University of Colorado
Denver

Business School

Campus Box 165
1475 Lawrence St
P.O. Box 173364
Denver, CO 80217-3364

business.ucdenver.edu

October 25th, 2023

Vanessa Countryman
Secretary
Securities and Exchange Commission
150 F St NE
Washington, DC 20549

Re: Regulation Best Execution (File No. S7-32-22) and Conflicts of Interest Associated with the Use of Predictive Data Analytics by Broker-Dealers and Investment Advisers (File No. S7-12-23)

Dear Ms. Countryman,

I and my co-authors, Professor Russell Cooper from the European University Institute and Dr. Mengli Sha from Bates White,[1] appreciate the opportunity to comment on the Securities and Exchange Commission's ("SEC") proposed rules regarding equity market structure (proposed Regulation Best Execution) and predictive data analytics, both of which reference dated studies by Barber and Odean addressing, among other things, harm caused to retail investors by active trading. As described in more detail below, we are concerned that the SEC's reliance on Barber and Odean's 2000 study, "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," ("Barber and Odean (2000)") may not be appropriate to support the above-referenced rule proposals for three primary reasons: (1) the validity of the results; (2) the relevance of the data used; and most importantly (3) the narrow nature of the study and our own research that presents counter results and reverses the key inference of Barber and Odean (2000) regarding the effects of active trading.

We start with the validity of the Barber and Odean (2000) results. While the study's sample includes around 78,000 brokerage accounts, it is not representative of the entire United States population due to the absence of sample weighting and because the sample is taken from only one retail brokerage, Charles Schwab. The sample weighting representation is critical, since most of the legitimate data sets that we use in our profession and research have sample weight, including the Survey of Consumer Finance by the Federal Reserve Board of Governors and Panel Studies Income Dynamics by the University of Michigan. Another issue with the validity of the data is the significant bias resulting from the data being taken ONLY from one brokerage headquartered at the time in San Francisco (now in Dallas). For example, it may oversample brokerage accounts in the West (42%) brokerage compared to the East (19%), the South (24%) and the Midwest (15%);

[1] Bonaparte, Y. Cooper, R. Sha, M. (2023). Rationalizing Trading Frequency and Returns: Maybe Trading is Good for You. Second Round at the *International Economic Review*

this is certainly not a reflection of the brokerage account national distribution. Certainly, the data oversamples the West and under samples the East.

The second big issue has to do with the relevance of the Barber and Odean (2000) data. Namely, the data sample was taken between 1991-1996, when the I internet was barely distributed with floor trading. Above all, this 27-year-old data reflects largely paper trading and not automated trade. Automated trading systems further empowered investors and ultimately led to no commission fees with possible smaller spreads. Most importantly, the profile of US investors has changed since 1991-1996. For example, based on the SCF that is most representative of household finance data in the US, if we compare SCF (1995) with the SCF (2022) we find that households are more educated (post high school) by 73.7%, more diverse by 13.3% (more minorities), and older by 8.2%. Basically, this is a different and more diverse and experienced America when it comes to investing.

Most importantly, the Barber and Odean (2000) study is very narrow, and only focuses on month-by-month trading without considering households' risk mitigation and consumption smoothing, which are the key factors when one makes stock trading. In our study titled "Rationalizing Trading Frequency and Returns: Maybe Trading is Good for You," we demonstrate that Barber and Odean (2000) find that households who trade more have a lower net return than others and attribute this pattern to irrationality, specifically overconfidence. In contrast, we find that household financial choices generated from a dynamic optimization problem with rational agents and portfolio adjustment costs can reproduce the observed distribution of (net) turnover rates as well as the fact that households with the highest turnover rates have the lowest net returns. Various forms of irrationality, modeled as beliefs about asset returns that are not data based, only marginally improve the ability of the baseline model to explain these turnover and net returns patterns. The potential for irrational choice appears in the decision to own stocks directly.

We caution the SEC about relying on the Barber and Odean (2000) inference due to the validity, relevance and narrow perspective of the study. Attached is our study, which is currently in the second round under review at an A* Journal, *International Economic Review*. We have another research paper that is an *NBER working paper 16022* titled "Rationalizing Trading Frequency and Returns," that also echoes the evidence of the paper.

Sincerely,

Yosef Bonaparte, Ph.D.
Associate Professor of Finance; Director for External Affairs
The University of Colorado at Denver
YOSEF.BONAPARTE@UCDENVER.EDU; Phone: +1 (303) 591-1878

Reference:

Barber, B., and T. Odean (2000): "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *The Journal of Finance*, 55(2), 773-806.

Bonaparte, Y. Cooper, R. (2010). Rationalizing Trading Frequency and Returns. *NBER Working Paper 16022*.

Bonaparte, Y. Cooper, R. Sha, M. (2023). Rationalizing Trading Frequency and Returns: Maybe Trading is Good for You. Second Round at the *International Economic Review*.

Rationalizing Trading Frequency and Returns*

Yosef Bonaparte† and Russell Cooper‡

November 17, 2010

Abstract

Barber and Odean (2000) study the relationship between trading frequency and returns. They find that households who trade more frequently have a lower net return than other households. But all households have about the same gross return. They argue that these results cannot emerge from a model with rational traders and instead attribute these findings to overconfidence. Using a dynamic optimization approach, we find that neither a model with rational agents facing adjustment costs nor various models of overconfidence fit these facts.

1 Motivation

Barber and Odean (2000) find that households who adjust their portfolio more frequently have a lower net return. They interpret this as evidence households are overconfident and thus not rational. According to Barber and Odean (2000):

> Our most dramatic empirical evidence supports the view that overconfidence leads to excessive trading ... On one hand, there is very little difference in the gross performance of households that trade frequently with monthly turnover in excess of 8.8 percent and those that trade infrequently. In contrast, households that trade frequently earn a net annualized geometric mean return of 11.4

*We are grateful to the National Science Foundation for supporting this research. We thank seminar participants at the Federal Reserve Banks of Dallas and Kansas City, Cambridge University, the CCER at Peking University, the EIEF in Rome, Melbourne University, Oxford University, the Pennsylvania State University and the Toulouse School of Economics for comments and suggestions on earlier work that underlies this paper. Guozhong Zhu provided outstanding research assistance.

†Robert Day School of Economics and Finance, Claremont McKenna College, 909-607-0756, ybonaparte@cmc.edu

‡Department of Economics, the European University Institute and Department of Economics, University of Texas at Austin, russellcoop@gmail.com

> percent, and those that trade infrequently earn 18.5 percent. These results are consistent with models where trading emanates from investor overconfidence, but are inconsistent with models where trading results from rational expectations.

This paper studies the implications of an optimizing model with costly portfolio adjustment for the relationship between frequency of trade and asset returns. We investigate two explanations for the findings of Barber and Odean (2000). The first looks the choice of rational agents faced with costs of portfolio adjustment. The second allows for overconfidence.

It seems natural to consider the differences in net returns as reflecting two forces: trading costs and a selection effect through household choice of whether to adjust their portfolio. Trading costs will drive a wedge between gross and net returns. Household choice, both on the extensive (to adjust or not) and intensive (turnover conditional on adjustment) margins, creates an endogenous relationship between asset returns and portfolio adjustment. Thus both ingredients are necessary to match the observations assuming agents are rational.

Building upon Bonaparte and Cooper (2009), we ask whether the presence of fixed and variable portfolio adjustment costs can generate the observed differences in returns based upon the frequency of trade. Our approach is to specify a dynamic optimization problem of a household and estimate its parameters. The uncertainty in the model comes from income shocks, which are partly household specific, as well as a stochastic return on the household portfolio. We generate simulated data from the estimated model to study the relationship between portfolio adjustment and returns.

Second, following the suggestion of Barber and Odean (2000), we study a series of models which relax the assumption of perfect rationality to model overconfidence. We consider models in which traders over-estimate the mean of the return process, under-estimate the variance or over-estimate the serial correlation.

Barber and Odean (2000) conclude with a powerful statement

> Our central message is that trading is hazardous to your wealth.

This conclusion reflects their finding that net returns are lower for agents who trade more actively without earning higher gross returns. This trading behavior is then viewed as irrational.

We do not concur. It is certainly possible for gross and net returns of traders to be below those of non-adjusters in an optimizing framework. The question is quantitative: do the costs of trade, the processes for income and returns and the rational choices of households generate the pattern of gross and net returns found in the data?

We find that none of these specifications are capable of matching the observed differences between gross and net returns as a function of trading frequency. While there are differences between returns earned by adjusters and non-adjusters, these are reflected in both the gross and net returns earned by traders. In contrast, the facts presented by Barber and Odean (2000) highlight differences in net but not gross returns.

Our results come from two sources. First, the estimated adjustment costs are not large enough to explain the observed differences between gross and net returns. Second, at the estimated parameters, the selection effects coming from the household choices on both the extensive and intensive margins are very powerful so that, in our baseline model, adjusters earn a higher net (and thus gross) return than non-adjusters. In the model with overconfidence, there are cases in which adjusters earn less than non-adjusters. But even here, the selection effect is the dominant factor as the difference appears in both the gross and net returns.

While trading is costly, how hazardous it actually is remains an open issue.

2 Household Behavior: Model and Estimates

Here we briefly review the model and estimates of Bonaparte and Cooper (2009) that is the basis of the household optimization problem we study. The key to the model is the household choice of whether to adjust its portfolio or not. Adjustment is costly due to the presence of fixed and variable trading costs. The household may choose not to incur these costs, in which case consumption is equal to its labor income. If the household adjusts, then it incurs a cost of portfolio adjustment. In this way, the model generates two types of turnover: the discrete choice of whether to adjust and the continuous choice of how much to adjustment conditional on having incurred fixed adjustment costs.

2.1 Household Optimization

Denote by $v(\Omega)$ the value of the household's problem in state $\Omega \equiv (y, s_{-1}, R_{-1})$ where y is current income, s_{-1} is holding of the single asset from the previous period and R_{-1} is the return from the previous period.[1] Total financial wealth this period is $R_{-1}s_{-1}$.

The household chooses between the options of adjusting or not:

$$v(\Omega) = \max\{v^a(\Omega), v^n(\Omega)\} \tag{1}$$

for all Ω. If the household chooses to adjust, then

$$v^a(\Omega) = \max_s u(c) + \beta E_{R,y'|R_{-1},y} v(\Omega') \tag{2}$$

[1] For the purpose of this exercise, the household has a single financial asset.

where s is the new holding of the asset and Ω' is the future value of Ω. Household consumption is

$$c = R_{-1}s_{-1} + y \times \psi - s - C(s_{-1}, s). \tag{3}$$

There are two costs of adjustment here. The first, given by the function $C(\cdot)$, represents direct trading costs. The second, parameterized by ψ, represents the lost income due to time spent on portfolio adjustment. As households will experience different realizations of income they will have different adjustment costs.

If there is no portfolio adjustment, then

$$v^n(\Omega) = u(y) + \beta E_{R,y'|R_{-1},y} v(\Omega'). \tag{4}$$

In this case, $c = y$ and $\Omega = (y', s_{-1}R_{-1}, R)$ so that gross proceeds from the existing portfolio create the portfolio for the current period without any cost, $s = s_{-1}R_{-1}$.[2]

The policy functions generated by household optimization include an extensive margin (adjust, no adjust) and an intensive margin indicating the magnitude of the adjustment. Note that optimizing households base their choices on realized income and returns. Their choice of portfolio turnover reflects net rather than gross returns.

2.2 Quantitative Analysis

Bonaparte and Cooper (2009) estimate trading costs, $C(\cdot)$, directly from the data set used by Barber and Odean (2000). A simulated method of moments approach is used to estimate other parameters.

2.2.1 Trading Costs

Bonaparte and Cooper (2009) assume:

$$C^b(s_{-1}, s) = \nu_0^b + \nu_1^b(s - s_{-1}) + \nu_2^b(s - s_{-1})^2 \tag{5}$$

if the household buys an asset, $s > s_{-1}$. If instead the household sells, $s < s_{-1}$, then

$$C^s(s_{-1}, s) = \nu_0^s + \nu_1^s(s_{-1} - s) + \nu_2^s(s - s_{-1})^2. \tag{6}$$

They use the monthly household account data Barber and Odean (2000) to estimate these parameters.[3] The trading costs are estimated in a regression where the dependent

[2] Bonaparte and Cooper (2009) discuss and estimate alternatives to this model of no adjustment. This specification fits the data best.

[3] Details on the estimation can be found in Bonaparte and Cooper (2009). Through this procedure, we are able to decompose the commission costs reported in Table 1 of Barber and Odean (2000) into fixed and variable components.

variable is the commission and the independent variables are trade value (the price of the share times the quantity of share) and trade value squared per stock. Bonaparte and Cooper (2009) report the estimates in Table 1.

Parameter	**Buying**	**Selling**
Constant ν_0^i	56.10	61.44
	(0.05)	(0.061)
Linear ν_1^i	0.0012	0.0014
	(1.63e-06)	(1.93e-06)
Quadratic ν_2^i	$-1.01e^{-10}$	$-1.28e^{-10}$
	(2.88e-13)	(9.26e-13)
Adj. R^2	0.251	0.359
Number of Observations	1,746,403	1,329,394

Table 1: Estimated Trading Costs

Though the linear and quadratic terms are statistically significant, the main cost of adjustment is the fixed cost per trade. While this cost may seem high relative to current trading costs, it is still small compared to the average trade of a household in the data set of about $12,500.

These estimates of trading costs do not include the bid-ask spread which, according to Barber and Odean (2000) are about 0.3% for purchases and 0.69% for sales. These additional costs are added to the linear terms reported in Table 1 when the trading costs are integrated into the household optimization problem.

2.2.2 Income and Returns

The appendix of Bonaparte and Cooper (2009) contains more detailed information on the data. Income and returns are modeled as AR(1) processes. Household income variable is the product of a common shock and a household specific shock.

We consider a couple of return processes. The first, as in Bonaparte and Cooper (2009), measures real returns, including dividends, and comes from Robert Shiller, available at `http://www.econ.yale.edu/~shiller/data.htm`. The mean return is 5.51% over the 1967-1992 period with a serial correlation of essentially zero and a standard deviation of 0.1589. We refer to this as the "baseline".

