Underconfidence and Stock-Market Participation*

Dou, Yiwei

Department of Economics
York University,
Toronto, Canada M3J 1P3
douyiwei@econ.yorku.ca

May, 2006

Abstract

Although stocks seem as an attractive asset with high average returns and a low covariance with consumption growth, limited stock-market participation still appears. Since risk aversion, and entry cost issues cannot explain this puzzle completely, I add underconfidence of investors into these lines. Through a simple model and empirical work using Canadian micro data (Survey of Consumer Finances 1984), my study shows that biased estimation of ability exacerbates entry cost and risk aversion effect to prevent underconfident households—those defined as not to hold mortgage, personal debt, or personal debt other than consumer debt—from stock-market. Moreover, This paper also indirectly lends a support to the possibility of existence of overconfidence in stock-market inasmuch as absence of underconfident investors.

Keywords: limited participation, risk aversion, entry cost, underconfidence

JEL Classification: G11, D19

* The author is grateful to Prof. Sule Alan for her helpful suggestion.. My thanks also go to Laine G.M. Ruus from Data library in University of Toronto, for detail description of data. Any remaining errors are my own.

1. Introduction

Mankiw and Zeldes (1991) report that in 1984, only 27.6% of households held any stock, and only 12% held more than $10,000 in stock in US. Bertaut and Starr-McCluer (2000) also find the limited participation of American households in the following years: 31.6%, 40.4 %, and 48.9% in 1989, 1995, and 1998 respectively. More generally, Guiso et al. (2002) document the proportion of direct and indirect stockholders in other countries during the 1990s, revealing even less participation in Netherlands, German and Italy like 11.5%, 10.5% and 4.0%. Why do many people appear reluctant to allocate money to stock with high mean return?

Understanding the underlying determinants of household portfolio choice has several reasons. First, as discussed by Constantinides et al. (2002), Vissing-Jorgensen (1999), Heaton and Lucas (1996), and Mankiw and Zeldes (1991), participation rate sheds light on equity premium puzzle by relaxing complete participation assumption in consumption CAPM[1]. If the majority of households do not own equity, the premium may be considerably higher than in an economy with full participation. Second, following the analysis of Abel (2001) and Campbell et al. (1999), exploring what ultimately deter a number of households from holding shares offers theoretical basis for investing social security funds in the stock market. In a full-information and frictionless environment, it is not necessary for government to invest behalf of citizens, since each one can construct her portfolio by herself.

The low stockholding rate, particularly when coupled with high equity premium in both short and long horizon, has stimulated a large volume of research. Risk aversion, the first factor economists resort to, first fail to explain it. Mehra and Prescott (1985) require the coefficient of relative risk aversion to be pushed as large 30 as consistent with tremendous equity premium. However, this number can only be near 1 in micro evidence[2]. Although Epstein and Zin (1989) using a utility function break the linkage between the risk aversion coefficient and elasticity of intertemporal substitution, which are reciprocal for each other in Mehra-Prescott model, their result can only explain about one third of the equity premium[3].

The second approach invokes entry cost of investing in the stock market[4]. Direct monetary costs

[1] However, Polkovnichenko (2004) argues that even limited participation is counted in, the model cannot deliver sufficiently high equity premium.
[2] Kandel and Stambaugh (1991) believe that even the values as high as 30 imply quite reasonable behavior when the bet involves a maximal potential loss of around 1% of the gambler's wealth.
[3] Vissing-Jorgenson and Attanasio (2003) also follow this line.
[4] Actually, entry cost is only one sort of transaction cost approach. See Vissing-Jorgensen (2000)

such as taxes, mutual fund fees and brokerage commissions and indirect opportunity cost like spending time understanding and learning investment strategy all reduce the net return to holding equity. Vissing-Jorgensen (2000) builds a model in which, fixed costs of participation are incurred per period, and estimates that cost about $200 per year to explain observed participation rate. Alan (2005) using a structural estimation procedure, estimate a one-time entry cost—approximately 2% of annual permanent income. Nevertheless, this approach suffers some challenge from the little discrepancy between emerging market and developed market. For instance, in 2003, the proportion of household directly holding stocks in China have reached approximately 15%[5], which almost comes up to the rate in European countries. Chinese stock market, established in 1990, appears more frictional as an emerging market, and households are less familiar with financial notion than European countries. Whichever kind of entry cost cannot give convincible reasons.

The third approach relies on the heterogeneity in individual psychology, and this is the one I focus on here. Behavioral finance thoughts: Prospect theory and ambiguity aversion have been applied to solve this participation puzzle[6]. My issue, although different from the three ways above, comes from another crucial behavioral discussion on the overconfidence phenomena in financial market. I propose that stock market participation is influenced by households' underconfidence. The underconfident people underestimate their ability of absorbing information from stock market, thus drop the real gain opportunity, namely do not hold stocks, even if they are remunerative. From this perspective, high participation rate in China can be attributed to less underconfidence than European countries. Actually, in psychology, a great many of studies have revealed cross-cultural variations in overconfidence, whereby respondents in some Asian cultures (e.g. Chinese) exhibit markedly higher degrees of overconfidence than respondents in other cultures (e.g. in the US and Europe)[7]. Yates et al. (1998) generalize this phenomenon into more common practical decisions and the result shows cross-cultural variances do indeed exist.

Griffin and Tversky (1992) find the determinants of confidence, and demonstrate the over- and underconfidence is not an artifact of item selection or a byproduct of test difficulty. They build the studies on the hypothesis that people focus on the strength or extremeness of the available evidence

[5] This proportion is estimated by account number in Chinese stock market and households number in China. Account number is reported by China Security Regulatory Commission July 2003.
[6] See Benartzi and Thaler (1995); Barberis, Huang and Santos (2001); Epstein and Wang (1994);
[7] See Lee et al. (1995); Whitcomb et al. (1995); Yates et al. (1989), and Wright et al. (1978).

