural estimation on two seperately categorized groups. Beyond the misclassification issue, my estimates suggest unobserved individual heterogeneity of consumption does not seem to play an evident role in estimating the version of CAPM considered here.

Since estimates from PSID data with the misclassification correction align with those from CEX without the misclassification correction, it is natural to ask whether there is any misclassification error in CEX and how it affects these estimates. It is not so obvious how the propensity score weighted scheme is adapted to cross-section averages, although misclassification error is similarly straightforward to obtain. This can be a future direction of research.

References

Anguiar, M., and E. Hurst (2008): "Deconstructing Lifecycle Expenditure," Manuscript, Graduate School of Business, University of Chicago.

Attanasio, O. P., J. Banks, and S. Tanner (2002): "Asset Holding and Consumption Volatility," *The Journal of Political Economy*, 110(4), 771–792.

Attanasio, O. P., and H. Low (2004): "Estimating Euler Equations," *Review of Economic Dynamics*, 7, 406–435.

Attanasio, O. P., and M. Paiella (2006): "Intertemporal Consumption Choices, Transaction Costs and Limited Participation to Financial Markets: Reconciling Data and Theory," NBER Working Paper 12412.

Daniel, K., and D. Marshall (1997): "Equity-premium and Risk-free-rate Puzzles at Long Horizons," *Macroeconomic Dynamics*, 1, 452–484.

Deaton, A. (1992): *Understanding Consumption*. Oxford University Press.

Epstein, L. G., and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57(4), 937–969.

Fama, E. F., and K. R. French (1988): "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22(1), 3–25.

——— (1989): "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics*, 25(1), 23–49.

Fillat, J., and H. Garduño (2005): "GMM Estimation of an Asset Pricing Model with Habit Persistence," Working paper, Department of Economics, University of Chicago.

Hall, R. E. (1988): "Intertemporal Substitution in Consumption," *The Journal of Political Economy*, 96(2), 339–357.

Hansen, L. P., and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50(5), 1269–1286.

——— (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *The Journal of Political Economy*, 91(2), 249–265.

——— (1984): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models: Errata," *Econometrica*, 52(1), 267–268.

Hausman, J. A., J. Abrevaya, and F. M. Scott-Morton (1998): "Misclassification of the Dependent Variable in a Discrete-response Setting," *Journal of Econometrics*, 87(2), 239–269.

Hsiao, C. (2003): *Analysis of Panel Data*, Econometric Society monographs: no. 34. Cambridge University Press, 2nd edn.

Ibbotson Associates (2007): *Stocks, Bonds, Bills, and Inflation 2007 Yearbook*. Morningstar Inc.

Lewbel, A. (2000): "Identification of the Binary Choice Model with Misclassification," *Econometric Theory*, 16(4), 603–609.

Mankiw, N. G., and S. P. Zeldes (1991): "The Consumption of Stockholders and Nonstockholders," *Journal of Financial Economics*, 29, 97–112.

Poterba, J. M., and L. H. Summers (1995): "Unemployment Benefits and Labor Market Transitions: A Multinomial Logit Model with Errors in Classification," *The Review of Economics and Statistics*, 77(2), 207–216.

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

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

——— (2002b): "Toward an Explanation of Household Portofolio Choice Heterogeneity: Nonfinancial Income and Participation Cost Structures," Manuscript, Department of Economics, University of Chicago.

Zeldes, S. P. (1989): "Consumption and Liquidity Constraints: An Empirical Investigation," *The Journal of Political Economy*, 97(2), 305–346.

Table 1. Summary statistics for stockholders and non-stockholders defined as positive stockholdings all over the years