The second measure of returns is computed directly from the household data used in Barber and Odean (2000).[4] This return measure is not used in Bonaparte and Cooper

[4]For the return process, we employ the Barber and Odean (2000) data. Our sample selection includes

(2009) but is used here to try to mimic the returns faced by the households whose behavior generated the differences in gross and net returns described by Barber and Odean (2000). For that process, the mean return is about 17.6% over the sample, the serial correlation is 0.107 and the standard deviation of the innovation is 0.2725. We refer to this case as "HH" since it is based upon the average of household returns in the Barber and Odean (2000) data.

2.2.3 Moments

Bonaparte and Cooper (2009) use three moments to identify the parameters (β, γ, ψ). All moments are from annual data.

The first moment comes from the SCF data set where in an average year 71% of households adjustment their portfolio. The second moment comes from the estimation of the log-linear approximation of a consumption Euler equation, drawing upon Hansen and Singleton (1983):

$$log(\frac{c_{t+1}}{c_t}) = \alpha_0 + \alpha_1 \times log(R_{t+1}) + \zeta_{t+1}. \tag{7}$$

The estimate of $\alpha_1 = 0.0878$ is from data on real consumption growth of non-durables and services of stock market participants and the Shiller measure of return for the 1967-92 period.

The final moment is the median financial wealth to income ratio, which is 1.03 in the data. This moment is quite informative about β.

2.2.4 Estimation Results

Estimation involves the solution of an agents dynamic optimization problem, (1)-(4), and the creation of a simulated panel data set with 500 households and 500 time periods. Households differ because of idiosyncratic income shocks which generates differences in trading patterns and returns.

Parameter estimates are obtained by minimizing the distance between the simulated and actual moments. Bonaparte and Cooper (2009) estimated (β, γ, ψ) using three moments: the frequency of inaction, the estimate of α_1 from (7) and the median financial wealth to income ratio.

The moments from the data and the parameter estimates are summarized in Table 2.[5] The column labeled "Data" shows the actual moments and the column labeled "Baseline" is

households who hold stocks consecutively for the whole period from February 1991 through December 1996. The selected sample contains 1,079,877 observations for 14,478 households; each household has 71 monthly consecutive observations.

[5] The weighting matrix was the identify matrix for this exercise so that the fit is simply the sum of the squared differences between the actual and simulated moments.

from the estimated model. The column labeled “HH” are results based upon the household return process.

Moments	**Data**	**Baseline**	**HH**(data)	**SC**
Portfolio Adjustment rate	0.71	0.72	0.61 (0.52)	0.74
α_1	0.088	0.110	-	0.0969
wealth income ratio	1.03	1.03	3.80 (3.80)	1.04
Estimated Parameters				
γ	-	3.51	-	3.51
ψ	-	0.974	0.834	0.974
β	-	0.88	0.797	0.88
fit	-	0.0006	0.0091	0.001

Table 2: Moments and Parameters

Baseline The data moments are summarized in the second column. The estimates are presented in the third column.[6] The moments are pretty well matched though the value of α_1 is not quite as low as in the data. The degree of risk aversion is considerably less than the inverse of α_1. Interestingly, there is evidence of adjustment costs beyond the trading costs: $\psi < 1$. The fit of 0.0006 is calculated as the sum of the squared differences between the simulated and data moments.

HH Returns The results using the return process estimated from household data are presented in the “HH” column. The portfolio adjustment rate and the wealth income ratio moments were recomputed for this sample and are indicated in the table as 0.52 and 3.80 respectively. Due to data limitations for these households, we cannot re-estimate α_1.

Given the portfolio adjustment rate and wealth income ratio as moments, we re-estimated (ψ, β) to match these two parameters, fixing $\gamma = 3.51$. The estimated β is much lower for this sample, reflecting the high average return on this portfolio. Further the cost of adjustment, ψ, is much lower as well to generate the lower adjustment rate of 52% in the household sample. The fit is not as good with this sample as it is for the baseline model. Still, this sample will provide additional information on the differential of returns for adjusting and non-adjusting households.

[6] Bonaparte and Cooper (2009) study the robustness of these estimates to different measures of return and other specifications of the model.

3 Returns and Trading

Given this model and estimates, we now turn to the main point of the analysis: the relationship between returns (both gross and net) and trading patterns. As noted above, Barber and Odean (2000) find an inverse relationship between net returns and trading frequency but no significant differences in gross returns.

We use our model to evaluate this evidence. In our model all agents have rational expectations so that any difference in returns associated with trading frequency comes from the optimal choices of households. For these households, their choices on the extensive and intensive margins depend on perceived net returns.

In theory, there are two ways in which the model can link differences in net return to trading frequency. The first is direct: the presence of a trading cost will reduce the net return. The second, more subtle link, can be generated by selection: agents choose whether to adjust or not (the extensive margin) and how much to adjust (the intensive margin) in a state contingent manner. Thus optimal behavior on the part of agents will itself generate a relationship between returns and trading decisions on both the extensive and intensive margins.

Table 3 presents calculations of different return measures for the baseline model using the parameters reported in Table 2. The gross return is simply the annual return on the portfolio. The second return measure nets out the financial costs, from Table 1, of trading. The third measure nets out both the financial cost and the income loss due to ψ.[7]

The timing here is important. The return is realized from period $t-1$ to t. The adjusters/no-adjusters distinction refers to the decisions taken in period t. Thus the table analyzes whether adjustment responds to higher or lower realized returns, as in the state vector of the optimization problem. In this sense, the dependence of the return on adjustment reflects the selection of whether to adjust or not.

The columns of the table relate to the extensive and intensive margins. The second and third columns of the table look at agents who adjust and those who choose not to adjust. The fourth, fifth and sixth columns compute return for the lowest, middle and highest quintiles of the turnover rate distribution.[8]

Baseline Looking first at the results from the baseline, the table reveals that neither the transactions cost explanation of the net return differential between adjusters and non-

[7] To be specific, the net return is the gross return minus the average cost of trading. The average cost of trading equals the total cost of trading divided by the quantity traded. For timing, the net return in period t is the gross return earned between periods $t-1$ and t minus the trading costs incurred in period $t-1$ as a fraction of the period $t-1$ asset trade.

[8] The turnover rate is the absolute value of the net change in the portfolio divided by its initial value.

Return	Adjusters	Non-adjusters	Lowest Turn.	Middle Turn.	Highest Turn.
Baseline					
Gross	1.060	1.049	1.050	1.061	1.065
Net FC	1.058	1.049	1.050	1.060	1.061
Net All	1.025	1.049	1.050	1.049	0.977
HH					
Gross	1.189	1.151	1.149	1.066	1.263
Net FC	1.189	1.151	1.149	1.066	1.262
Net All	1.134	1.151	1.149	1.046	1.157

Table 3: Returns and Trading

adjusters nor the selection story fits the facts. For both cases, the key observation is that the gross return for adjusters, 1.060, exceeds that of the non-adjusters of 1.049. This is inconsistent with the findings of Barber and Odean (2000).

To understand these results, consider the transactions costs explanation for the differentials in returns. Looking at adjusters, the return net of transactions costs is only slightly lower than the gross return. From Table 1, the main adjustment cost is the fixed component of around $60.00. For these costs to create a differential in return of 6 percentage points, as reported by Barber and Odean (2000), the average trade would have to be about $1,000. In fact, the average trade in the data is about $12,500 and is about $13,800 in the simulated data. Thus the fixed cost is much too small relative to the trade size in both the actual and simulated data to explain the differential in returns.

The selection effect can create differences in net return based on trading frequency. But, that explanation runs into two problems. First, as can be seen from Table 3, the selection effect generates differences in gross as well as net returns. This runs counter to the evidence from Barber and Odean (2000). Second, the selection effect creates a higher, not lower, return for the adjusters. This is because agents adjust more frequently in high return states than in low return ones. We return to this point below.

Barber and Odean (2000) study the returns for low and high turnover households, thus focusing jointly on the extensive and intensive margins. Turning to the model's implications for the relationship between return and turnover, we see the same basic patterns. We split our data into quintiles based upon the absolute value of the turnover rate. Table 3 reports the returns for the lowest and highest turnover groups.[9] The gross return of the high turnover

[9]The lowest group has zero turnover as does a fraction of the second lowest group. As the return for non-adjusters is not identical, the return for the lowest turnover group is not the same as the return for the non-adjusters.

group exceeds that of the low turnover group. Further, the adjustment costs are not large enough to overturn that ordering for net returns.

HH Returns The lower section of Table 3 shows the measured returns from the HH simulations. Recall that these simulations are based on a return process calculated for individual households from the Barber and Odean (2000) study.

As seen in Table 3, the results follow the same pattern as the baseline. The returns (gross and net) are higher for adjusters and for the highest turnover group, compared to the low turnover group. Despite using the return process in the data from the Barber and Odean (2000) study, the pattern of gross and net returns that they uncovered does not appear in the simulated data based upon optimizing behavior. Once again, the key is the selection effect since trade is slightly more frequent in high return states.

There is one important difference between our model and the underlying data studied by Barber and Odean (2000). Our model is specified and estimated at an annual frequency. In contrast, Barber and Odean (2000) have monthly data and their return differentials are geometric annualized monthly returns. Thus relatively small differences in monthly returns can become large differences in annual returns. If an agent trades frequently enough to incur a transactions cost each month which reduces the net return in that month by one-half of a percentage point, then the cost is over 5 percentage points on an annual basis. So while it might be that the transactions costs could be large enough at a higher frequency to explain the lower net return of adjusters, this point does not explain the differences in gross returns which reflects the selection effects.[10]

Alternative Parameters We consider a couple of extensions of our model to study the impact of variations in risk aversion and trading costs on returns. If agent's are more risk averse, then they are more likely to trade for the purposes of consumption smoothing. To study that, we increase γ from 3.51 to 10.0. We find households trade more often but the differential in the gross rate of return remains: the gross return for the adjusters is nearly 4 percentage points higher than the gross return for non-adjusters. But, as before, the difference between gross and net returns for traders is negligible.

Another interesting possibility is that agents actually prefer to trade so that $\psi > 1$. Though the estimate of ψ is less than one, it is interesting to see what return patterns are produced by this model. At $\psi = 1.01$, all households adjust their portfolios. This reflects the fact that the financial costs of trading are low enough that they do not by themselves

[10] In some simulations at the monthly frequency, the returns of adjusters exceed those of the non-adjusters, as in the results reported in Table 3. But due to the discount factor near unity and the lack of high frequency moments, estimation of the model is not possible.

create inaction: the gross and net returns are about the same for the adjusters.

However, there are interesting patterns of returns across the turnover rate. The net and gross return is a concave function of the turnover rate, reaching 9% for the middle quintile and falling to nearly 6.2% for the highest turnover group. When ψ is bigger than one, some of the trades are undertaken by high income households simply for "the joy of trading" even though returns are not high enough to warrant these trades. Still these are differences in both net and gross returns, not in net returns alone.

Finally, we looked at adjustment costs in excess of those estimated in Table 1. For this experiment, the fixed cost of trading was increased to $1000. The gross return to adjusters exceeded that of the non-adjusting households while the net returns were about the same.

In another experiment, the linear term was increased to 5%. This additional cost also did not influence the basic findings of our model: gross returns and net returns are higher for adjusters (high turnover households) than non-adjusters (low turnover households).

Differences in Expected Utility As a final point, note that for all of these treatments this model has completely rational households. All trades are consistent with the maximization of discounted expected utilities. Investors who trade do so precisely because the expected utility from trading **exceeds** that of not trading.

In contrast, Barber and Odean (2000) say:

> The two models yield different predictions about the gains of trading. The rational expectations model predicts that investors who trade more (i.e., those whose expected trading is greater) will have the same expected utility as those who trade less. The overconfidence model predicts that investors who trade more will have lower expected utility.

This theme that traders who trade more have the same expected utility as those who trade less is not a property of the model. The choice of whether to trade comes from the optimal choice of the agents in (1). There is no presumption of indifference. Further, Barber and Odean (2000) contend that in a model with rational investors (they refer to the model of Grossman and Stiglitz (1980)) more active traders will have higher gross returns but no difference in net returns. The gross return differential is a property of this model but net returns also differ across traders even though they are rational. A key aspect of our model is the heterogeneity across agents in income and wealth. In an optimizing model, these differences induce the utility differentials between adjusters and non-adjusters as well as the selection effects which underly the reported relationships between adjustment, turnover and returns.

4 Evaluating Models of Overconfidence

Given that the presence of transactions and opportunity costs are not enough to create the pattern of gross and net returns found by Barber and Odean (2000), we turn to their favored explanation: overconfidence. We model overconfidence in three ways: (i) a higher than actual return, (ii) a lower than actual standard deviation of the return and (iii) more persistence in the return process.

4.1 Models of Overconfidence

Consider the following process for the beliefs of agents about returns:

$$R_t = \tilde{\mu} + \tilde{\rho} R_{t-1} + \varepsilon_t \tag{8}$$

where ε is normally distributed with a mean of 0 and a standard deviation of $\tilde{\sigma}$.[11] The mean return is denoted $\tilde{\mu}$ and the serial correlation is $\tilde{\rho}$. This process may not coincide with the true process for returns. Indeed, our interest is in studying the relationship between beliefs and the true process for trading strategies and portfolio returns.

4.2 Approach

From (8), our specification permits three types of deviations through the: (i) mean, (ii) standard deviation and (iii) persistence of the return process. For each of these specifications, we parameterize the deviation from the true process and solve the model assuming household's hold these beliefs. The realized net and gross returns are calculated from simulated data where the actual process is used for the exogenous returns process.

The first deviation from truth comes from excessive optimism about the mean of the process, so that $\tilde{\mu}$ is higher than the actual mean of the return process. The second deviation allows the household to believe that the standard deviation of the process is smaller than truth: i.e. $\tilde{\sigma}$ is less than the true standard deviation of the innovation. In this case, the household perceives less uncertainty than reality.