(e.g. the size of an effect) with in sufficient regard for its weight or credence (e.g. the size of the sample), and claim this judgment mode yields underconfidence when strength is low and weight is high. When a people facing thousands of stocks to choose, the performance of each stock in screen can give her full weight but low strength unless an authority in investment could give her some recommendation. It is entirely possible for her to be underconfidence. Bjokman et al. (1993) document a very pervasive underconfidence bias in the area of sensory discrimination. Using subjective distance theory of confidence based on the law of comparative judgment, they explain that people should perform better than they express in their confidence assessments and this bias is practically inevitable. Koriat et al. (2002) construct a paired cycles study, which display a slight overconfidence on the first cycle, and robust underconfidence from the second cycle. After several experimental manipulations, this underconfidence still appears. Considering the miscellaneous preparation before purchase stocks, people are likely to fall in the underconfidence cycle.

Under the circumstance with co-existence of over- and underconfidence, Odean (1998) give three reasons to support that overconfidence really happen in the stock market, one of which, named *selection bias*, indicates that only the people overestimating their ability to trade choose to participate into financial market. Thus, underconfidence, as inverse of overconfidence may be another factor to explain the limited participation.

In this paper, I begin with a simple theoretical model to inspect how underconfidence works based on risk aversion and entry cost effect as well as the relation between them, and then test this idea using the data from Survey of Consumer Finance (1984). Matching my story, the result is consistent with my argument that there are higher participation rates among non-underconfident people than among underconfident under the control of other factors. The rest of the paper is structured as follows. The next section unravels how underconfidence prohibits more people from stock. Section 3 describes my data set and empirical test model. In Section 4, I present result and corresponding analysis. Section 5 concludes.

2. A Simple Theoretical Model[8]

In an economy, each household initially consists of half a dollar in cash, and is considering the

[8] The spirit of this model comes from Daniel et al. (1998)

possibility to invest that money in a representative stock portfolio. All portfolios return one or zero with equal probabilities. The value of stock ends up with cash flow by $\tilde{v}$, and the risk of this portfolio is completely idiosyncratic[9]. I assume the discount rate and risk-free rate is zero for simplicity. Therefore, the expectation of return for holding stock equals to zero in first period.

The potential value from a representative portfolio comes from the possibility of acquiring information on it. The households can gather information signal about the stock's payoff, and make their decision whether to participate into stock market. The imperfect signal is like

$$\tilde{s} = \varepsilon\tilde{v} + (1-\varepsilon)\tilde{u}$$

where $\tilde{u}$ has the identical independent distribution with $\tilde{v}$, and ε takes a value of one with probability $a \in (0,1)$ or zero otherwise. This signal $\tilde{s}$ appears more informative for larger value of a, as the true value of the portfolio is then observed more often. The parameter a can be interpreted as the ability of individual making the investment decision.

The sequence of event is illustrated in Figure 1. The households have three options during the period 1, to choose whether to participate, do not, or gather some information. If information is gathered, the households make their second-stage decision: invest or do not again. Initially, the utility of each household from investment is assumed equal to the value received in the end, and tally with Von-Neumann and Morgenstern utility function. Given that the risk of the representative stock available to the firm is idiosyncratic and risk-free rate is zero, we know that the choice of households should be the one with maximal expectation value.

Lemma 1 *Suppose that each household is risk neutral and have no entry cost, all the households first gather information and half of them do participate.*

However, when we add risk aversion, entry cost and underconfidence into this model, it shows some different pictures

Case 1. Risk Aversion

To keep the analysis tractable, I model risk aversion as a utility cost $r \geq 0$ that is incurred by the

[9] Previous research has found that for most households, stock market risk has a correlation close to zero with other important risks, such as labor income risk, proprietary income risk and house price risk. (see Heaton and Lucas, 2000)

value of stock end up with nothing[10]. This cost effectively makes households risk aversion: the three original potential outcomes of stock participation decision, $\{0,\frac{1}{2},1\}$, will respectively yield $\{-r,\frac{1}{2},1\}$ in utility to the households, inflecting their utility function convex to the end-of-period value. Suppose the utility cost r of households is distributed over $[0,\infty)$. Note that, with this three-outcome specification, assuming more traditional utility functions will not change any of my result.

Lemma 2 *Besides the households receiving* $\tilde{s}=0$, *the ones whose* $r>\frac{a}{1-a}$ *will evade stock market.*

Therefore, no matter which signal is observed, the households with $r>\frac{a}{1-a}$ will not participate into stock markets all the time, and collecting information or not is the same for them

Case 2. Entry Cost

Hold all the conditions unchanged in Case 1, entry cost joins. Following Alan (2005), I treat entry cost as a one time fee, that muse be paid for first participation. Entry cost, as she states, is begot by acquiring information and is linked to the opportunity cost of time like wage. Here, every household should undertake $e>0$ entry cost for gleaning information, which makes non-difference for non-participation people in virtue of risk aversion in Case 1. Actually, this entry cost is not mandatory for purchasing stocks. Households can invest money into stock direct like gambling, but they never will do that due to risk aversion. The expectation of direct investment is always less than saving money. Owing to heterogeneity of each household, e is distributed over $[0,\infty)$

Lemma 3 *Besides the households receiving* $\tilde{s}=0$ *or with* $r>\frac{a}{1-a}$, *the ones whose entry cost* $e>\frac{a-r(1-a)}{4}$ *will neither collect information nor participate stock market.*

Case 3. Underconfidence

Following the work of Daniel et al. (1998), Odean (1998), Gervais and Odean (2001) and Gervais

[10] This definition resembles loss aversion rather than risk aversion, but does not matter my visualization. see Tversky and Kahneman (1992); Benartzi and Thaler (1995); Barberis, Huang and Santos (2001).

(2003), I treat the underconfidence to be the perception of private information if more precise and more reliable than it really is. I assume that the underconfidence household thinks that a is equal to $\tilde{a} \in (0, a]$, where the difference $a - \tilde{a} \in (0, a]$ measures the degree of underconfidence. Clearly, the underconfident households put too much weight on their information and, as a result, under-adjust their beliefs toward private information. They therefore think that the stock is worse (better) than it really is after receiving $\tilde{s} = 1$ ($\tilde{s} = 0$). Underconfidence, although can placate the disappointment when $\tilde{s} = 0$ is observed, still pale in comparison with quitting stock market.