Variable	Mean	Std. Dev.	Min	Max
Definition of D_3[(a)]	*"non-stockholders" (425 obs.)*			
wealth in 1984 (1,000 dollars)	21.68	68.47	-8.70	887.00
age at 1984	44.6	11.7	27	75
education as of 1984	10.8	3.0	0	17
family size of 1984	3.39	1.75	1	11
food consumption of 1984	6093.8	3564.2	16.2	25387.8
aggregate consumption of 1984[(b)]	20712.8	11173.3	6.4	72698.44
	"stockholders" (121 obs.)			
wealth in 1984 (1,000 dollars)	209.06	519.31	-2.60	4720.00
age at 1984	45.4	10.8	30	70
education as of 1984	15.0	2.1	8	17
family size of 1984	3.05	1.28	1	7
food consumption of 1984	9842.2	8906.8	2054.4	94659.1
aggregate consumption of 1984[(b)]	32221.9	23234.1	6539.18	240475.3
stockholdings (1,000 dollars)[(c)]				
1984	50.00	94.54	0.10	750.00
1989	137.52	57.37	0.60	5000.00
1994	318.46	964.91	0.00	9999.99
1999	345.17	742.46	1.00	7000.00
2001	443.90	1058.34	0.00	9500.00
2003	608.06	1767.21	1.00	15000.00
2005	517.40	952.73	0.00	7000.00

Note: (a) D_3 =1 if observations of the PSID longitudinal sample own positive stockholdings in all years from 1984 to 2005, and D_3 =0 if observations of the PSID longitudinal sample own zero stockholdings in all years from 1984 to 2005; (b) aggregate consumption of each year is predicted by individual consumption components of PSID with weights taken from regressions on CEX data of aggregate consumption on the same set of individual consumption compoeenents for the same year; both measures of food consumption and aggregate consumption are deflated by CPI respectively; (c) including top-coded cases (coded as 9,999,997 if 9,999,997 or more), and not including bracket-value cases. From 2001 the question only pertains to non-IRA stocks.

Table 2. Conventional estimates of EIS without correction for misclassification error

Food Consumption	OLS	OLS (clustered by year)	2SLS	GMM
Definition D_3				
Stockholders	1.2137	1.2137	1.8395	2.0194
(Obs. = 121)	(0.1664)	(0.4559)	(1.0813)	(0.9278)
Non-stockholders	0.8201	0.8201	1.0001	0.9843
(Obs. = 425)	(0.1073)	(0.3673)	(0.6531)	(0.6082)
Definition D_4				
Stockholders	1.0694	1.0694	1.5317	1.6826
(Obs. = 398)	(0.0862)	(0.4092)	(0.9254)	(0.7940)
Non-stockholders	0.9544	0.9544	1.0305	0.9863
(Obs. = 816)	(0.0682)	(0.3651)	(0.6405)	(0.5891)
Aggregate Consumption				
Definition D_3				
Stockholders	0.8062	0.8062	1.0235	1.0605
(Obs. = 121)	(0.1042)	(0.3788)	(0.8421)	(0.6605)
Non-stockholders	0.5763	0.5763	0.1193	0.2469
(Obs. = 425)	(0.0742)	(0.3307)	(0.6054)	(0.5516)
Definition D_4				
Stockholders	0.6958	0.6958	0.7780	0.8229
(Obs. = 398)	(0.0569)	(0.3431)	(0.7053)	(0.5479)
Non-stockholders	0.6827	0.6827	0.3126	0.4067
(Obs. = 816)	(0.0459)	(0.3408)	(0.5774)	(0.4540)

Notes: D_3=1 if observations of the PSID longitudinal sample owning positive stockholdings in all years from 1984 to 2005, and D_3=0 if owning no stockholdings in all years from 1984 to 2005; D_4=1 if observations of the PSID longitudinal sample owning positive stockholdings in year 1984, and D_4=0 if observations of the PSID longitudinal sample owning no stockholdings in year 1984. Instrumental variables are employed to account for the potential endorgeneity of stock market returns. Results in the second, third, and last columns are all adjusted for clustering year effects in standard errors.

TABLE 3. Probit regression of stockholding status with and without correction for misclassfication error