The final case allows a deviation between the perceived and actual serial correlation of the return process, $\tilde{\rho}$ may differ from the actual serial correlation parameter.[12] When returns are

[11] Here we start from a standard AR(1) model and draw on the discussion in DeLong, Shleifer, Summers, and Waldmann (1991). While there are numerous papers in the literature using the concept of overconfidence, there are relatively few which point to a particular model of overconfidence. Gervais and Odean (2001) study overconfidence in a learning model that is beyond the scope of our study. Guiso and Jappelli (2006) study the effects of overconfidence on information acquisition.

[12] See Daniel, Hirshleifer, and Subrahmanyam (2001) for a discussion of models of overconfidence in which agents overestimate the informativeness of signals.

serially correlated, the realized return has an effect on current wealth and on the distribution of future returns. The latter effect is like a signal. For our estimated return process, there is no evidence of serial correlation over the sample period. Thus our return process is iid and the current return provides no information about the future. If, however, $\tilde{\rho}$ is positive, then households are more confident that current returns provide information about future returns.

4.3 Results

Return	Baseline	Mean		Std.		Ser. Corr.	
		$\mu * 1.01$	$\mu * 1.05$	$Std * 0.9$	$Std. * 0.8$	$\rho = 0.1$	$\rho = 0.3$
Non-Adjusters							
Gross	1.0494	1.0495	1.0495	1.0488	1.0484	1.0697	1.0750
Adjusters							
Gross	1.06	1.06	1.0579	1.06	1.0602	1.0515	1.0502
Net FC	1.0583	1.0585	1.0574	1.0584	1.0584	1.0500	1.0465
Net All	1.025	1.0312	1.0477	1.0264	1.022	1.0196	0.9870
Highest Turnover							
Gross	1.0652	1.0608	1.0975	1.0666	1.0667	1.0670	1.0690
Net FC	1.0608	1.0569	1.0959	1.0625	1.0616	1.0629	1.0573
Net All	0.9768	0.9858	1.0628	0.9829	0.9669	0.9855	0.8749
Lowest Turnover							
Gross	1.0501	1.0505	1.0137	1.0536	1.0677	1.0685	1.0712

Table 4: Models of Overconfidence

Our findings for the various experiments are summarized in Table 4. To be clear, these calculations are all at the baseline parameters, including the stochastic processes for income and interest rates, allowing for only one form of overconfidence at a time.[13] These deviations from truth are used to index the columns of the table. The first set of rows report the gross returns for adjusters and non-adjusters, which indicate whether portfolio adjustment occurred in a given period or not. The second set of rows reports returns for the highest and lowest turnover groups. As the lowest turnover group did not trade, only gross returns are reported.

[13] Hence in the simulations the distribution from which the returns are drawn differs from the beliefs of agents.

There are two types of net returns reported, corresponding to the objects calculated in Table 3. The first, "Net FC", calculates the net return accounting for direct financial costs of the actual trade, using the estimates in Table 1. The second "Net All" uses both the direct financial costs as well as the opportunity cost, parameterized by ψ, to calculate trading costs.

We first discuss results based on the extensive margin of adjust or not. We then discuss the relationship between return and turnover rates.

The results from the baseline appear in the second column of the table. Recall that there is a gross rate of return differential in favor of the adjusters. This reflects the selection effect. The net return, though of course lower than the gross return, is still higher than the return of the non-adjusters.

The first form of misperception is in the mean of the return process, $\tilde{\mu} > \mu$, shown under the columns labeled "Mean". If there is overconfidence about the mean of the return, then the adjustment rate increases. When the mean is viewed as higher than truth by a factor of 1.01, there is relatively little change in the returns though the adjustment rate is higher. At a mean return increased by a factor of 1.05, the same pattern of net and gross returns remained.

The second form of misperception is in the standard deviation of the innovation, $\tilde{\sigma} < \sigma$, shown under the columns labeled "Std." If the perceived standard deviation is lower than truth, the gap in gross returns widens with the adjusters having the higher gross return. As in the other specifications, the costs of trade reduce the net return, but it still exceeds the return of the non-adjusters.

The final case of misperception in the serial correlation, $\tilde{\rho} > \rho = 0$, shown under the columns labeled "Ser. Corr" is quite different.[14] Here the agents are more confident that the returns will persist and are responsive to this signal. In this case, the gross, and hence net return, to the adjusters is below the gross return of the non-adjusters. This is much closer to the pattern highlighted by Barber and Odean (2000). But, as before, this pattern appears in the differential on gross returns.

Though these models of overconfidence do not match the findings of Barber and Odean (2000), they do have other interesting implications. The column in Table 3 labeled "SC" presents the moments when households believe $\rho = 0.3$. In this case, the overall fit of the model, though not as good as the baseline of 0.0006, is 0.001 using the baseline parameters.[15] The goodness of this fit comes from the reduction in α_1 from 0.11 to 0.0969. With more perceived persistence in the interest rate, the amount of saving is more responsive and thus consumption less responsive to interest rate movements.

[14] Daniel, Hirshleifer, and Subrahmanyam (2001) also discuss the implications of this last form of overconfidence for asset trades. They focus on the intensive margin.

[15] The fit can be improved further by additional estimation.

Instead of looking at the margin of adjust or not, it is informative to look at the patterns of returns for different turnovers rates. The results for the highest and lowest turnover groups mimic the results based on adjust/no adjust. That is, except for the of misperceptions about the serial correlation, the gross returns are higher for the highest turnover group of agents. But this ordering is switched in the "Ser. Corr. =0.3" case where the returns for the lowest turnover group are slightly higher than the highest turnover group. In this case, however, it is important to note that the gross returns are lowest for the middle turnover group, at 1.0356, so that there is not a monotone relationship between returns and turnover rates in this case. In contrast, the relationship is monotone in the results of Barber and Odean (2000).

It is useful to see how overconfidence impacts on the likelihood of trade. To do so, compare the choices of whether to adjust or not for two types of households. The first, rational trader, has preferences and beliefs based on the baseline model. The second, overconfident trader, has the same preference but with beliefs that $\tilde{\rho} = 0.3$. We focus on the sensitivity of trade to variations in the gross return on the asset.

The results are summarized in Table 5. For the baseline model of a rational agent, the probability of making a trade is about 71% in the high return state but only 66% in the low return state. Thus, as noted in the discussion of Table 3, the gross return of adjusters is lower than the gross return of the non-adjusters.

But, in this model of overconfidence, the results flip. As is evident from Table 5, the adjustment rate is higher in the low return states for the overconfident agents. The perception of a persistent low returns leads them to adjust their portfolio.

specification	High Return	Low Return
baseline	0.71	0.66
OC	0.66	0.74

Table 5: Adjustment Rate Dependence on Gross Return

The results in Table 5 complement the findings on gross returns reported in Table 4. When the adjustment rate is higher in high return states, as in the baseline, then gross returns are higher for adjusters compared to non-adjusters. But, when the adjustment rates are lower in high return states, as in the case with a belief in a serial correlation of 0.3, then the realized returns of the adjusters are lower than those of the non-adjusters.

5 Conclusion

The goal of this paper was to assess the claim in Barber and Odean (2000) that the patterns of returns as a function of the frequency of trade was consistent with overconfident agents and inconsistent with rational traders. In our model, the frequency of trade translates into whether households incur a cost to adjust their portfolio or not.

Using parameter estimates which match moments of adjustment rate, the sensitivity of consumption growth to interest rate movements and the volume of trade, we found that the model of rational agents produced differential in both gross and net returns in which adjusters earned more than non-adjusters. Since agents are utility maximizing, trading choices are optimal *ex ante* though in any dynamic stochastic model, there can be *ex post* regret.

Models of overconfidence in either the mean, the standard deviation or the persistence of shocks, did not match the observations of Barber and Odean (2000) either. In particular, these models also created differences in net returns from differences in gross returns. This reflects the power of the selection effects which dominate trading costs in determining the pattern of net and gross returns.

We have focused on only a few of potentially many models of overconfidence. We are confident only that the models we have looked at are at odds with the data.

References

Barber, B., and T. Odean (2000): "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *The Journal of Finance*, 55(2), 773–806.

Bonaparte, Y., and R. Cooper (2009): "Costly Portfolio Adjustment," NBER Working Paper # 15227, revised 2010.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (2001): "Overconfidence, arbitrage, and equilibrium asset pricing," *Journal of Finance*, pp. 921–965.

DeLong, J., A. Shleifer, L. Summers, and R. Waldmann (1991): "The survival of noise traders in financial markets," *The Journal of Business*, 64(1), 1–19.

Gervais, S., and T. Odean (2001): "Learning to be overconfident," *Review of Financial Studies*, 14(1), 1.

Grossman, S., and J. Stiglitz (1980): "On the impossibility of informationally efficient markets," *The American Economic Review*, 70(3), 393–408.

Guiso, L., and T. Jappelli (2006): "Information Acquisition and Portfolio Performance," mimeo.

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

Rationalizing Trading Frequency and Returns: Maybe Trading is Good for You*

Yosef Bonaparte† and Russell Cooper‡ and Mengli Sha§

June 9, 2023

Abstract

Barber and Odean (2000) find that households who trade more have a lower net return than others and attribute this pattern to irrationality, specifically overconfidence. In contrast, we find that household financial choices generated from a dynamic optimization problem with rational agents and portfolio adjustment costs can reproduce the observed distribution of (net) turnover rates as well as the fact that households with the highest turnover rates have the lowest net returns. Irrationality, modelled as beliefs about the serial correlation of asset returns that are not data based, only marginally improve the ability of the model to explain these turnover and net returns patterns. The potential for irrational choice appears in the decision to own stocks directly.

1 Motivation

In a very influential contribution, Barber and Odean (2000) find that households with higher stock turnover have a lower net return. They interpret this as evidence households are overconfident and thus not rational. They state:

> *Our most dramatic empirical evidence supports the view that overconfidence leads to excessive trading ... On one hand, there is very little difference in the gross performance of households that trade frequently with monthly turnover in excess of 8.8 percent and those that trade infrequently. In contrast, households that trade frequently earn a net annualized geometric mean return of 11.4 percent, and those that trade infrequently earn 18.5 percent. These results are consistent with models where trading emanates from investor overconfidence, but are inconsistent with models where trading results from rational expectations.*

*We are grateful to the National Science Foundation for supporting the early stages of this research. This paper is a major revision of one circulated earlier as NBER Working Paper #w16022 entitled "Rationalizing Trading Frequency and Returns: Trading is Good for your Wealth" by Yosef Bonaparte and Russell Cooper. We thank seminar participants at the Federal Reserve Banks of Dallas and Kansas City, Cambridge University, the CCER at Peking University, the EIEF in Rome, Melbourne University, Oxford University, the Pennsylvania State University and the Toulouse School of Economics for comments and suggestions on earlier work that underlies this paper. Guozhong Zhu provided outstanding research assistance on the earlier paper. Comments and suggestions from seminar participants at the European University Institute and Universitat Bonn on the most recent incarnation of the research are greatly appreciated. Suggestions from the editor and referees led to major improvements in this paper.

†University of Colorado Denver, Yosef.bonaparte@ucdenver.edu

‡Department of Economics, European University Institute; Department of Economics, the Pennsylvania State University; Research Associate, NBER, russellcoop@gmail.com

§Department of Economics, the Pennsylvania State University and Bates White Economic Consulting, mengli.sha0124@gmail.com

This is a very provocative finding and one that has rightly received considerable attention. But there are questions left open concerning the interpretation of these trading patterns.

First, the results of Barber and Odean (2000) come from the trading activity at a discount broker of about 66,000 households. These are atypical in that these households hold stocks directly, trade in a particular fashion and have higher income than average.[1] Should we interpret their findings as indicative of some universal form of overconfidence among traders? Or is there a significant fraction of agents whose trades are rationalized by standard intertemporal choice models, leaving only a subset of irrational types? That is, do the average households in the US suffer from the overconfidence that Barber and Odean (2000) see as an explanation for the patterns in their data?

Second, focusing on the sample from Barber and Odean (2000), to what extent is some form of irrationality needed to match data moments? And, what is the form of the irrationality? These questions are not convincingly addressed in their paper as there are no formal models of rational nor irrational traders that are specified and taken to the data.

With ideal data, these questions would be answered jointly. Imagine a panel of households with the following information. First, there is an indicator of asset market participation including households with: (i) direct and indirect stock holdings, (ii) only indirect stock holdings and (iii) non-participants in the stock market.[2] Second, the data includes the realized income for each household in each period, allowing the estimation of the income process. Finally, there is detailed information on household portfolios, including: (i) holdings of all assets, (ii) returns on these assets and (iii) all purchases and sales. With this complete data set we could specify and estimate the parameters of a household choice problem, including asset market participation (direct and indirect), portfolio adjustment and the share of stocks in total financial assets. The model could be broad enough to allow and identify some forms of irrational behavior.

In reality, we are quite far from this ideal. From Barber and Odean (2000) we have detailed data on the trades of households within a particular discount broker. We lack the information needed to estimate the income processes for these households and also do not have information on asset holdings outside of the discount broker. While the PSID and the SCF contains additional information, neither have detailed information on turnover rates. Moreover, estimating the dynamic problem with this rich set of choices combined with the prospect of irrational agents is computationally very demanding.

Consequently our analysis is more modest. We focus on variants of the two issues raised earlier, addressing them through quantitative exercises. Our approach is based upon the estimation of a structural model of household choice through a simulated method of moments methodology. Key ingredients of the household problem are fluctuations in income and costly portfolio adjustment. The estimation strategy finds the parameters of the structural model that match standard household finance data moments. These data moments are informative about the underlying structural parameters. The method allows for forms of household irrationality.

A key value of this approach is the feasibility of generating moments from the estimated model that are directly observed in the data. Given that we do not have detailed data on turnover patternss, as Barber and Odean (2000) do for their particular sample, we can nonetheless simulate these turnover moments as a way to evaluate hypotheses about the relationship between turnover and net returns. With this methodology we address two questions.