Proposition *Suppose all the households are risk aversion, underconfident and need to pay entry cost. Besides the ones receiving* $\tilde{s} = 0$*, with* $r > \frac{a}{1-a}$*, or* $e > \frac{a - r(1-a)}{4}$*, the households who satisfy* $r \in \left[\frac{\tilde{a}}{1-\tilde{a}}, \frac{a}{1-a}\right]$ *or* $e \in \left[\frac{\tilde{a} - r(1-\tilde{a})}{4}, \frac{a - r(1-a)}{4}\right]$*, do not participate in stock market.*

The influence of underconfidence should be built on the risk aversion or entry cost, both of which produce a frictional circumstance for underconfidence to exaggerate frictions. In an ideal assumption proposed in the beginning of this section, underconfidence fails to change the participation rate eventually.

To sum up, synthesized with risk aversion and entry cost, underconfident people by means of underestimating the precision of their private information, drop the beneficial opportunity to invest into stock market. In next section, I identify whether to hold *mortgage, personal debt* or *personal debt except consumer debt* to surrogate underconfidence, to test my propositions above.

3. Empirical Work

3.1 the Model

The debt for individual in traditional economic theory is deemed as a method to smooth intertemporal utility in a lifecycle. Not much literature investigates other, perhaps even more ultimate reasons for getting in debt in personal level. A great deal of work in psychology have indicated those household heads who have outstanding credit, and who have higher amount of such debt, are significantly less likely to report complete psychological well-being, and behave increasing distress

with the augment of debt[11]. Brown et al. (2005) using the British Household Panel Survey strengthen the relation between personal debt and psychological well-being amongst heads of households. Their results highlight the psychological cost in debt. Livingstone and Lunt (1991) try to explain the underlying factors related to debt. Besides some attitudinal facors like pro-credit rather than anti-debt, they mention the confidence of handling future affaire as a psychological account. Golby (1990) even test overcommitted as a part of self-control to explain the level of debt. He concludes the most powerful explanation of the level of debt appears to be the degree of self-control (as captured by the overcommitted variable).

From intuition, the people who do not underestimate their ability are confidence for the future cash flow, and more easily incline to take advantage of debt to smooth their consumption or overcome transitory shock than underconfident ones. The debt in personal level can be classified as three main categories: mortgage, consumer debt and other personal debts. Considering the consumer debt convolutes us more attentions to the lifecycle hypothesis, and do not clearly represent underconfidence, I use *mortgage*, *personal debt* (sum of consumer debt and others) and *personal debt except consumer debt* (*PD except CD* henceforth) as the proxies separately for my underconfidence. My model has two types of investors: (1) "underconfidence" (the indicator is equal to 1), when the household do *not* have mortgage, personal debt or PD except CD; (2) "non-underconfidence" (the indicator is equal to 0) otherwise.

Since there are no direct questionnaires designed to measure risk aversion, it is very difficult to choose proxies to indicate the degree of risk aversion. Here, I try to use the *marital status* to proxy risk aversion. Married individuals may be more risk aversion because they have more than just themselves to provide for[12]. Apropos of entry cost, there also exist difficulty to find appropriate proxy. Several series of papers try to solve the participation puzzle through entry cost approach, and so far do not produce a consensus conclusion. A generally accepted issue ascribes entry cost to the familiarity both in psychology and financial knowledge. The entry cost, working as a participation block, depends on the unfamiliarity with financial market. In my model, the *area* dummy variable is taken as the proxy of entry cost. Thinking of the area where households live includes a whole bunch of factors to influence

[11] See Marmot et al. (1997); Weich and Lewis (1998).
[12] Using the China data, Weinhold and Zak (2005) show a significant relation between marital status and risk aversion. In addition, they provide several other interesting proxies for risk aversion like whether to move to city, whether to engage in a self-financed training and family size. The former two data are unavailable. The latter I add into regression later on.

participation decision, for example culture, acquaintance with stock market and physical distance to exchange. As a matter of fact, education level also can do a good job in capturing entry cost, but it contains quite a few personal characteristics, such as ability. Howbeit, I count *education level* in my regression for control variable.

The other control variables are *age*, *age square*, *wealth*, *income*, *ratio of cash to net wealth*, and *head labor status*. To scrutinize the effect of explanatory variables listed above, I plug them into regression by several groups. It is important to stress that since the survey defines the heads' sex as male for households in marriage automatically, the sex is virtually highly correlated with marital status. Therefore, I delete this popular explanatory variable out from my regression.

3.2 The Data

My data comes from Survey of Consumer Finances 1984 (SCF), *Microdata File on income (1983), assets and debts (1984) of economic families and unattached individuals*, administered by Household Surveys Division, Statistics Canada. The sample for the SCF conducted in May, 1984 consisted of approximately 17,100 dwellings selected from the Labour Force Survey sampling frame.[13] Briefly, this sample, designed, represents approximately 98% of the populations, excluding groups—residents of the Yukon, Northwest Territories, Indian reserves, and military barrocks and inmates of institutions. Interviewers listed all persons in the selected households and asked person 15 years of age and over questions concerning their labour activity during the previous week and work experience during the previous year. The interviewers then asked these persons for detailed income information from for the calendar year 1983 and their asset and debt position and pension plan participation of family members at the time of the survey. During the survey period, data were collected by personal interview with at least one visit to the household.[14] The overall family unit response for income data was 78.4%. Since a multi-stage stratified clustered sample design was employed, the revised weight based on updated population counts from the 1991 census should be introduced in statistic description. However, thinking that My study concentrate on the role of underconfidence played in households' stock market participation decision, all the regression analysis do not adopt weight.

To create my measure of underconfidence, I focus on three fields on households' debt. My

[13] A detailed description of this survey design, which is a multi-stage stratified clustered probability sample, can be found in Methodology of the Canadian Labour Force Survey, 1076, Catalogue No. 71-526.
[14] The questionnaires used in this survey are available for request.

mortgage dummy variable takes on the value of one for households not holding mortgage, and zero for remainder [15]. My *personal debt* dummy variable is set to one for households without any personal debt, including consumer debt and multifarious loans, and zero for the rest. Personal debt constitutes the majority of total debt for household, and if it is none, can indicate some underconfidence for this household. My *PD except CD* dummy variable is one for households who have no personal debt without consumer debt, and zero for the remainder.