	No Correction	With Correction	No Correction	With Correction
Definitions of Stockholders	(D_3) (546 observations)		(D_4) (1214 observations)	
α_0	——	0.0112	——	0.0141
		(0.0076)		(0.0137)
α_1	——	0.1595	——	0.3536
		(0.0450)		(0.0374)
age of 1984	0.1156	0.1559	0.0859	0.0210
	(0.0774)	(0.1214)	(0.0395)	(0.0713)
age of 1984 (squared)	-0.0006	-0.0015	-0.0005	-0.0003
	(0.0006)	(0.0010)	(0.0003)	(0.0006)
education level of 1984	0.3514	2.3806	0.4800	0.1981
	(0.3945)	(1.5113)	(0.1748)	(0.2999)
education level of 1984 (squared)	0.0068	-0.0656	-0.0097	-0.0018
	(0.0115)	(0.0521)	(0.0052)	(0.0085)
age × education	-0.0042	-0.0027	-0.0016	0.0007
	(0.0032)	(0.0061)	(0.0016)	(0.0031)
log of wealth level in 1984	2.5664	9.5141	0.6466	9.1984
	(0.3859)	(1.9496)	(0.1209)	(2.1610)
(constant)	-40.7364	-143.7744	-15.1116	-119.8121
	(5.8980)	(29.6752)	(2.2813)	(27.5089)
maximized loglikelihood	-151.5622	-142.0576	-668.0536	-628.1550
Pseudo R^2	0.4752		0.1302	

Note: probit prediction of stockholding ownerships for D_3 (=1 if owning positive stockholdings in all years from 1984 to 2005) and D_4 (=1 if owning positive stockholdings in year 1984), with correction of misclassification and without; α_0 is defined as probability of true state being of D=0 yet misclassified as D=1; α_1 is defined as probability of true state being of D=1 yet misclassified as D=0; standard errors in parenthesis.

Table 4. Comparision of direct and misclassification corrected EIS estimates for stockholders and non-stockholders defined as positive stockholdings all over the years (D_3)

EIS Estimates	OLS	2SLS
	(all clustered by year)	
D_3: dummy variable		
Stockholders	0.8062	1.0235
(Obs. = 121)	(0.3788)	(0.8421)
Non-stockholders	0.5763	0.1193
(Obs. = 425)	(0.3307)	(0.6054)
D_3: propensity score (corrected for misclassification error)		
Stockholders	0.8037	**0.3442**
	(0.3596)	(0.5048)
Non-stockholders	0.6171	**0.0248**
	(0.3305)	(0.4609)

Notes: D_3=1 if observations of the PSID longitudinal sample are reported to have owned positive stockholdings in all years from 1984 to 2005; D_3=0 if observations of the PSID longitudinal sample are reported to have owned zero stockholdings in all years from 1984 to 2005. The propensity score-weighted regression is $E(Y)=pX\beta_1+(1-p)X\beta_0$, where $p=Pr(\Delta=1 \mid Z)$ is the consistent propensity scrore after correcting for misclassification in the probit regression. Insturment variables (lagged dividend-price ratios, lagged bond default premia, and lagged bond horizon premia) are used to account for the endorgeneity of stock market returns in X, also adjusted by propensity scores in the bottom panel. See text for details on instrumental estimation for the misclassification case.

TABLE 5. Comparision of direct and misclassification corrected EIS estimates for stockholders and non-stockholders defined as positive stockholdings in 1984 (D_4)

EIS Estimates	OLS	2SLS
	(all clustered by year)	
D_4: dummy variable		
Stockholders	0.6958	0.7780
(Obs. = 398)	(0.3431)	(0.7053)
Non-stockholders	0.6827	0.3126
(Obs. = 816)	(0.3408)	(0.5774)
D_4: propensity score (corrected for misclassification error)		
Stockholders	0.7834	**0.3583**
	(0.3517)	(0.4850)
Non-stockholders	0.5830	**-0.0712**
	(0.3322)	(0.4828)

Notes: D_4=1 if observations of the PSID longitudinal sample are reported to have owned positive stockholdings in all years from 1984 to 2005; D_4=0 if observations of the PSID longitudinal sample are reported to have owned zero stockholdings in all years from 1984 to 2005. The propensity score-weighted regression is $E(Y)=pX\beta_1+(1-p)X\beta_0$, where $p=Pr(\Delta=1 \mid Z)$ is the consistent propensity scrore after correcting for misclassification in the probit regression. Insturment variables (lagged dividend-price ratios, lagged bond default premia, and lagged bond horizon premia) are used to account for the endorgeneity of stock market returns in X, also adjusted by propensity scores in the bottom panel. See text for details on instrumental estimation for the misclassification case.