[1] Looking, for example, at the SCF 2001 wave (using proper weights), the stock market participation rate (both direct and indirect holdings) is about 45% with about 21.3% of households held stocks directly.

[2] By direct holdings we mean stocks held in an individual's account, such as the discount broker accounts studied in Barber and Odean (2000). Stocks held through a mutual fund are a form of indirect holdings.

Q1: What is the quantitative contribution of Liquidity Trades? A natural starting point for explaining evidence on turnover and net returns is the tradeoff between liquidity and return. This tradeoff naturally emerges in household finance models with costly portfolio adjustment. In these models, rational households hold a portfolio of consisting of: relatively liquid low return bonds and higher return but less liquid stocks. The presence of non-convex adjustment costs and fluctuating incomes leads to infrequent portfolio rebalancing in order to smooth consumption.

Barber and Odean (2000) argue that these forces are not sufficient to explain the patterns of turnover and net return:

> they seem implausible as an explanation of the 75 percent annual turnover that we document for the average individual investor and belie common sense as an explanation of the more than 250 percent annual turnover of the households who trade most. Investors facing rapidly fluctuating liquidity needs can, in most cases, find less expensive means to finance these than rapid trading in and out of stocks.

Even if they are correct and liquidity motives alone cannot match the excessive turnover, it is important to understand how far it goes in matching their facts. Doing so generates provides a more precise sense of the role of irrational trades. It also makes clear whether the irrationality pertains to only a subset of households or a subset of trades. And, as indicated in the last sentence, the puzzle may be more about the choice of participation in direct asset holdings rather than trading behavior conditional on participation.

Our approach to this question has two important restrictions. First, we focus on rebalancing between stocks and bonds (liquid assets), not on portfolio rebalancing. Second, the baseline model has only rational agents.

The moments chosen for the estimation are standard from the household finance literature. Importantly, we focus on asset market participants, though without distinguishing between direct and indirect holdings. The moments include a measure of inaction in portfolio adjustment from the PSID, but do not include the entire distribution of turnover rates from the Barber and Odean (2000) sample. An extension adds monthly turnover in the highest quintile as an additional moment.

This exercise leads to one of our findings. The estimated model is able to reproduce both the large turnover rates and the negative correlation between turnover rates and net return driven solely by the interaction of fluctuating income and non-convex adjustment costs. These patterns are generated without irrational traders.

Q2: What is the contribution of Irrationality? The model and estimation is extended to include irrational agents, defined as agents whose beliefs about the return process are inconsistent with the estimated process from the data. We argue that as long as choices depend on beliefs, these can be inferred through the estimation.

Focusing on net trades, we find some limited evidence that irrationality matters. Introducing beliefs about positively serially correlated returns we are able to modestly improve the fit of the model. But we argue that the results are so close to the rational model that, in effect, little is added by the presence of the irrational agents.

A second component of irrationality has to do with the participation decision, i.e. the opening and closing of an account. From a comparison of balanced and unbalanced panels, exit matters for the measures of turnover and net return. In particular, the turnover rate in the highest quintile almost triples when looking at net trades. Nearly 50% of the trades in the top quintile are in excess of 100% in the unbalanced panel. From the data, exit and excessive turnover are related: the exit comes largely from households with high turnover. An open puzzle is the motivation for direct holdings. Does it reflect overconfidence in the ability to predict stock returns or a merely a value attached to the direct control of a portfolio? Even if overconfidence lies behind the motivation to open an account, the data suggest that agents do respond to adverse outcomes of high turnover and low returns through exit.

We conclude that overconfidence is not needed to explain the trading patterns in Barber and Odean (2000), excluding portfolio churning. For our measurement of moments and estimation, net returns are indeed lower for more active traders. But this is fully consistent with the choices of rational optimizing agents. While a model with rational agents matches key moments, it leaves open extensive margin choices regarding the open and closing of accounts.

2 Facts: Turnover and Net Returns

This section presents the patterns of turnover and net return that underlie the Barber and Odean (2000) findings. The data is a panel, indicating trades by household over their period of activity with this particular discount broker.

There are a couple of important issues in calculating the turnover and return moments. First, there is the distinction between gross and net trades. We present turnover moments as net trades between common stock holdings and another group of liquid financial assets, hereafter called a bond. The table that follows presents these turnover moments which are used in the estimation and to address our research questions. A second measure, as in Barber and Odean (2000), is to look at gross trades, thus counting both purchases and sales within the stock account.[3] These are ultimately not used in the estimation as the model does not focus on portfolio rebalancing. Averaging over households and time, net trades are about 58% of total (gross) trades.

Second, there are two ways to create turnover quintiles from these monthly measures of turnover. The first approach to measuring turnover, hereafter called "time series", calculates the quintiles of the time series average of the turnover rates by household.[4] That is, the monthly turnover rates are averaged over time by household. The quintiles are calculated based on these time series averaged turnover rates. This measure captures persistent differences in turnover rates across households. At the extreme, if households were distributed in, say, the utility gain (cost) from trading, then that dispersion would be seen through the time series approach.

The second approach, hereafter called "cross section" turnover, calculates the time series average of the turnover rate, by quintiles, for each month. That is, for each month, the turnover rates are ranked to create average turnover rates for each quintile. Then these rates are averaged over time. With this measure, a household may be in the high turnover quintile in some periods, and not in others. Moments from this measure are not central to the analysis and are presented in Appendix 8.3.

Third, the moments depend on the treatment of exit from the sample. In the analysis that follows, "balanced" moments are those from households who were in the sample for a complete 71 months, starting in January 1991.[5] In contrast,"unbalanced" moments are from households who started in January 1991 but

[3] These moments are presented in Appendix 8.3.

[4] Though not stated in Barber and Odean (2000) this seems to be their methodology as the moments calculated in this way are closer to their summary tables than the cross section approach.

[5] Specifically, this requires 71 months of reports in the position data. Focusing on households with an account in January

did not necessarily maintain an account through the entire sample.

2.1 Measures

The facts about turnover and net return build upon the Barber and Odean (2000) study of household trading activity and return. These include the monthly portfolio turnover rate as well as the net stock return. Here we focus on net trades, thus ignoring portfolio rebalancing.

We report a number of key findings. First, despite the focus on net trades, there is a large amount of turnover in the highest quintile, ranging from a mean of 11.8% in the balanced panel) to nearly 32% in the unbalanced panel) Second, there is a net return gap: traders with turnover rates in the highest quintile have lower net return than those in the lowest turnover quintile. From this we conclude that the patterns uncovered by Barber and Odean (2000) are also present in net trades, not just in the rebalancing of stock portfolios.

Calculating Turnover Rates Portfolio turnover by a household over a month is defined as the absolute difference between end of period and beginning of period stock wealth, divided by the beginning of period stock wealth. In this way, the turnover measure **excludes** portfolio rebalancing, i.e. the trading of one stock for another. Specifically, net turnover for household i in period t is

$$T_{i,t} \equiv |(\frac{A_{i,t} - A_{i,t-1}R_t^s}{A_{i,t-1}R_t^s})| \tag{1}$$

where $A_{i,t-1}$ is the end of period $t-1$ stockholdings for household i.[6] Note that the turnover rate depends on the household's stock wealth at the start of period t, $A_{i,t-1}R_t^s$, inclusive of current stock returns, R_t^s.[7]

Calculating Net Returns We calculate the net return for household i, denoted R_i^n,on the stock portfolio as

$$R_{i,t}^n = \frac{A_{i,t-1}R_t^s - C(A_{i,t}, A_{i,t-1}R_t^s)}{A_{i,t-1}}. \tag{2}$$

Here the cost function comes from the results in Table 2. In calculating the net return costs of trade due to the bid and ask spread are included as well. Note that this is not the net return on an individual trade but rather the net return on the entire stock portfolio.

Clearly, part of the relationship between high turnover and low net return is mechanical. We estimate a quadratic cost function to capture $C(A_{i,t}, A_{i,t-1}R_t^s)$. The linear component directly implies an inverse relationship between turnover and the net return. But the fixed cost and the quadratic cost impact this relationship. In particular, low initial stock holdings imply a large differential between gross and net returns. Thus it is not solely turnover that determines this differential, but the state of the household undertaking the trade as well.[8]

As in Barber and Odean (2000), we compute the return differential on the portfolio of the lowest and highest turnover quintiles, based either on the cross section or time series approach. Either way, this is

1991 is consistent with Barber and Odean (2000). From their sample size, we infer that their statistics come from an unbalanced panel. We present patterns for both balanced and unbalanced panels to evaluate the effects of closing accounts on these patterns of turnover and net returns.

[6] Throughout, we use T when denoting turnover and t as an index of time.

[7] Given the state space representation of the household choice problem, this is the natural way to calculate net turnover.

[8] From the data, the median monthly net return with the fixed, linear and quadratic costs in the T5 quintile is 1.0125 and with the fixed only it is almost the same, 1.0139.

a monthly differential in return. If, for example, one portfolio earns 1.0146 per month and another earns 1.0046, then the difference in returns is 13.34% over a year.

2.2 Time Series Moments

Table 1: Net Trade Balanced Panel

Value	T1	T2	T3	T4	T5	DQ
# of Obs.	2924	5310	2654	2655	2655	-269
Avg. Turnover	0	0.0017	0.0135	0.0296	0.1182	0.1182
Med. Turnover	0	0	0.0134	0.0288	0.0712	0.0712
Avg. Gross Return	1.0134	1.0135	1.0136	1.0136	1.0145	0.0016
Avg. Net Return	1.0134	1.0135	1.0137	1.0130	1.0128	-0.0006
Med. Net Return	1.0127	1.0129	1.0129	1.0128	1.0129	0

This table reports data moments using the time series approach, for the balanced panel. These are the averages across households of the average turn over rates and returns during their period in the sample. Here "Ti" is the turnover rate for quintile "i" and "DQ" is the difference between the highest and lowest turnover rate quintiles. "Avg" means average and "Med" is the median.

Table 1 shows the time series moments based upon the net trade measure.[9] To construct these measures, the turnover rates for each household are averaged over time and then the households are placed into the 5 quintiles.[10] This approach captures variations across households.

For the balanced panel, the lowest quintile of households have a mean and median turnover rate of zero. This is true for the median and almost the case on average for the second quintile as well. In contrast the turnover rate for those households in the highest quintile is 11.8% on average, with a median of 7.1%. The average net return is -0.001 lower for these high turnover households. On an annual basis, this is a difference of nearly 1.2 percentage points.

Importantly, this is not the sample nor the methodology used by Barber and Odean (2000) for calculating turnover and return differential. Yet, the patterns they report appear even in net trades: (i) there is large variation in net trades across households and (ii) higher return households have a lower net return.

2.3 Outliers

There are outliers in the sample, particular in terms of turnover rates. From Table 1, the moments in the fifth quintile, T5, may be driven, in part, by outliers. The median turnover and return differential in that cell is 7.1% and 0% respectively, quite different from the mean values.

This is supported by the top row of Figure 1 which shows the distribution of turnover rates within the T5 quintile for a balanced panel, using the time series approach.[11] The panels differ in terms of the turnover rate, with the right side illustrating the 5% right tail of the T5 quintile distribution, essentially turnover rates in excess of 100%. From the top right panel, there are some extremely high turnover rates in the net trade data that drive the difference between the mean and the median.

[9]To make the sample as close as possible to that of Barber and Odean (2000), there were choices made here regarding the treatment of households who reentered the sample (allowed) and also allowing all households with positive holdings of stocks to be in the sample. These selection issues are discussed in more detail in the Appendix.

[10]Note that by our construction the quintiles do not have an equal number of observations. Because more than one fifth of the sample had 0 turnover, the 0-turnover individuals are both included in T1 and T2, so that T2 has more observations compared to other quintiles. The reason to include 0 -turnover households in both quintiles is that two individuals with the same 0-turnover can have different returns due to portfolio differences, and we don't want to impose assumptions here as to who to include in the 1st vs 2nd quintile.

[11]The bottom panel shows the unbalanced panel, which we return to below.

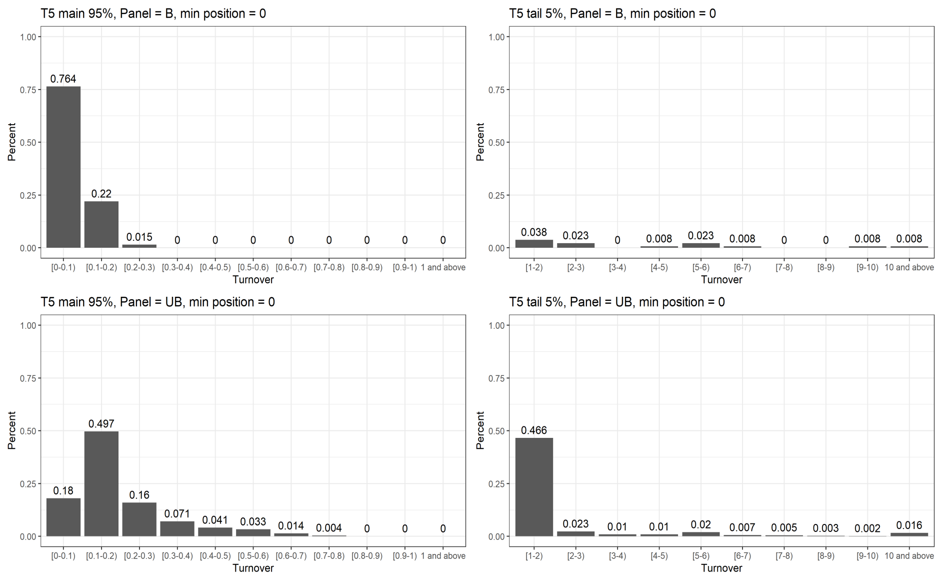


Figure 1: Turnover Distribution: T5

From inspection of the data, the households involved in the trades underlying these very high turnover rates have relatively low stock positions. That is, the high turnover rates seem to reflect more of a low denominator rather than an extremely large trade *per se*. From the data, the mean stock position of a household in the T5 quintile with a turnover rate less than 400% is nearly 5 times larger than the mean stock position of a household with a turnover rate over 400%.