Marital Status dummy is defined one if the head is married to represent more risk aversion, and zero otherwise. *Area* dummy, created to indicate entry cost, is identified one for urban centers with a more than 30, 000 population, and zero for otherwise rural place. For age and education level, I may have two responses per household—one for the man and one for the woman of the house. I take the "*age*" and "*education level*" of a household to be the higher of the two values[16]. The *education level* dummy is one for educated more than 13 years, and zero for less.

Heads labor status dummy is taken one for unemployed heads, and zero otherwise. Theodossiou (1998) using the 1992 British Household Panel Study (BHPS), claims unemployed individuals are found to suffer significantly high odds of experiencing a marked rise in anxiety, depression and loss of confidence and a reduction in self-esteem after controlling for a number of individual characteristics while six different measures of mental distress are used. Therefore, this dummy variable plays a very important role to control the underconfidence derived from specific individual transitory awful conditions.

Table 1 and 2 provide an overview of some basic facts about the data. Panel A of Table 1 breaks down participation rates across different demographic groups. Overall, in the whole SFC sample, 13.26% of households participate in the stock market. Participation increases sharply with wealth, going from 3.02% in the lowest quartile of the wealth distribution to 28.61% in the highest wealth quartile. There are also strong differences between high education households and low groups, as well as urban and rural households. These differences remain stark even controlling for wealth. For example, in the fourth quartile of the wealth distribution, the participation rate is 38.66% for more than 13 years educated households, and only 19.90% for less educated ones. Likewise, in the same quartile, the

[15] To purify this underconfidence indicator, If only part of the property containing the residence is used as their residence and part is used for other purpose (i.e. for rent or business purpose), only the mortgage outstanding on the proportion use as family's residence is counted in, because catchall mortgage may be contaminated by business risk.

[16] None of my results is sensitive to how I choose to handle these details.

participation rate is 32.78% in comparison with 21.32% for urban households to rural ones.

In Panel B of Table 1, I take a crude first look at the effect of underconfidence indicators on participation, with a two-way sort based on wealth and each of my three measures of underconfidence. Although this approach is obviously not a substitute for the more carefully controlled regressions that follow, it makes it clear that the basic patterns emerge in even the simplest tabulations of the data. Using the third quartile of wealth distribution, the participation rate of households with *mortgage* is 16.64%, while 14.48% of ones without mortgage do. In the same quartile, with *personal debt*, the corresponding figures are 15.49% and 9.61%, while with *PD except CD*, they are 18.72% and 12.43%.

In Table 2, I look at the correlation between our three measures of underconfidence, as well as between these measures and the other independent variables that will enter my specifications. No surprisingly, *personal debt* and *PD except CD* are highly correlated, with a correlation coefficient of 0.33, and *mortgage* is correlated with *personal debt*. On the other hand, *mortgage* and *PD except CD* are weaklier correlated with coefficient of only 0.027. Thus, it seems that there is some independent information on underconfidence between them.

No astonishingly, *marital status* is highly correlated with *mortgage*, because, married people are more prone to own their house[17]. *Income after tax* seems be correlated with quite a few independent variables. Hence, I add it into regression finally. None of the underconfidence measures is all that highly correlated with the other proxies, though there are some noteworthy differences.

4. Empirical Results

Table 3 presents my results. All regression are run by OLS, with the dependent variable an indicator that takes on the value one when a household owns stock including direct and indirect holding, and zero otherwise[18]. There are nine columns, corresponding to three different specifications with each of our three underconfidence measures—*mortgage*, *personal debt* and *PD except CD*. In column (1), (4) and (7), the underconfidence variable enters along with an *area* dummy, a *marital status* dummy, an *education level* dummy, *age*, *age square* and three dummies corresponding to the second third and fourth quartiles of the wealth distribution.

[17] This correlation actually does not mean underconfidence and risk aversion are dependent with reference to our definition. Here the confusion results from proxies choosing.
[18] Given the dichotomous nature of left-hand-side variable, I have redone all the tests using probit and logit specification, with very similar result.

In column (2), (5) and (8), the only second modifications are: first, I use 9, rather than three wealth dummies, which means that I have now chopped up the wealth distribution into 10-percent increment, and these allow me to get a tighter control for wealth and education, but make it impractical display the individual coefficients on all these dummies; Second, I add 9 *income* dummies, the *ratio of cash to net wealth* to capture liquid need, a *head labor status* dummy. Finally, in column (3), (6) and (9), the two interactions between underconfidence indicator and area and marital status respectively are counted in. Recall my theoretical claim, underconfidence prevents participation only through amplifying the effect of risk aversion and entry cost, and it does not work when risk aversion and entry cost are valued zero. So the ideal results we want to see are: when I put two interactions into regression, the coefficient of underconfidence becomes no prominent and interaction coefficients display significant instead.

The results offer a consistent picture. Take the first look at *mortgage* measure of underconfidence. The negative coefficients in all three regressions imply that underconfident households less probably participate stock market. However, with more control variables added into regression, the coefficients are less and less statistically significant. Especially, when *income* appears in the regression. That means whether a household owns house by mortgage is mainly influenced with their income. However, the other underconfidence indicators do well. Both of their negative coefficients are statistically significant at the 1% level.

As anticipated, the negative coefficient of the *marital status* dummy is prominent in 1% significant level after more control variables coming in. That shows the more risk averse, the less participation. The *area* dummy also works in explaining the participation rate. The positive coefficient evinces less entry cost will increase the participation possibility prominently. Moreover, almost all the null hypotheses that coefficient is equal to zero are rejected in a 1% significant level. Furthermore, the performance of interaction variables should be paid an attention. Although the *mortgage* and *personal debt* do not follow the way I expect, when interactions enter, the *PD except CD* performance provides a favorable support. In the column (9), the interactions replace the significance of underconfidence indicator. That is consistent with my theoretical claim.