TABLE 6. 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	
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 7. Non-durable goods and services consumption prediction regressions from CEX samples of various years (continued)

Year	Obs.	Food		Utility		Transportation		(Constant)		R squared
		Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	
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

Note: CEX sample for each year (see Aguiar and Hurst (2008) 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.

FIGURE 1. Graph of EIS estimates by the dummy variable and by the propensity score for stockholders and non-stockholders defined as positive holdings in all the years (D_3)

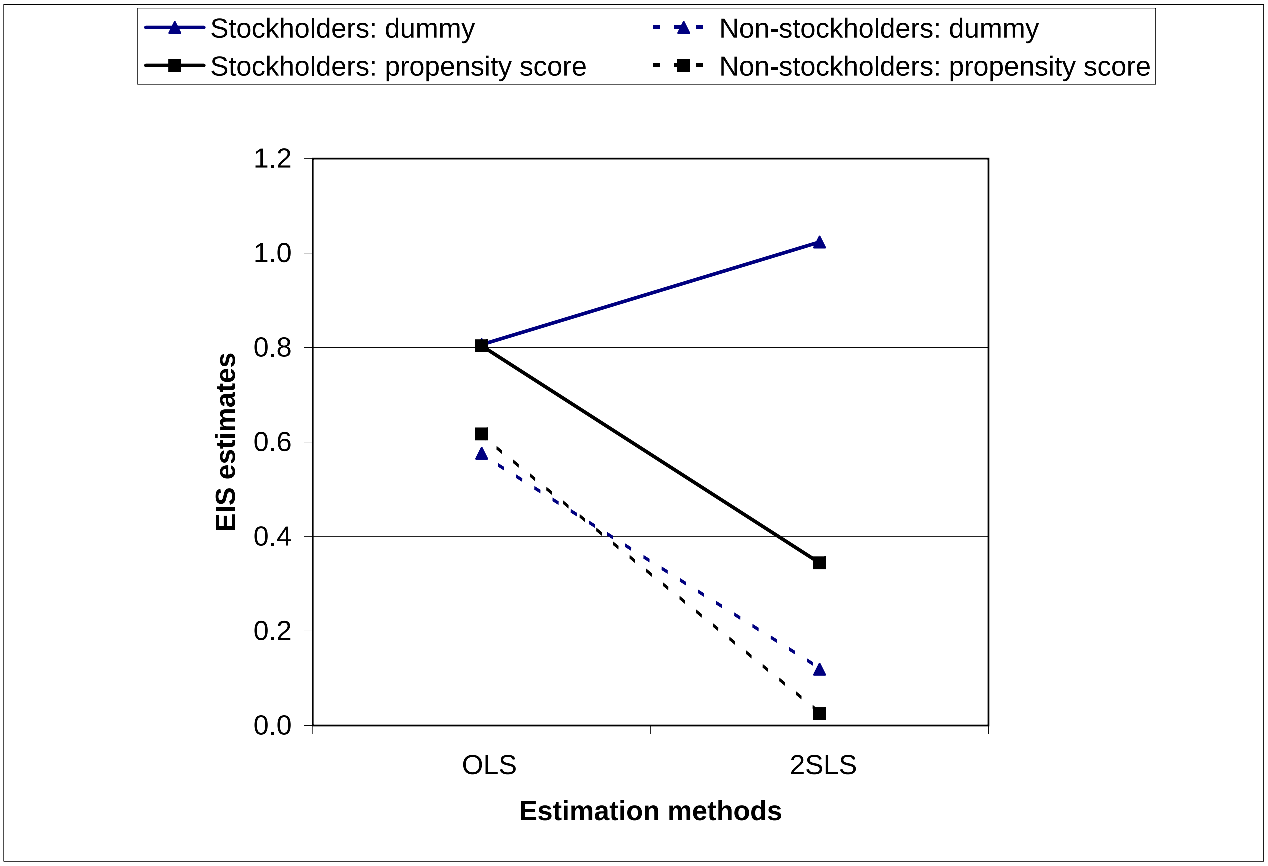

Figure 2. Graph of EIS estimates by the dummy variable and by the propensity score for stockholders and non-stockholders defined as positive holdings in 1984 (D_4)