Interestingly, on average, nearly 80% of these trades in the tail are purchases rather than sales. This is consistent with the large turnover rate being driven by the relatively low denominator.

3 Household Optimization

Here we present the model of household optimization, drawing upon Bonaparte, Cooper, and Zhu (2012) and Cooper and Zhu (2015), that is the basis of the parameter estimation.[12] The household is infinitely lived and has two assets: bonds and stocks. Bonds are costless to adjust, and have a certain return. Stocks yield a random return, higher on average than bonds. Stock holdings, by assumption, are costly to adjust.

The key to the model is the household choice of whether to adjust its portfolio or not. Adjustment is costly due to the presence of fixed and variable trading costs. The household may choose not to incur these costs, in which case consumption smoothing is achieved solely through adjustment in bond holdings. If the household adjusts, then it incurs costs of portfolio adjustment. In this way, the model generates two types of turnover: the discrete choice of whether to adjust and the continuous choice of how much to adjust conditional on having incurred fixed adjustment costs.

The model is general enough to study portfolio rebalancing as well as the movement of wealth between liquid and illiquid assets. The restrictions used to study rebalancing and consumption smoothing separately, in order to address the key questions of the paper, are imposed as needed.

[12] An earlier version of the paper had only a single asset, as in Bonaparte and Cooper (2009). With that specification, portfolio adjustment and adjusting the margin between consumption and savings was not distinct.

3.1 Rational Household

Let $\Omega = (y, S, R^s)$ represent the state of the household where y is current labor income, $S = (b, s)$ is the current value of the holdings of liquid assets (bonds) and stocks respectively. At this level of generality, s is a vector of stocks and R^s is a vector of returns. The return on bonds, R^b is deterministic. A household chooses between (i) portfolio adjustment and (ii) no portfolio adjustment. This choice is given:

$$v(\Omega) = \max\{v^a(\Omega), v^n(\Omega)\} \tag{3}$$

for all Ω.

A household choosing to **adjust** selects the amount of stocks and bonds to solve:

$$v^a(\Omega) = \max_{b' \geq 0, s' \geq 0} u(c) + \beta E_{\Omega'|\Omega} v(\Omega')$$

$$s.t.$$

$$c = \psi y + R^b b + R^s s - b' - s' - F - C(s, s'). \tag{4}$$

In this problem, there is no borrowing and short sales of stocks are not allowed. Throughout, assume $u(c)$ is strictly increasing and strictly convex. For the empirical analysis, $u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$.

There is a time cost of stock adjustment represented by $\psi \leq 1$ in (4). Second, the model allows a fixed cost of adjustment, F. Further, the model includes direct trading costs, explained further below, captured by $C(s, s')$. In addition to the frequency of adjustment these costs also generate a demand for bonds and thus impact the stock share. These costs are important for matching household finance moments and for generating turnover.

If the household choses **not to adjust** its portfolio, then the trading and opportunity costs are avoided. There is re-optimization over bond holdings alone. The household chooses bonds to solve:

$$v^n(\Omega) = \max_{b' \geq 0} u(c) + \beta E_{\Omega'|\Omega} v(\Omega')$$

$$s.t.$$

$$c = y + R^b b - b' \tag{5}$$

$$s' = R^s s. \tag{6}$$

Here we assume that if there is no portfolio rebalancing, any return on stocks is automatically put into the stock account, i.e. $s' = R^s s$.

The policy functions generated by household optimization include an extensive margin (adjust, no adjust) and an intensive margin indicating the magnitude of the adjustment. Due to the adjustment costs, the model can produce both inaction in portfolio adjustment as well as large turnover rates.[13] The incentive for portfolio adjustment comes from large persistent shocks to income and returns.

Consider a large income shock, such as a job loss. A household will draw on both liquid and illiquid assets to smooth consumption. Because of the fixed portfolio adjustment cost, much of the consumption smoothing will occur through variations in liquid assets. But eventually the household will adjust the holdings of liquid assets by liquidating some of the stock holdings. Once the adjustment cost is paid, the household will choose to liquidate a large amount of stocks to avoid having to pay adjustment costs in the near future.

[13] Bonaparte, Cooper, and Zhu (2012) contains a lengthy presentation of the properties of the policy function resulting from a model with costly portfolio adjustment.

This produces an episode of large turnover. The trade costs themselves are partly measured, as discussed in Section 4.3, and thus produce a reduction in the net return of these traders compared to those who do not adjust. Likewise, a large positive return shock may create a large enough wealth gain that households choose to rebalance their portfolios. This would generate a positive correlation between gross returns and trades. Similarly, a large adverse return shock might also cause financial wealth to fall so that rebalancing is worthwhile. In this case, a negative correlation between trading and gross return is created.

A key element of the model are the fixed trading costs. These generate periods of inaction in portfolio adjustment, complemented with occasional large adjustment to rebalance the liquid and illiquid holdings. Further, these costs generate an incentive, despite the return differentials, to hold liquid assets. These model outcomes are matched, through the estimation, with the complementary data moments.

3.2 Irrational Households

In our setting, irrationality is limited, in various ways, to the manner in which conditional expectations are formed. We are not considering here the implications of, say, time inconsistencies nor intransitive preferences.

3.2.1 Misperceptions

Consider the following representation of beliefs of agents about the return on an individual stock j:

$$R_t^j = (1 - \tilde{\rho}_{R^j})\tilde{\mu}_{R^j} + \tilde{\rho}_{R^j} R_{t-1}^j + \varepsilon_t^{R^j} \tag{7}$$

where the the parameters $(\tilde{\mu}_R^j, \tilde{\rho}_R^j, \tilde{\sigma}_R^j)$, where $\tilde{\sigma}_R^j$ is the standard deviation of the innovation in the returns process of asset j. These are perceived parameters that may not coincide with the true process for returns. Indeed, our interest is in studying the relationship between beliefs and the true process for trading strategies and portfolio returns.

From (7), the specification permits three types of deviations through the: (i) mean, (ii) standard deviation and (iii) persistence of the return process. Beliefs about a positive serial correlation in the return captures the frequently noted belief in stock market “momentum”. In addition, for a model with multiple assets, there can be misperceptions about the covariance of returns.

3.2.2 Noisy Signals

Here we consider another form of irrational beliefs associated with a signal provided, say, by a financial advisor about future returns. From our specification of the stock return process, future returns are not predictable. But an agent may believe in a signal, such as the advice of an advisor, leading to excessive turnover and relatively low net returns.

To study this formally, assume there is an iid signal, denoted z, that the household believes is correlated with future returns. The discrete choice is again given by (3) and the options of adjustment and non-adjustment given by (4) and (5) respectively with the modified state vector of (Ω, z).

This choice is given by:

$$v(\Omega, z) = \max\{v^a(\Omega, z), v^n(\Omega, z)\} \tag{8}$$

for all (Ω, z). The revised options are:

$$v^a(\Omega, z) = \max_{b' \geq 0, s' \geq 0} u(c) + \beta E_{\Omega'|\Omega, z} \int_{z'} v(\Omega', z') dG(z')$$

$s.t.$

$$c = \psi y + R^b b + R^s s - b' - s' - C(s, s') - F \tag{9}$$

if adjustment. If the household does not adjust, it solves

$$v^n(\Omega, z) = \max_{b' \geq 0} u(c) + \beta E_{\Omega'|\Omega, z} \int_{z'} v(\Omega', z') dG(z')$$

$s.t.$

$$c = y + R^b b - b' \tag{10}$$

$$s' = R^s s. \tag{11}$$

In these expressions, $G(z')$ is the cdf of z'. The conditional expectation in these expressions highlights that the sole role of z is to provide information about Ω'.

As households believe z is informative about future returns, their decisions will depend on this random variable. This source of irrationality is similar to overconfidence about the serial correlation about returns but realizations of z only influence household beliefs, not their budget sets. This allows z to be, at least in the mind of the household, a predictor of future stock returns, $R^{s'}$.[14]

4 Simulated Method of Moments Estimation

The parameters, including perceptions, are estimated by matching moments. This section discusses the estimation approach and the moments that are used to identify the parameters.

4.1 Approach

Specifically, the estimation finds the vector of parameters that minimizes the distance between actual and simulated moments:

$$J = min_{\Theta} \left(M^s(\Theta) - M^d\right)' W \left(M^s(\Theta) - M^d\right). \tag{12}$$

Here $M^s(\Theta)$ are the simulated moments that depend on the parameter vector Θ, M^d are data moments and W is the identity matrix. Results with a weighting matrix are reported as well.

The parameter vector is $\Theta = (\beta, \gamma, \psi, F)$ in the case of rational households. In the estimation we study the opportunity and fixed cost cases separately. When the estimation allows for some form of irrationality, the parameter vector is supplemented to include household beliefs. Importantly, the estimation with irrational agents includes (β, γ, ψ, F) along with beliefs.

The analysis assumes the household choices are made on a monthly basis. The monthly model allows a direct link to the high frequency household account data.

Given a parameter vector, a simulated panel data set with 4000 households and 4000 time periods (after dropping the first 1000 periods) is created from the solution of the household's dynamic optimization

[14] Gervais and Odean (2001) study a related model where agents are learning about their ability in an environment with noisy signals. In their setting, agents update their beliefs about own ability, putting excessive weight on successes relative to failures.

problem. The simulated moments are calculated from this panel, just as in the actual data.[15] Households differ because of idiosyncratic income shocks which generates differences in trading patterns and returns. Trading costs are calculated from the observed trades using the estimates given in Table 2.

4.2 Moments

The moments used for the first part of the estimation are reported in the top row of Table 3.[16] Our approach is to estimate the key parameters by matching these household finance facts. Then the model is extended to study turnover rates.

The first moment is a measure of portfolio adjustment. For all stock market participants it is a bi-annual rate taken from the PSID. Over the sample period of 1999 to 2007, this is about 47%. The wealth to income ratio is the mean of average financial wealth over average income. For our sample, this is 2.43 from the SCF. Finally, the stock share is the median of the ratio of stocks to financial assets. This share is 0.684 in the data from the SCF.

Though all parameters influence all of these moments, it is possible to see some linkages. The cost of portfolio adjustment will impact the adjustment rate as well as the stock share. To understand the second effect, keep in mind that liquidation of stocks is costly which, all else the same reduces the stock share below the level it would obtain if there were only return differences between the assets. The impatience parameter has a direct impact on the wealth to income ratio. Finally, risk aversion will influence the desire for consumption smoothing and thus the frequency of adjustment. It also impacts the portfolio composition both in terms of a demand for liquidity and an aversion to the riskiness of stock holdings.

4.3 Trading Costs

The monthly household account data set from Barber and Odean (2000) is used to determine trading costs and thus calculate net returns at the individual level. It provides information on common stock trades of about 78,000 households through a discount brokerage firm from January 1991 to December 1996.[17]

Bonaparte and Cooper (2009) estimate trading costs, $C(\cdot)$, directly from this data set. Assume:

$$C^b(q, q_{-1}) = \nu_0^b + \nu_1^b(q - q_{-1}) + \nu_2^b(q - q_{-1})^2 \tag{13}$$

if the household buys an asset, $q > q_{-1}$. If instead the household sells, $q < q_{-1}$, then

$$C^s(q, q_{-1}) = \nu_0^s + \nu_1^s(q_{-1} - q) + \nu_2^s(q - q_{-1})^2. \tag{14}$$

Bonaparte and Cooper (2009) use this monthly household account data to estimate these parameters.[18] The trading costs, measured in dollars, are estimated in a regression where the dependent variable is the commission and the independent variables, denoted q, are trade value (the price of the share times the quantity of share) and trade value squared per stock.

[15] Our results do not change if the size of the simulated panel is increased. For these results, the (coarse) fine state space had $(20 \times 25 \times 10 \times 3)$ $100 \times 200 \times 10 \times 3$ elements. The solution entailed piecewise cubic hermite interpolation, with convergence for the value function defined on the fine grid.

[16] These moments are presented and discussed in detail in Bonaparte, Cooper, and Zhu (2012).

[17] The following section as well as the Appendix provide additional details about the data set and the calculation of these costs and moments.

[18] Details on the estimation can be found in Bonaparte and Cooper (2009). Through this procedure, we are able to decompose the commission costs reported in Table 1 of Barber and Odean (2000) into fixed and variable components.

Table 2: Estimated Trading Costs

Variables	(1)	(2)	(3)	(4)
		Sales		
Linear	141.735	142.788	141.735	142.788
	(0.19199)	(0.19056)	(3.83737)	(3.82560)
Quad.	-1.553	-1.590	-1.553	-1.590
	(0.01042)	(0.01033)	(0.59998)	(0.60397)
Constant	57.475	67.938	57.475	67.938
	(0.05724)	(0.21439)	(0.43417)	(0.53864)
Observations	1,329,394	1,329,394	1,329,394	1,329,394
R-squared	0.354	0.365	0.354	0.365
		Purchases		
Linear	117.539	118.319	117.539	118.319
	(0.15906)	(0.15770)	(8.32375)	(8.27930)
Quad.	-0.969	-0.977	-0.969	-0.977
	(0.00277)	(0.00275)	(0.17325)	(0.17562)
Constant	52.669	65.297	52.669	65.297
	(0.04717)	(0.18124)	(0.82009)	(0.81541)
Observations	1,746,403	1,746,403	1,746,403	1,746,403
R-squared	0.242	0.256	0.242	0.256
Year, Month FE	No	Yes	No	Yes
Account Cluster	No	No	Yes	Yes

This table presents regression results for (13) and (14). The trades values were divided by $100,000. There are controls for time (year and month) and an account cluster.

The estimates are reported in Table 2.[19] There are four specifications for each type of trading depending on the inclusion of month and account fixed effects. The estimated adjustment costs, particularly the fixed cost, depends on the monthly fixed effect while the estimates do not depend on the account fixed effect.