The coefficients on some of the other controls are worth a brief mention. To begin with, I confirm earlier work by finding very powerful effects of education and wealth on stock market participation. For instance, the estimates suggest that more than 13 years educated households have about a 6% greater participation rate than other groups, all else be equal. The effect of age always appears negative,

while age square influences positive. This maybe implies a convex u-shape function of age to participation. In addition, the head labor status, in spite of non-prominence, consists with both Theodossiou (1998) and intuition that unemployed heads are less willing to participate capriccioso stock market.

Beyond what is displayed in Table 3, I have also experimented with a number of variations of my specification, in order to further check the robustness of my result. First, I have redone everything with alternative definitions of stock market participation that require a household to have some minimum level invested in the market (I have tried thresholds of $100, $500, and $1000[19]) as opposed to just anything over zero, in order to be counted as a participation. This procedure helps me to discriminate small stakeholder and considerable investment, which could be more correlated with underconfidence. As it turns out, these alternative definitions of participation lead to results that are very similar to those in Table 3. Hence, these results do not appear to be driven by investors with only trivial stakes in the market.

Second, I have also tried to add several further controls, beyond those reported in columns (3), (6) and (9) of Table 3, to my specifications. These controls include: *family size*[20], *number of persons with income*, and *heads immigration status*[21]. These control variables can be motivated as attempts to capture some earning and consumption ability, which perhaps influence our result like what income dummies do. However, I find that none of the additional controls has any noticeable impact on the coefficients of the underconfidence indicators. In addition, *lifecycle* dummies also are tried to control some endogenous motivations for holding debt, and the results still are very similar.

5. Conclusions

This paper adds underconfidence after risk aversion and entry cost, in order to provide another approach to explain the limited stock market participation puzzle. I first review some psychology and behavioral finance literature to enlighten this road, and then construct a simple model to show underconfidence is nothing but a type of subjective underestimation on ones' ability. Under a perfect

[19] Among those households who participate in the market, the median amount invested is $3000. the 25th percentile of the distribution is $600, and the 75th percentile is $10,000

[20] By inspecting the Survey of Consumer Finance data from 1990-1995, Crook (2001) find besides age, relatively risk aversion, whether to own their home, and whether head is working, when the family size is larger, the household demand more debt.

[21] Aydemir and Skuterud (2005) concentrate on the immigrant wage differentials, and find the wage gaps may be entirely independent of the actual or perceived skills or quality of immigrants themselves.

complete market, where there is no risk aversion and entry cost, underconfidence cannot influence households' participation decision. Using the data from Survey of Consumer Finance 1983, I employ whether to hold mortgage, personal debt and personal debt except consumer debt, three indicators as proxy for underconfidence, and find the negative relation between participation rate and underconfidence, after controlling other personal traits. This paper also supports the claim as behavior finance scholars insist on, that there really exist overconfidence in financial market, because the underconfident people are censored out in participation process.

While the SCF data allow me to address my theory in a straightforward way, they also suffer from several important drawbacks. First of all, the measure of underconfidence that I take from the SCF—whether households have mortgage, personal debt and PD except CD—reflect endogenous choices, and hence may capture information not just about the degree of underconfidence per se, but about other personality traits that may be associated with the propensity to invest in the stock market. For example, pessimistic people are afraid to be involved in debt, and they also have lower expectations for future stock market returns. Second, the difference between debtor and nondebtors, to some extent, can be attributed to their subjective estimation on earning ability, but there is no guarantee to show that they over- or underestimate their stock trading ability in the same way. The paradox is that only their performance in stock market can be the perfect evidence for over- or underconfidence, whereas underconfident people never participate. Third, although the stock market performance data for all the households are unavailable, I, at least, obtain some panel data of the same object households. Through analyzing the historical behavior—like whether to reducing risky assets, which earning return last year, there may be some more convincible proxies for underconfidence.

Appendix

Proof of Lemma 1

$$\Pr\left\{\tilde{s}=1\right\}=\Pr\left\{\tilde{v}=1,\tilde{s}=1\right\}+\Pr\left\{\tilde{v}=0,\tilde{s}=1\right\}$$

$$=\Pr\left\{\tilde{s}=1\,|\,\tilde{v}=1\right\}\Pr\left\{\tilde{v}=1\right\}+\Pr\left\{\tilde{s}=1\,|\,\tilde{v}=0\right\}\Pr\left\{\tilde{v}=0\right\}$$

$$=\left[a+(1-a)\times\tfrac{1}{2}\right]\times\tfrac{1}{2}+(1-a)\times\tfrac{1}{2}\times\tfrac{1}{2}=\tfrac{1}{2}$$

$$\Pr\left\{\tilde{v}=1\,|\,\tilde{s}=1\right\}=\frac{\Pr\left\{\tilde{v}=1,\tilde{s}=1\right\}}{\Pr\left\{\tilde{s}=1\right\}}=\frac{\left[a+(1-a)\times\frac{1}{2}\right]\times\frac{1}{2}}{\frac{1}{2}}=\frac{1+a}{2}>\frac{1}{2}$$

$$\Pr\left\{\tilde{v}=0\,|\,\tilde{s}=1\right\}=\frac{\Pr\left\{\tilde{v}=0,\tilde{s}=1\right\}}{\Pr\left\{\tilde{s}=1\right\}}=\frac{(1-a)\times\frac{1}{2}\times\frac{1}{2}}{\frac{1}{2}}=\frac{1-a}{2}<\frac{1}{2}$$

If household gather information and purchase stock when $\tilde{s}=1$, do not when $\tilde{s}=0$

$$\text{Expectation}=\Pr\left\{\tilde{s}=1\right\}\left[\Pr\left\{\tilde{v}=1\,|\,\tilde{s}=1\right\}(1)+\Pr\left\{\tilde{v}=0\,|\,\tilde{s}=1\right\}(0)\right]+\Pr\left\{\tilde{s}=0\right\}\frac{1}{2}$$

$$=\frac{1}{2}\times\left[\frac{1+a}{2}\times1+0\right]+\frac{1}{2}\times\frac{1}{2}=\frac{1}{2}+\frac{a}{4}>\frac{1}{2}$$

Hence, all the households are willing to gather information in the first step, and only ones who observe $\tilde{s}=1$ with one half probability, participate into stock market. ***QED.***

Proof of Lemma 2

For the households whose $r>\dfrac{a}{1-a}$, when they observe $\tilde{s}=1$

$$\text{Expectation}=\Pr\left\{\tilde{s}=1\right\}\left[\Pr\left\{\tilde{v}=1\,|\,\tilde{s}=1\right\}(1)+\Pr\left\{\tilde{v}=0\,|\,\tilde{s}=1\right\}(-r)\right]+\Pr\left\{\tilde{s}=0\right\}\frac{1}{2}$$

$$=\frac{1}{2}\times\left[\frac{1+a}{2}\times1+\frac{1-a}{2}\times(-r)\right]+\frac{1}{2}\times\frac{1}{2}=\frac{1}{2}+\frac{a-r(1-a)}{4}<\frac{1}{2}.$$ ***QED.***

Proof of Lemma 3

For the households whose $e>\dfrac{a-r(1-a)}{4}$, $r<\dfrac{a}{1-a}$, when they observe $\tilde{s}=1$

$$\text{Expectation}=\Pr\left\{\tilde{s}=1\right\}\left[\Pr\left\{\tilde{v}=1\,|\,\tilde{s}=1\right\}(1)+\Pr\left\{\tilde{v}=0\,|\,\tilde{s}=1\right\}(-r)\right]+\Pr\left\{\tilde{s}=0\right\}\frac{1}{2}-e$$

$$=\frac{1}{2}\times\left[\frac{1+a}{2}\times 1+\frac{1-a}{2}\times(-r)\right]+\frac{1}{2}\times\frac{1}{2}-e=\frac{1}{2}+\frac{a-r(1-a)}{4}-e<\frac{1}{2}.$$ ***QED.***

Proof of Proposition 1

Underconfident households make their decision based on $\tilde{a}$ not a. Substituting all the a with $\tilde{a}$ in the threshold in Case 1 and Case 2, only the ones with $r<\frac{\tilde{a}}{1-\tilde{a}}$, $e<\frac{\tilde{a}-r(1-\tilde{a})}{4}$ and gathering $\tilde{s}=1$ simultaneously dare to participate in stock market. ***QED.***

Reference:

Alan, S, 2005, "Entry Costs and Stock Market Participation Over the Life Cycle", Working Paper", York University

Abel, A., 2001, "The effects of investing social security funds in the stock market when fixed costs prevent some households from holding stocks", *American Economic Review, vol*91, 128-148.

Aydemir, A., Skuterud, M., 2001, "The immigrant Wage Differential Within and Across Establishments", Working Paper, University of Waterloo.

Barberis, N., and M. Huang, 2001, "Mental accounting, loss aversion and individual stock returns", *Journal of Finance, vol*56: 1247-1292.

Benartzi, S., and R. Thaler, 1995, "Myopic loss aversion and the equity premium puzzle", *Quarterly Journal of Economics, vol*110: 75-92.

Bjokman, M., Juslin, P., and Winman, A., 1993, "Realism of confidence in sensory discreimination: The under confidence phenomenon", *Perception & Psychologysics, vol*54: 75-81.

Bertaut, C., and Starr-McCluer, M.,, 2000, "Household portfolios in the United States", Finance and Economics Discussion Series 2000-26, Board of Governors of the Federal Reserve System.

Campbell, J., Cocco, J.,Gomes, F., and Maenhout, P., 1999, "Investing retirement wealth: A life-cycle model", Working paper, Harvard University.

Cohn, R., Lewellen, W., Lease, R., and Schlarbaum, 1975, "Individual Investor, Risk Aversion and Investment Portfolio Composition", *Journal of Finance, vol*39: 605-620

Golby, S., 1990, "An Economic Analysis of Personal Debt", Bulletin of Economic Research, vol42: 241-247.

Constantinides, G., Donaldson J., and Mehra, R., 2002, "Junior Can't Borrow: A New Perspective on the Equity Premium Puzzle", *Quarterly Journal of Economics, vol*117: 269-296.

Crook, J., 2001, "The demand for household debt in the USA: evidence from the 1995 Survey of Consumer Finance", *Applied Financial Economics, vol*11: 83-91.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A., 1998, "A Theory of Overconfidence, Self-Attribution, and Security Market Under- and Over-reactions," *Journal of Finance, vol*53: 1839-1885.

Epstein, L., and Zin, S., 1989, "Substitution, Risk aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework", *Econometrics, vol*57: 937-969

Epstein, L., and Wang, T., 1994, "Intertemporal asset pricing under Knightian uncertainty", *Econometrica, vol*62: 283-322.

Friend, I., and Blume, M., 1975, "The demand for risky assets", *American Economic Review, vol*65: 900-921.

Gervais, S., and Odean, T., 2001, "Learning To Be Overconfident," Review of Financial Studies, vol14: 1-27.

Gervais, S., Heaton, J. and Odean, T., 2003, "Overconfidence, Investment Policy, and Executive Stock Options", Working paper, Duke University.

Guiso, L., Haliassos, M., and Jappelli, T., 2002, "Household Portfolios", (Cambridge, MA: MIT Press)

Griffin, D., and Tversky, A., 1992, "The Weighing of Ecidence and the Determinants of Confidence", *Cognitive Psychology, vol*24: 411-435.

Heaton, J., and Lucas, D., 1996, "Evaluating the Effect of Incomplete Markets on Risk Sharing and Asset Pricing", *Journal of Political Economy, vol*104: 443-487.

Heaton, J., and Lucas, D., 2000, "Portfolio Choice in the Presence of Background Risk", *Economic Journal, vol*110: 1-26.

Livingstone, S., and Lunt, P., 1992, "Predicting personal debt and debt repayment: Psychological, social and economic determionants", *Journal of Economic Psychology, vol*13: 111-134

Kandel, S., and Stambaugh, R. F., 1991, "Asset Returns and Intertemporal Preferences", *Journal of Monetary Economics, vol*27: 39-71.