Though the linear and quadratic terms are statistically significant, the main cost of adjustment is the fixed cost per trade. While this cost may seem high relative to currently advertised fixed trading costs, it is still small compared to the average trade of a household in the data set of about $12,500. A purchase of that magnitude would entail a total cost of $75, with the fixed cost comprising almost 77% of the total cost.[20]

These estimates of trading costs do not include the bid-ask spread which, according to Barber and Odean (2000) are about 0.31% for purchases and 0.69% for sales. These additional costs are added to the linear terms reported in Table 2 when the trading costs are integrated into the household optimization problem.

4.4 Income and Returns

Solving the model requires inputting the stochastic processes for income and returns. For rational households these are taken from the data. From the irrational households, the parameters for the processes are estimated through the moments characterizing household choices.

[19] As suggested by a referee, it is also useful to control for both time (year and month) and household account fixed effects. The results reported in Bonaparte and Cooper (2009) are for the treatment without fixed effects for time and household. As demonstrated in Section 5.2 the key is the fixed cost.

[20] Using specification (1), this comes from $57.475 + 141.735 * (12500/100000) - 1.553 * ((12500/100000))^2$.

4.4.1 Income Process

The income process for stockholders is annual. It is estimated from the Panel Study of Income Dynamics (PSID). The serial correlation is 0.842 and the standard deviation of the innovation is 0.29.[21] Importantly, there is not sufficient information in the data set used by Barber and Odean (2000) to estimate the income process for individuals in that sample. Nor is it possible to extract a sample of households who directly own stock from the PSID to mimic those in the Barber and Odean (2000) sample.

As the frequency of the household choice problem is monthly, it is necessary to convert the annual income process to this higher frequency. This is done in two ways, distinguished by the presence of unemployment risk. The first simply converts the annual process into a monthly one without adding any higher frequency unemployment risk. In this case, the monthly serial correlation is 0.9781 and the standard deviation of the innovation to income is 0.1165. The second adds unemployment risk to the income process. As emphasized in Carroll (1992), it is important to recognize that, particularly, at the monthly frequency, households face significant risk of job loss. Thus, this second specification adds a zero labor income state to the process. These flows between employment and unemployment are taken from the Bureau of Labor Statistics.[22]

Specifically, each month an employed agent becomes unemployed with probability $\delta = 0.014$. Each month an unemployed agent finds a job with probability 0.27. The replacement rate for an unemployed agent is set at 40% of average income. If an unemployed agent finds a job, the wage is assumed to be the mean of the income process. Given these flows and the estimated annual serial correlation and standard deviation of the innovation at the annual level, the monthly income process is estimated through a simulated method of moments approach, discussed in the Appendix. From this analysis, the monthly serial correlation is estimated at 0.9959 and the standard deviation of the income innovation is estimated at 0.0839, conditional on employment.

4.4.2 Returns

The real stock return, which includes capital gains and dividends, is measured as the S&P index monthly return from 1967-94. It is taken from CRSP (`http://wrds-web.wharton.upenn.edu/wrds/index.cfm`). The average monthly return is set at 1.0094 (an annual return of 1.119) with a standard deviation of 0.0446. The estimated serial correlation of annual returns is not significantly different from zero.[23] We set the mean return of the liquid asset to yield an average equity premium of 3.5 percentage points.[24]

5 Liquidity Traders: The PSID HHs

This section addresses the question of the turnover pattern generated by a household holding stocks. Here the focus is on liquidity trade so that the portfolio is reduced to two assets, stocks and bonds. The optimization problem is specified in section 3.1 with the restriction to a single stock.

The analysis focuses on a typical household which holds stocks both indirectly and perhaps directly as well. This is not intended to be a household holding an account in a brokerage firm, as in the Barber

[21] This is the same process as used in Bonaparte, Cooper, and Zhu (2012) and is described in the Appendix of that paper. Variables are normalize by the mean income of 6000 dollars per month income taken from the Barber and Odean (2000) sample.

[22] Specifically, these probabilities characterizing this additional state are calculated from the seasonally adjusted flows taken from `https://www.bls.gov/webapps/legacy/cpsflowstab.htm` for 1990 to 2009.

[23] The point estimate is 0.045 with a standard error of 0.046.

[24] This is lower than the average return differential between stocks and bonds but is consistent with our interpretation of bonds as liquid assets.

and Odean (2000) data. Throughout we refer to this as a "PSID HH", as the estimation is for a PSID household holding stocks and bonds, both directly and indirectly. This is in contrast to a "BO HH", which is a household in the BO sample of direct holders.

Given the incomplete data on BO HH, we instead focusing on a more typical household where we study basic financial choices based on moments from the PSID and the SCF. These moments are completed with an income process estimated from the PSID. In this way, we have the ingredients and moments for the estimation and simulation of the choices by this type of household.

Given these parameters, household choices are evaluated relative to the turnover and net return moments based on the Barber and Odean (2000) data. This can be thought of as a counterfactual exercise where we place the PSID HH into the BO HH environment. In this way we can see how atypical is the behavior of the households in the BO sample.

5.1 Results

This sub-section reports results for model estimation based upon household financial moments alone. Building on this, estimation and simulation results including turnover and net trades are included as well.

5.1.1 Matching Household Finance Moments

Table 3: Data and Simulated Financial Moments

case	AR	WI	Sh	fit
Data	0.467	2.430	0.684	na
	Fixed Cost			
NO UR	0.370	2.534	0.552	0.082
UR	0.443	2.257	0.584	0.029
	Opportunity Cost			
NO UR	0.576	1.914	0.652	0.101
UR	0.581	1.972	0.626	0.103

This table reports moments from various estimations. "AR" is the (biannual) adjustment rate for the stockholders and the annual adjustment rate for the direct holders, "WI" is the wealth to income ratio and "Sh" is the stock share, "T5" is the turnover rate for quintile "5". "UR" refers to an income process allowing unemployment risk and "NO UR" excludes this risk.

The moments for the estimated models are presented in Table 3. There are two cases depending on the nature of the portfolio adjustment costs. The "Fixed Cost" case imposes a fixed cost of portfolio adjustment, i.e. $F \geq 0$, while the alternative allows an opportunity cost, i.e. of $\psi \leq 1$. These are both present in (4) but treated independently in the estimation.

This estimation exercises matches moments for all shareholders, not just those with direct holdings. Thus the trading costs estimated from the Barber and Odean (2000) sample are not included in the model for this part of the estimation exercise. Also, these moments are computed using the time series approach with a balanced panel.

A couple of findings emerge. First the best fitting models are those with a fixed rather than opportunity cost of trade. The best fitting opportunity cost model seems unable to match the relatively low biannual adjustment rate. Second, the additional risk of unemployment brings the models closer to the data. With this added variation in income, the adjustment rate is higher and so closer to the data compared to the estimated model without unemployment risk.

Table 4: Parameter Estimates

case	β	γ	F	ψ
Finance Moments				
	Fixed Cost			
NO UR	0.970	5.064	0.0002	na
UR	0.970	5.060	0.0002	na
	Opportunity Cost			
NO UR	0.973	4.015	na	0.828
UR	0.971	4.428	na	0.831
Finance + Turnover Moments				
	Fixed Cost			
NoUR	0.970	5.071	0.0002	na
Baseline: UR	0.964	5.107	0.0002	na
	Opportunity Cost			
NO UR	0.973	4.614	na	0.819
UR	0.972	4.391	na	0.830

This table reports parameter estimates of the discount factor, β, the curvature of utility γ, the fixed adjustment cost, F, and the proportional adjustment cost, ψ.

Table 4 reports parameter estimates. The top block reports the parameter estimates for the estimation matching the household finance moments only. The highlighted row is the one that matches the moments best. On an annual basis, the estimated discount factor is only 0.69, the estimated risk aversion is around 5, not outside of standard calibrations. The fixed adjustment cost F is about \$12, so only a small fraction of monthly income. But this additional cost constitutes an increase of nearly 20% beyond the estimated fixed component of the trading cost reported in Table 2.

Comparing these estimates to those reported in Bonaparte, Cooper, and Zhu (2012), the estimated discount factor is very close with a point estimate of 0.686 in that study, the risk aversion a bit lower. But there are two differences to keep in mind. First, the estimation in Bonaparte, Cooper, and Zhu (2012) included the elasticity of intertemporal substitution as a moment. Second, the decision period that model was a year. Here, the decision period is a month, allowing us to include higher frequency movements in employment status. Finally, that model focused on an adjustment cost proportional to income rather than the fixed cost that matches the moments best in this study.

5.1.2 Matching Turnover and Net Return Moments

Table 5: Financial and T5 Turnover Moment

case	AR	WI	Sh	T5	fit
Data	0.467	2.430	0.684	0.118	na
	Fixed Cost: Simulated				
HH Fin.	0.367	2.539	0.552	0.091	0.131
	Fixed Cost: Estimation				
NO UR	0.359	2.475	0.550	0.106	0.098
Baseline, UR	0.426	2.010	0.591	0.126	0.064
	Opportunity Cost: Estimation				
NO UR	0.618	2.334	0.600	0.226	1.045
UR	0.602	2.105	0.624	0.221	0.945

This table reports simulation and estimation results. "AR" is the (biannual) adjustment rate for the stockholders and the annual adjustment rate for the direct holders, "WI" is the wealth to income ratio and "Sh" is the stock share, "T5" is the average turnover rate for quintile 5. The HH Fin. row simulates using the parameter estimates for the best fit of the HH finance moments alone. The other cases are new estimation results.

Given the estimated household model from the PSID, we now study the net return and turnover moments produced by this type of household. Essentially this is a counterfactual exercise where we take the PSID

Table 6: All Turnover Moments

case	T1	T2	T3	T4	T5	T5(med)	DR
Data	0.000	0.002	0.013	0.029	0.118	0.071	-0.0006
Model	0.004	0.009	0.013	0.015	0.126	0.031	-0.0007

This table reports data and simulation results using the baseline parameters. For the turnover moments, "Ti" is the turnover rate for quintile "i" and "DR" is the difference in the net return between the highest and lowest turnover rate quintiles.

household and put them into the discount broker of Barber and Odean (2000). In this way, we can see if the turnover and net return moments can be well matched by a typical rational household.

The first block of Table 5 reports the simulation results for the model that best fit the financial moments, i.e. the specification with fixed costs and unemployment risk. For that parameterization, the turnover rate in the top quintile is 9%, indicating that a model parameterized to match standard household moments can generate large turnover rates.[25]

The remainder of Table 5 reports the moments from an estimation exercise in which the turnover rate in the fifth quintile (T5) was added as an additional moment. The model with this additional moment is over-identified. For this estimation, the trading costs reported in Table 2 were included.

Again, the model with fixed adjustment costs and unemployment risk matched the data moments best. That model is able to come quite close to the T5 turnover rate while still matching the basic household financial moments.

The models with opportunity rather than fixed costs both overstate inaction and large trades. Evidently, the adjustment cost must be high enough to match the share of stocks but that cost produces both inaction and large trades. Note with this parameterization, the model has no difficulty in generating a large turnover rate in T5, almost twice that in the data.

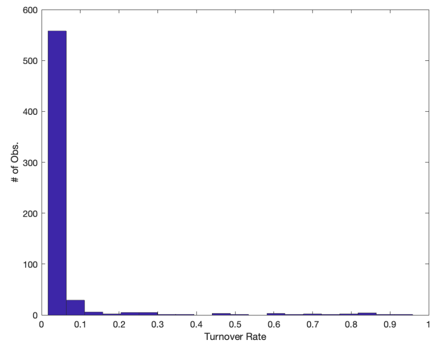
From the bottom part of Table 4, the parameters that best fit the model with turnover are quite close to those that fit just the household financial facts. The discount factor is slightly lower and the risk aversion slightly higher. Thus matching the net turnover moments requires only a modest deviation in parameter values. In the presentation that follows, we refer to these as the **baseline** estimates and moments

Table 6 goes further to look at all of the time series turnover moments. These are from a simulation of the model using the baseline parameters. The simulation matches quite well the near zero turnover in quintiles T1-T3. Importantly, the model matches the differential in net returns between T5 and T1 seen in the data. Though important for the discussion of turnover and return patterns, this net return differential was not among the moments we target.

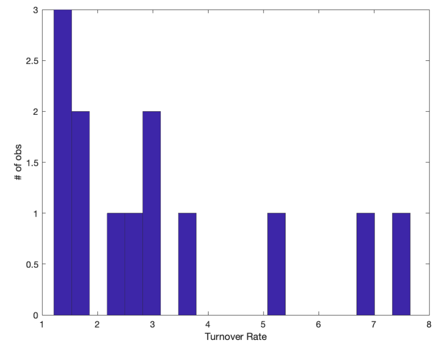
Figure 2 shows the distribution of turnover rates in T5. The left panel shows this distribution conditional on less than 100% turnover. This is about 98% of the observations in T5. There are almost 90% of the observations with less than a 10% turnover rate, lower than the 76% in this group in the data, shown in Figure 1. The right panel shows the distribution of turnover rates in excess of 100%. Further, as shown in Table 6 the median turnover is much lower in the simulated data compared to the actual data.

Before proceeding, it is useful to put the moments and the theory together to understand the mechanism that allows the model to match the high turnover rate in T5.[26] In the estimated model, the presence of the fixed adjustment costs generates three effects. First, it allows the model to match the inaction rate in the data. Without the adjustment costs, household would response to both income and return shocks and

[25]For this simulation we imposed the estimated trading costs, thus impacting the household finance moments.
[26]In a similar model, Bonaparte, Cooper, and Zhu (2012) provide graphical representations of the policy functions.



(a) T5: Turnover < 100%



(b) T5: Turnover ≥ 100%

Figure 2: T5 Turnover Distribution

there would be no inaction, aside from numerical approximation. Second, when the household does adjust its portfolio, the adjustment rate can be quite large since it is replenishing its liquid account infrequently. These bursts create the high turnover rates. Finally, the adjustment costs create a demand for liquidity so that the stock share is low enough to match data. This intuition is supported through the simulations reported in the robustness exercises, Section 5.2.