Koriat, A., Sheffer, L., and Ma'ayan, H., 2002, "Comparing Objective and Subjective Learning Curves: Judgments of Learning Exhibit Increased Underconfidence With Practice", *Journal of Experimental Psychology, vol*131: 147-162.

Lee, J., Yates, J., Shinotsuka, H., Singh, R., Onglatco, M., Yen, N., Gupta, M., and Bhatnagar, D., 1995, "Cross-national differences in overconfidence", *Asian Journal of Psychology, vol*1: 63-69.

Mankiw, N., and Zeldes, S., 1991, "The Consumption of Stockholders and Non-stockholders", *Journal of financial Economics, vol*29: 97-112.

Mehra, R., and Prescott, E., 1985, "The equity premium: A puzzle", *Journal of Monetary Economics, vol*15: 145-161.

Odean, T., 1998, "Volume, Volatility, Price, and Profit When All Traders Are above Average", *Journal of Finance, vol*53: 1887-1934.

Polkovnichenko, V., 2004, "Limited Stock Market Participation and the Equity Premium", *Finance Research Letters, vol*1: 24-34

Siegel, F. W., and Hoban J., 1982, "Relative risk aversion revisited", *Review of Economics and Statistics, vol*64: 481-487.

Theodossiou, I., 1998, "The effects of low-pay and unemployment on psychological well-being: a logistic regression approach", *Journal of Health Economics, vol*17: 85-104

Vissing-Jorgensen, A., 1999, "Limited stock market participation and the equity premium puzzle", Working paper, University of Chicago

Vissing-Jorgensen, A., 2000, "Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures", Working paper, University ofChicago.

Vissing-Jorgensen, A. and Attanasio, O., 2003, "Stock-Market Participation, Intertemporal Substitution, and Risk-Aversion", *American Economic Review*, vol93: 383-391.

Weinhold, D., and Zak, P. J., 2005, "The Choice of Institutions: The Role of Risk and Risk-Aversion", Working Paper. London School of Economics.

Whitcomb, K., Onkal, D., Curley, S.,and Benson, P., 1995, "Probability judgment accuracy for general knowledge: Cross-national differences and assessment methods", *Journal of Behavioral Decision Making, vol*8: 51-67.

Wright, G., Phillips, L., Whalley, P., Choo, G., Ng, K., Tan., and Wisudha, A., 1978, "Cultural differences in probabilistic thinking", *Journal of Cross-Cultural Psychology, vol*9: 285-299.

Yates, J., Zhu, Y., Ronis, D., Wang, D., Shinotsuka, H., and Toda, M., 1989, "Probability judgment accuracy: China, Japan, and the United States", *Organizational Behavior and Human Decision Processes, vol*43: 145-171.

Yates, J., Lee, J., Shinotsuka, H., Patalano, A., and Sieck, W., 1998, "Cross-Cultural Variations in Probability Judgment Accuracy: Beyond General Knowledge Overconfidence?", *Organizational Behavior and Human Decision Processes, vol*74: 89-117

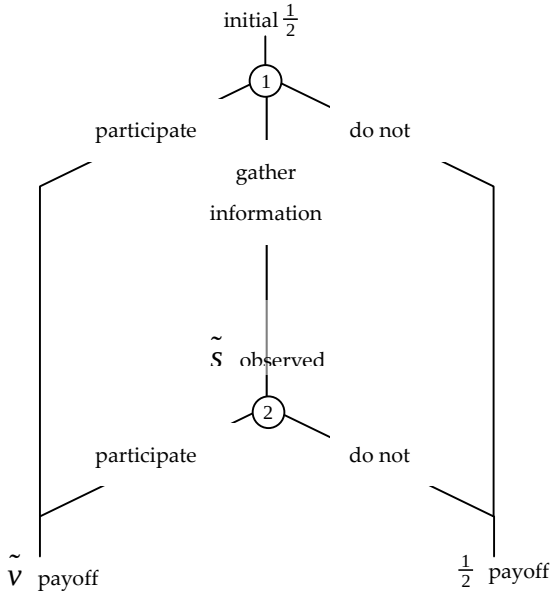

Figure. 1. Sequence of Choice.

The open circles denote stage at which the households must make decision. At each of these stages, the individual must decide whether to participate stock market, do not (in the first stages, they have one more choice to gather more information). At the end of the period, the households get their payoff either from the representative stock ($\tilde{v}$) if participation is chosen, or from the safe holding ($\frac{1}{2}$)

Table 1

Stock-Market Participation Rates for Different Categories of Households

Table 1 gives the stock market participation rates for different categories of households in the Survey of Consumer Finances 1984. Panel A reports these rates by Education level, area and wealth of households. Panel B reports these rates by three underconfidence indicators (Whether to hold mortgage, personal debt, or personal debt other than consumer debt) and wealth. There are 14,029 household observations

Panel A: Stock-Market Participation Rates by Education Level, Risk Aversion, Area and Wealth

		Education Level		Area	
	All	⩽13 yrs	>13 yrs	Rural	Urban
	(1)	(2)	(3)	(6)	(7)
All Household	13.26%	9.05%	19.94%	10.15%	15.20%
1th Quartile of Wealth Distribution	3.02%	2.06%	5.14%	3.35%	2.88%
2nd Quartile of Wealth Distribution	8.27%	6.26%	11.32%	5.04%	10.54%
3rd Quartile of Wealth Distribution	13.15%	10.15%	18.20%	10.21%	15.77%
4th Quartile of Wealth distribution	28.61%	19.90%	38.66%	21.32%	32.78%

Panel B: Stock-Market Participation Rates by Underconfidence Indicators and Wealth

		Mortgage		Personal Debt		PD except CD	
	All	No Mortgage	Mortgage	No Debt	Debt	No Debt	Debt
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All Household	13.26%	11.63%	17.43%	10.95%	14.79%	12.56%	17.35%
1th Quartile of Wealth Distribution	3.02%	2.97%	5.09%	1.27%	4.28%	2.42%	5.35%
2nd Quartile of Wealth Distribution	8.27%	8.98%	6.90%	5.91%	9.49%	7.45%	13.70%
3rd Quartile of Wealth Distribution	13.15%	10.48%	16.64%	9.61%	15.49%	12.43%	18.72%
4th Quartile of Wealth distribution	28.61%	27.89%	30.04%	25.56%	30.93%	27.13%	38.31%

Table 2

Mean and Correlations of Underconfidence Indicators and Other Independent Variables

Table 2 gives summary statistics on various characteristics of households in the Survey of Consumer Finances 1984. Mortgage equals one for the households who do not own any mortgage, and zero otherwise. Personal Debt equals one for the households who do not hold any personal debt, and zero otherwise. PD except CD equals one for the households who do not hold any personal debt without consumer debt, and zero otherwise. Additional household characteristics include indicator variable for risk aversion (marital status), for entry cost (area), and for ratio of liquid need (ratio of cash to net wealth). Other characteristics include an indicator for education level, head labor status income after tax and wealth. There are 14029 observations.

	Mortgage	Personal Debt	PD except CD	Area	Marital Status	Education Level	Ratio of Cash to Net Wealth	Head Labor	Income after Tax	Wealth
Mean of sample	0.7166	0.4048	0.8580	0.5674	0.3723	0.6509	0.0319	0.0652	24335.14	87057.91
Correlations										
Mortgage	1.0000									
Personal Debt	0.2210	1.0000								
PD except CD	0.0270	0.3331	1.0000							
Area	-0.0078	-0.0347	0.0023	1.0000						
Marital Status	-0.2902	-0.1664	-0.0118	-0.0870	1.0000					
Education Level	-0.1551	-0.1843	-0.1223	0.1337	0.0851	1.0000				
Ratio of Cash to Net Wealth	0.0390	0.0613	0.0291	0.0014	-0.0362	-0.0673	1.0000			
Head Labor	0.0389	-0.0135	-0.0367	-0.0052	-0.0360	0.0040	0.0110	1.0000		
Income after Tax	-0.2710	-0.2055	-0.0371	0.1105	0.2443	0.4066	-0.0542	-0.0922	1.0000	
Wealth	-0.0374	0.0248	0.0143	0.0041	0.0698	0.1639	-0.0287	-0.0726	0.4167	1.0000

Table 3
Effect of Underconfidence on Whether a Household Owns Stocks

The sample comprises households in the Survey of Consumer Finances 1984. The dependent variable is an indicator that the household owns stock. The independent variables are: an underconfidence indicator (either Mortgage, Personal debt or PD except CD), risk aversion indicator (marital status), entry cost indicator (area), education level, age, age square, ratio of cash to net wealth, head labor status, and two interactions. Column (1) (4) (7) includes dummies for wealth quartile. The dummies for wealth deciles, for income deciles, head labor status and liquid conditions are counted in column (2) (5) (8). In column (3) (6) (9), I add two interactions. Robust standard errors are in parentheses. There are 14,029 observations. (*, **, *** are Significant at the 10%, 5%, 1% level respectively.

	Mortgage			Personal Debt			PD except CD		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Underconfidence Indicator	-0.0186***	-0.0085	-0.0227	-0.0340***	-0.0222***	-0.0184	-0.0494***	-0.0447***	-0.0117
	(0.0067)	(0.0068)	(0.0169)	(0.0060)	(0.0062)	(0.0121)	(0.0078)	(0.0077)	(0.0170)
Area	0.0418***	0.0274***	0.0284***	0.0421***	0.0281***	0.0214**	0.0436***	0.0290***	0.0254***
	(0.0055)	(0.0056)	(0.0066)	(0.0055)	(0.0056)	(0.0086)	(0.0055)	(0.0056)	(0.0060)
Marital Status	-0.0069	-0.0390***	-0.0362***	-0.0085	-0.0389***	-0.0364***	-0.0048	-0.0385***	-0.0421***
	(0.0063)	(0.0069)	0.0075	(0.0063)	(0.0068)	(0.0096)	(0.0062)	(0.0068)	(0.0072)
Age	-0.0009	-0.0031***	-0.0032***	-0.0007	-0.0033***	-0.0032***	-0.0013	-0.0030***	-0.0030***
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0012)	(0.0011)	(0.0011)	(0.0011)
Age Square	0.0000	0.0000***	0.0000***	0.0000	0.0000***	0.0000***	0.0000	0.0000***	0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Education Level	0.0887	0.0660***	0.0660***	0.0865***	0.0649***	0.0646***	0.0861***	0.0633***	0.0632***
	(0.0059)	(0.0060)	(0.0060)	(0.0059)	(0.0060)	(0.0060)	(0.0059)	(0.0060)	(0.0060)

Table 3—*Continued*

	Mortgage			Personal Debt			PD except CD		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2nd Quartile of Wealth	0.0413***			0.0446***			0.0500***		
Distribution	(0.0082)			(0.0080)			(0.0080)		
3rd Quartile of Wealth	0.0809***			0.0867***			0.0911***		
Distribution	(0.0089)			(0.0086)			(0.0086)		
4th Quartile of Wealth	0.2309***			0.2360***			0.2381***		
Distribution	(0.0093)			(0.0091)			(0.0091)		
Ratio of Cash to Net Wealth		-0.0008	-0.0008		0.0002	0.0003		0.0003	0.0001
		(0.0058)	(0.0058)		(0.0058)	(0.0058)		(0.0058)	(0.0058)
Head Labor Status		-0.0122	-0.0124		-0.0116	-0.0117		-0.0132	-0.0134
		(0.0111)	(0.0111)		(0.0111)	(0.0111)		(0.0111)	(0.0111)
Underconfidence indicator			-0.0028			0.0113			0.0257*
×Area			(0.0120)			(0.0110)			(0.0153)
Underconfidence indicator			-0.0153			-0.0040			-0.0272*
×Marital Status			(0.0165)			(0.0118)			(0.0164)
9 Wealth Dummies	NO	YES	YES	NO	YES	YES	NO	YES	YES
9 Income Dummies	NO	YES	YES	NO	YES	YES	NO	YES	YES
Adjusted-R Square	10.21%	13.20%	13.20%	10.36%	13.27%	13.27%	10.41%	13.40%	13.41%