5.1.3 Irrational Agents

This sub-section introduces irrational agents into the estimation exercise matching the financial moments and the T5 turnover. We proposed a number of the models of irrational beliefs in sub-section 3.2. Here we explore in detail a leading cases of serial correlation in returns. This represents a form of overconfidence in that investors are "smarter" than the market. An agent believing in serially correlated returns will hold on to winning stocks and sell losers.[27]

Table 7: Parameter Estimates

case	β	γ	F	IR
Baseline	0.964	5.107	0.0002	na
$\tilde{\rho}_R$	0.982	3.011	0.0002	0.005

This table reports parameter estimates of the discount factor, β, the curvature of utility γ, the fixed adjustment cost, F, and and irrational beliefs, IR.

By construction, we are assuming that all agents hold the same beliefs, rather than estimating a fraction of irrational agents, so that we are adding only a single parameter.[28] This case could increase turnover and thus help to match the T5 moment. The challenge is whether it can do so without deteriorating the match of the household finance moments.

The estimation did cover evidence of irrationality. As shown in Table 7, there is a tiny amount of serial correlation in the return process, estimated at 0.005, that improves the model fit as reported in Table 8. If we simulated the model with these parameters and $\tilde{\rho}_R = 0$, the moments are not close to the data moments.

[27]We did not find evidence of other forms of irrationality with respect to the return process. For the case of noisy advice with a small probability of good news, the fit was essentially the same as the baseline.

[28]An alternative would be to fix some degree of irrationality and estimate the fraction of agents with those beliefs. It seems impossible to jointly identify beliefs and the fraction of agents holding them.

Table 8: Financial and T5 Turnover Moment

case	AR	WI	Sh	T5	Fit
Baseline	0.426	2.010	0.591	0.126	0.064
$\tilde{\rho}_R$	0.446	2.237	0.630	0.117	0.015

This table reports simulation results for two cases of irrational beliefs. "AR" is the (biannual) adjustment rate for the stockholders and the annual adjustment rate for the direct holders, "WI" is the wealth to income ratio and "Sh" is the stock share, "T5" is the turnover rate for quintile "5".

In particular, the T5 turnover rate is about one-fourth of that in the data. Clearly, the belief in the serial correlation of the return creates additional turnover so that the estimated model can match that and other moments at different parameter values.

Is this evidence sufficiently strong to argue that irrationality is needed to match the patterns of high turnover and low return? We think not. The estimated value of $\tilde{\rho}_R = 0.005$ is quite close to zero so that these are not really irrational expectations. As noted earlier, our point estimate of the serial correlation of returns was 0.045 with a standard error of 0.046.

There is another issue to consider: does this value of $\tilde{\rho}_R = 0.005$ point to irrational expectations or just the role of serially correlated returns in improving the model fit? If we simulate the model with rational expectations, thus allowing both the belief thatt $\rho_R = 0.005$ and the reality of the serial correlation to be the same, then we obtain the same moments as reported in Table 8. So one might interpret this as simply as an alternative parameterization, within a standard error of the estimated serial correlation of the return process, that leads to an estimated model which fits the data moments better.

5.2 Robustness

We performed two robustness exercises, both revolving around trading costs. For both of these exercises, these are simulation results at the baseline parameters, with fixed rather than opportunity costs and with unemployment risk.

Table 9: Data and Simulated Financial Moments: Robustness

case	AR	WI	Sh	T	fit
Baseline	0.426	2.010	0.591	0.126	0.064
Only Fixed Costs	0.428	2.000	0.591	0.107	0.067
Low Fixed Costs	0.425	1.99	0.590	0.135	0.081

This table reports simulation results for two robustness exercises. "AR" is the (biannual) adjustment rate for the stockholders and the annual adjustment rate for the direct holders, "WI" is the wealth to income ratio and "Sh" is the stock share, "T5" is the turnover rate for quintile . "UR" refers to an income process allowing unemployment risk and "NO UR" excludes this risk.

In the first experiment, labeled "Only Fixed Costs", we eliminated all but the fixed cost of trading from the estimates given in Table 2. This experiment allows us to isolate the effects of the linear and quadratic costs. As shown in the table, the turnover rate in T5 is a bit reduced by the elimination of the quadratic and linear adjustment costs. The other moments are relatively unaltered and the fit remains very good.

In the second experiment, labeled "Low Fixed Costs", we reduced the fixed costs of trading by 50% to illustrate how advances in the technology of trading, reflected in lower adjustment costs, would impact these moments. Here we see even more action in the high turnover category and a slightly deteriorated fit. One role of adjustment costs is to temper the response of agents to shocks.

6 Unbalanced Panel: Participation

Our analysis thus far has ignored an important dimension of the data: the participation choice of owning direct stocks or not. As established below, this margin is important for understanding trade patterns.

Further, it is quite relevant for evaluating our findings. The choice of participating as a direct holder is distinct from the choice of trading patterns. The former is a choice on the extensive margin while the latter is one on the intensive margin. Consistent with the evidence we have presented, it is conceivable that, conditional on being a direct shareholder, the trading choices are reflect the decisions of rational agents. Yet at the same time it might be that some agents are overconfident in their ability to trade stocks, motivating them to hold and manage them directly.[29] Barber and Odean (2000) find no significant difference in the returns for the average households in their sample compared to a market index.[30] But overconfident agents might see a return differential requiring their direct control of stocks.

This section has three components. The first presents empirical evidence looking at unbalanced panels to better understand the participation margin. The second extends the model to allow a participation choice.[31] The third outlines a framework to interpret the evidence.

6.1 Evidence

Table 10: Un-Balanced Panel

Value	T1	T2	T3	T4	T5	DQ
		Net Trades				
nobs	16550	24439	12220	12219	12220	-4330
ave_to	0	0.002	0.022	0.055	0.319	0.319
med_to	0	0	0.022	0.053	0.154	0.154
ave_GR	1.014	1.014	1.013	1.014	1.025	0.011
ave_NR	1.014	1.014	1.012	1.006	1.01	-0.004
med_NR	1.012	1.012	1.012	1.012	1.015	0.003
ave_exit	0.659	0.606	0.546	0.627	0.647	-0.013

This table reports data moments using the time series approach, for the unbalanced panel. These are the averages across households of the average turn over rates and returns during their period in the sample. Here "Ti" is the turnover rate for quintile "i" and "DQ" is the difference between the highest and lowest turnover rate quintiles.

Turnover rates and the return differential for the unbalanced panel, allowing traders to leave the sample, are presented in Table 10. The patterns are quite different compared to the balanced panel. This panel has almost 56,000 observations, about 4 times larger than the balanced panel, indicating the importance of entry and exit.[32] Looking at net trades, the turnover rate in the T5 cell is nearly 3 times that of the balanced panel.[33] Thus it seems that a large fraction of the high turnover rates and negative net return differential reported in Barber and Odean (2000) reflected decisions on the extensive rather than intensive margin.

A second piece of evidence comes from the exit behavior of traders in the Barber and Odean (2000) sample.[34] Consider a logit model of the dependence of the exit decision on: (i) how long the household has been in the sample (tenure), (ii) beginning of the month position, (iii) individual-month net return, or gross

[29] Even the evidence on turnover needs to take into account the participation choice. This is noted in Odean (1998) without any corrections for selection.

[30] Specifically, from their Table II and related discussion, the average return of the households in the sample is not statistically different from a NYSE/AMEX/Nasdaq value weighted index.

[31] There is essentially no literature to draw on for this choice.

[32] Comparing to the sample size in Barber and Odean (2000), it is clear that they studied the unbalanced panel.

[33] As shown in Appendix subsection 8.3, the average turnover for net trades rises to almost 32%.

[34] The distinction between gross and net trades is immaterial for these regressions.

Table 11: Exit Regression Results

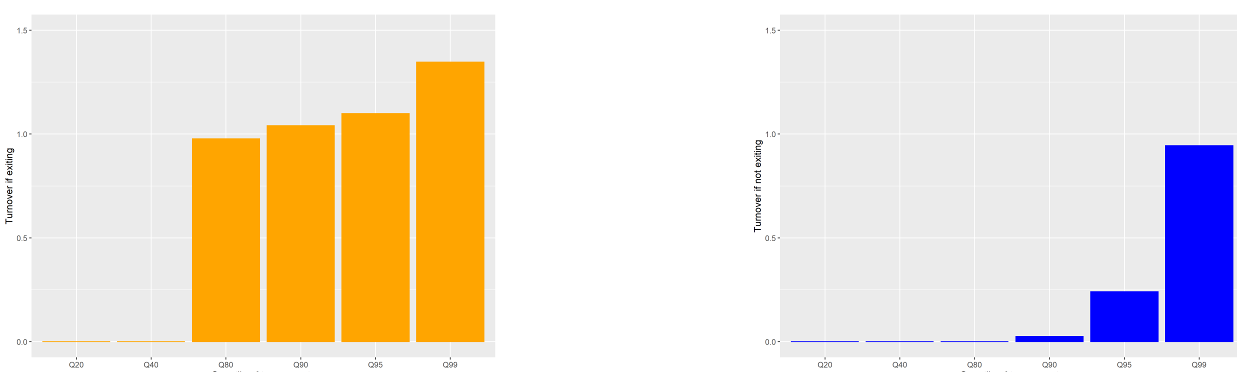
variable	coef.	std. error	coef.	std. error
constant	-1.58	0.012	-1.238	0.013
tenure	-0.058	0.0002	-0.062	0.0003
turnover	2.05e-04	9.874e-05	na	na
position	1.455e-07	3.075e-08	1.586E-07	3.30E-08
net return	na	na	-0.001	0.0004

Results from logit estimation with exit as dependent variable.

return and (iv) the turnover rate. As they are highly correlated, one specification includes turnover and the other net return. As shown in Table 11, the coefficient on tenure is always negative and the monthly position are positive: exit is more likely if the position is large and the account is relatively new. The coefficient on turnover is positive and significant and that on the net return is negative and significant. The probability of exit is decreasing in the net return and increasing in turnover. So, all else the same, the households experiencing the motivating patterns of high turnover and low net return are the most likely to exit. These effects are statistically significant. Given the very smalll coefficients on turnover and net return, the response of the odds ratio to a unit increase in them is about 1.

Figure 3 shows the distribution of turnover by exit status. For those exiting, left panel, turnover is near zero up to the 80th quantile, but averages over 100% for the 80th quantile and above.[35] In contrast, for those not exiting, right panel, the 80th percentile of the (net) turnover rate distribution within a given month is 0 and the large turnover does not occur until the 99th quantile. In other words, if a household is remaining in the market, it is likely to be inactive.

Finally, recall the turnover in the T5 quintile. From Figure 1, the bottom right panel illustrates the 5% tail of the T5 distribution. Compared to the balanced panel, there are more than 10 times the observations in the 100 to 200% turnover rate.



(a) Turnover if Exiting (b) Turnover if Not Exiting

Figure 3: Turnover By Exit

6.2 Extended Model

To study exit, the optimization problem is modified to include a fixed flow cost of **direct** stock market participation, presented as a fraction of average income, denoted Γ. Thus the budget constraint for a

[35] Some of the exits with near zero turnover can arise if the account closes without any reported trades.

participating, adjusting household is given by:

$$c = \psi y + R^b b + R^s s - b' - s' - C(s, s') - F - \Gamma \bar{y}. \tag{15}$$

If the participating household does not adjust its portfolio it must still incur the flow cost to maintain an account and thus the option for future adjustment. In this case consumption becomes

$$c = y + R^b b - b' - \Gamma \bar{y}. \tag{16}$$

A household can choose to sell out of this account to avoid the fixed costs. All assets are converted into bonds. There is an option to reenter the stock market as an indirect participant. An indirect stockholder solves the optimization problem without the costs of trading that are specific to direct holders and with $\Gamma = 0$.

Given that direct participation is more expensive without a gain in financial return, why is that option ever chosen?[36] One possibility is that individuals enjoy the control over their portfolio. In the model this might be viewed as either trading giving direct utility, i.e. $\psi > 1$ or $F < 0$, or participation itself generating direct value, as if $\Gamma < 0$.

Another is that irrational agents with erroneous beliefs about the underlying stochastic processes will believe that they can obtain a higher payoff from direct holdings compared to a rational agent. The irrationality could take the form of a perceived high serial correlation in returns or the existence of a signal that impacts portfolio choices. The key is that by holding stocks directly, these agents are able to act on their beliefs. In contrast, with indirect holdings, the portfolio is professionally managed and thus out of the control of these agents.

6.3 Empirical Implications

The model allows two competing hypotheses for direct participation: (i) tastes and (ii) beliefs. It is beyond the scope of this paper, with its focus on the intensive margin, to formally evaluate these alternatives. Instead we take this opportunity to informally evaluate them.

For the first explanation, the case of $\Gamma < 0$, so that there is a direct benefit of direct ownership, is of particular interest. It would be consistent with two pieces of evidence. First, individuals hold stocks directly. Second, conditional on direct ownership, their trading behavior is not that different from those not participating directly. The problem with this approach is that it would have difficulty explaining the exit patterns.

As for beliefs, overconfident agents, believing they have some insights in the returns through their signal, have an incentive to directly hold stocks. This is the case even if there are no differences in the distributions of returns, consistent with the aforementioned evidence of Barber and Odean (2000). Given their decision to hold assets directly, these agents use this market to meeting liquidity needs as well as betting on stocks. That is, they are not direct asset holders as a means of providing liqudity. Rather, given their presence in these markets, these are also traded to meet liquidity needs. This irrationality may drive some of the extreme observations in the T5 cell. But for the most part, the decision to hold stocks directly is impacted much more by the irrationality compared to the trading strategy given participation.

As for the evidence on exit, this points to some form of learning from experience is important in under-

[36] We have not found an answer in the literature.

standing the exit choices summarized by the above regressions. Agents who believe strongly in their ability to receive informative signals about returns will have an incentive to directly hold stocks. As formalized in Gervais and Odean (2001), overtime, these agents will respond to what they believe are informative signals allowing them to believe they can beat the market. In reality, this is not the case. Responding to the signals they receive contributes, along with liquidity trades, to the observed patterns of turnover and net return differentials. Eventually, many of these agents realize from their experience and perhaps given their reluctance to confront reality, that in fact they are not beating the market. This is what leads these agents to close their accounts. This is consistent with the pattern of first entry and then exit from direct holdings.

Bernile, Bonaparte, and Delikouras (2023) offer an alternative theory and evidence about overconfident agents. They measure overconfidence through a measure of agents beliefs that they will beat the market. These agents are natural candidates for direct stock holding. But, in contrast to Gervais and Odean (2001), Bernile, Bonaparte, and Delikouras (2023) provide evidence that overconfident agents will put excessive weight on failures rather than successes. If so, these agents are again likely to exit the status of direct stock holders.

7 Conclusion

The paper assesses the claim of Barber and Odean (2000) that observed patterns of large turnover and low net returns were indicative of overconfident households. The challenge was to provide a clear rational alternative to what Barber and Odean (2000) concluded had to be irrational.[37]

Due to data limitations, we are unable to study the households in the Barber and Odean (2000) sample directly. Our indirect approach estimates the preferences of a rational household using standard household finance moments. We construct counterfactuals based on that estimated model. If we focus solely on net trades, thus excluding portfolio rebalancing, with a balanced panel, we still observe the presence of trades with large turnover and low net returns. We argue that we can match these patterns with our model of the rational household. The key is the presence of portfolio adjustment which, along with stochastic income and returns, that generates trading pattern seen in the data. Adding forms of irrationality can only slightly improve the fit of this model.

We also studied the extensive margin choices of exit and participation. The Barber and Odean (2000) data included households who exited. We find that the turnover rates are considerably larger in the unbalanced panel. Further, household's exit decisions are impacted by turnover and net returns: the likelihood of exit is increasing in turnover and decreasing in the net return. Taking versions of these models, perhaps with learning, to match both the participation and intensive margin in unbalanced panel, is left for further study.

References

Barber, B., and T. Odean (2000): "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *The Journal of Finance*, 55(2), 773–806.

Bernile, G., Y. Bonaparte, and S. Delikouras (2023): "Stock Market Experience and Investor Overconfidence: Prior Beliefs or Learning?," mimeo.

[37]We are grateful to a referee for this phrasing.

Bonaparte, Y., and R. Cooper (2009): "Costly Portfolio Adjustment," NBER Working Paper # 15227, revised 2010.

Bonaparte, Y., R. Cooper, and G. Zhu (2012): "Consumption Smoothing and Portfolio Rebalancing: The Effects of Adjustment Costs," *Journal of Monetary Economics*, 59, 751–68.

Carroll, C. D. (1992): "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence," *Brookings Papers on Economic Activity*, 23(1992-2), 61–156.

Cooper, R., and G. Zhu (2015): "Household Finance: Education, Permanent Income and Portfolio Choice," *Review of Economic Dynamics*, 20, 63–89.

Gervais, S., and T. Odean (2001): "Learning to be overconfident," *Review of Financial Studies*, 14(1), 1.

Odean, T. (1998): "Are investors reluctant to realize their losses?," *The Journal of finance*, 53(5), 1775–1798.

8 Appendix

8.1 Income

Notation

The following uniform notation is used throughout this note. For income, denote the absolute level by Y, the log by y, and the residuals by $\tilde{y}$. The superscript a indicates annual and m indicates monthly data.

Annual Income

From Bonaparte, Cooper, and Zhu (2012), henceforth BCZ, the annual income process for stock holders obtained by the following steps (as described in the BCZ appendix): pool all the observations together, regress (log) income on age, age^2, education attainment, gender and marital status. Take the residuals from the regression and use the residuals to run the AR(1) process

$$\tilde{y}_t^a = \rho^a \tilde{y}_{t-1}^a + \epsilon_t^a. \tag{17}$$

The persistence of the income shock is estimated to be $\rho^a = 0.84224$, and the standard deviation of the innovation is $\sigma_\epsilon = 0.29027$.

Monthly Income

From Bureau of Labor Statistics CPS survey, `https://www.bls.gov/webapps/legacy/cpsflowstab.htm`, data on flows and levels are obtained to calculate the monthly probability of being separated from employment, the probability of finding a job conditional on being unemployed. Our definition of unemployment includes unemployment and not in the labor force. The probabilities are calculated from averaging over monthly flows over stocks. For example, the probability of being unemployed is $\frac{flow(e \to u + e \to n)}{employmentlevel}$.

We make the following assumptions on the monthly income process. Suppose a person was unemployed last month, then with prob p_{ue}, he is employed this month, and receives the average income, 6.[38] In logs,

[38] This corresponds to the monthly average income of \$6000, annual \$72000 for stock holders.

$y^m_{ue} = log(6)$. With prob $1 - p_{ue}$, this person is still unemployed, he receives the unemployed benefit, with is 0.4 times the average monthly income, so this translates to $y^m_{uu} = log(0.4 \times 6)$.

Suppose a person was employed last month, then with probability p_{ee}, he is still employed this month and receives

$$\tilde{y}^m_t = \rho^m \tilde{y}^m_{t-1} + \epsilon^m_t. \tag{18}$$

In levels, this corresponds to

$$y^m_t = \mu_m(1 - \rho^m) + \rho^m y^m_{t-1} + \epsilon^m_t \tag{19}$$

Since we assume that the innovation ϵ_t follows normal distribution of mean 0 and standard deviation σ_ϵ, Y^m follows log normal distribution of parameters (μ_m, σ^m_y), where $\sigma^2_y = \frac{\sigma^2_\epsilon}{1-\rho^2}$. Hence,

$$6 \equiv \mathbb{E}Y^m = e^{\mu_m + \frac{\sigma^{m2}_y}{2}} \rightarrow \mu^m = log(6) - \frac{\sigma^{m2}_y}{2} \tag{20}$$

With prob $1 - p_{ee}$, this person gets unemployed and get $y^m_{eu} = log(0.4 \times 6)$.

8.1.1 Estimation

Given these momthly flows into and out of unemployment, it is necessary to estimate the parameters, (ρ^m, σ^m_ϵ), of the monthly income process. This is essentially a SMM exercise with the annual parameters, $(\rho^a, \sigma^a_\epsilon)$ as moments to match. The following algorithm was used for this purpose:

1. guess a vector of parameters ρ^m, σ^m_ϵ
2. simulate a panel of level of monthly incomes, aggregate to a panel of annual incomes (level)

$$y^a = log(\sum_{i=1}^{12} exp(y^m_i)) \tag{21}$$

3. use the simulated annual income panel to do the following AR(1) regression

$$y^a_t = \mu_a(1 - \rho^a) + \rho^a y^a_{t-1} + \epsilon^a_t \tag{22}$$

   and calculate the variance of the residual.

4. compare the ρ^a and σ^a_ϵ with the estimates from BCZ and go back to step 1 if not close enough.

As indicated in the text, the serial correlation of the monthly income process conditional on employment is estimated to be 0.9959, and the standard deviation of the innovation is 0.0839 under the mean recovery case. For the income process without unemployment risk, we go through the procedures described above, except that in Step 2, we simulate a panel of monthly incomes assigning 0 probability to unemployment. For the case without unemployment risk, the serial correlation is estimated to be 0.9781, and the standard deviation of the innovation is 0.1165.

8.2 Data Construction

We discuss the data processing assumptions here.

Sample The time period of the data used in Barber and Odean (2000) ("BO data" here after) is from January 1991 to December 1996. To minimize assumptions regarding the first and last month, we computed our moments based on data between February 1991 and November 1996. It is at the monthly frequency by individual and security.

Multiple Entries For the same individual and security, we sometimes observe multiple records in the same month. When multiple records occur, sometimes they are identical duplicates, sometimes they exactly cancel out, but there are also incidents when the records are different, and it is unclear which one is the right record to use. For identical duplicates, we only keep one record, but for individuals that have different duplicates, we exclude these individuals from our analysis (including the other months without conflicting records). Third, when we merge return data onto trade and position data by cusip, there are observations without merged return. These observations are excluded from our analysis.

Re-entries In BO data, sometimes we observe disconnections in an individual's profiles. For example, one individual would have position and trade records until month t, and no records between t and t + k but records appear again after k periods. Based on the absence of trade records in month t, it is unclear whether the individual exited and re-entered, or it is simply missing data. Therefore, when we exclude re-entries, we stop using those individuals' records after the first time their records stopped. When we include "in-entries", we include their records after their records re-appeared.

Threshold In BO data, sometimes we observe very small positions (including ones that are below 1)$thatleadtolargeturnover$ dollars on the beginning of the month total position, and only include these records with a position above 60$inthecalculation.Evenwhenwedon'timposethresholds(oralternatively,thethresholdis$0) we discard records with nonpositive positions, which we do observe in the data.

Balanced vs. unbalanced panel . For every individual, we count the number of months that this individual has records. Individuals with 70 months of records (i.e., there is a record every month during the data periods) constitutes the balanced panel.

8.2.1 Data Sets Used

The first step is to merge pieces of information in each data set together. We first start with the description of each raw data set that was used.

1. newtrades3.dta: This data set documents the trade information of each account number. Notice that each household can have multiple accounts. In each record, there is information of trade date *TraD*, account number *AccN*, security number *SecN*(each account can trade various securities), type of trade *BS*, the commission charged *Commision*, Price, Quantiy and Principle (it always hold that $Prin = Quan \times Price$) Cusip, and product code *Proco* indicating the type of the account. As in BO, we are restricting attention to common stocks. A trade is only reported when it happens. In other words, if a household did not trade in month t, then there will be no report.

2. newposit91(92,...96).dta: These data sets document the position information for each security *SecNum* at each month*PosDat* under each account. In each entry there is information on product code *ProdCod*, position equity *PosiEq* and position quantity *PosQuan*. We merge all these years of position records by

simply appending newposit91 to newposit96, and generate positall.dta. The position data should be documenting the end-of-the-month position since that is consistent with the trade record. For example, for SecN 1328808 under AccN 629, the first record of position is equal to 300 in quantity in May 1992 when a quantity of 300 is purchased. For the following months (until December 1992) the position quantity was always 300 though the position equity changed with price changes until October 1995 when there is no position, with a quantity of 300 being sold.

3. newbase.dta: This data set has the household ID *HHN* and account ID *AccN*, as long as the date on which the household opened the account *HHOD*, account types, Client segment *CliSeg* which tells us whether the account is general **G**, affluent **A** or active **T** Since the position data only have account number but we want to know how long the household was in the sample to construct the balanced panel, we merge AccN with HHN. One HHN can have multiple AccN's. There are 77927 households from the raw data, and on average each *HHN* has 2.03 accounts with minimal of 1 and maximum of 392. Dofile called "genkeeplisthhod.do" generates a keep list that has 1 for households whose latest account was opened before or in February 1st 1991 and 0 otherwise. In other words, this keep list selects households who opened all their accounts before 1991 February.

The following are generated from the raw data described above.

- "baseAccNunique.dta": is generated from newbase.dta after dropping the same account numbers that belong to different households. Out of 158034 records of *HHN-AccN* pairs, there are 416 pairs where the same account number belongs to different households. We dropped these.
- "SecN_Cusip.dta": Starting from positall.dta, we only keep Security number and Cuisp and take its first 8 digits. It turns out the 77.3% of securities have Cusip.
- "cusip_initialvalue.dta": Since the position data is end-of-month, we need to generate a beginning of the month position data by taking the end-of-month position of last month. What we do here is to start with positall.dta, only keeping the common stocks, and generate a trade month one after the record. For example, if the position record is for May 1995 then the trade date generated is June 1995. Rename the position equity variable as "Posi_inital". There are 0.73% observations that have multiple reports for a security in one month. We dropped all the multiple records and only keep these records that have one unique report for the position of a security of an account in one month.[39] Then we match *SecN* with *Cusip* using "SecN_Cusip.dta" and only keep the matched ones. (13,335,790 were matched while 332,985 were not, in the master data)
- "myreturn.dta" is created from CRSP return data. The variable called "ncusip" is the correct cusip to use from the CRSP data, as well as "ret" for return.[40] The sample period is from December 1991 to November 1996.
- "nm0927.dta" is generated from positall.dta. First match that with *HHN* and only keep common stocks. For each household, count how many months there are at least one report of positions under one account for any security. The list assigns 1 to households who have at least 1 reported position for all the 71 months (Jan 1991 to Nov 1996) and 0 otherwise. Out of 65,514 observations, 20.78% of them always reported a position each month, On average, households have reports of 43.7 months.

[39] This leaves us with 13,668,775 observations.

[40] webpage http://www.crsp.com/products/documentation/monthly-data-items-0 gives detailed definition of the variables. Ret: Month-end to month-end change in total investment of a security, with ordinary dividends reinvested at the month-end.

8.2.2 Merging Trade and Position Data

First, starting with newtrades3.dta, we only keep product codes that are common stocks. We generate the trade month *traM* from the trading date information *TraD*. Then by each trade month *traM* and each security *SecN* under each account *AccN*, we add up all the purcases (sales) for quantity, principal and commission. This is to make the trades at the monthly basis because there can be multiple purchases (sales) within the same month. Reshape the data so that each row has the purchase principal, commission as well as the sales principal and commission. This data then is saved to "data0712.dta".
We merge position data with trade data by matching the trade data into the position data by *AccN*, *SecN* and *traM*. For the unmatched position data, we understand that there is no trade, so we replace the missing values with 0 for the purchase and sales principal. This merged data set is then saved as "merge_final0820.dta"
Working from "merge_final0820.dta", we first match *AccN* with *HHN* for each household using baseAccNunique.dta. Then we match each security with its cusip by "SecN_Cusip.dta" because the return data is based on cusip, not the security number. We only keep the records where the securities have a cusip. [41]
After this, merge the data through cusip with the return data to obtain monthly return for each security. Last, we merge this through cusip with the record of the position of last month, and only keep the records that are matched (in other words, those that has a position information of last month). Now we are ready to move on and calculate important statistics.

- *nettrade* of each security under each account in each month: is defined as the sum of purchases minus sum of sales of this security
- *mycost* is computed for each security under each account in each month plugging the *nettrade* computed above into the quadratic cost function (Table 1 in the paper). **Bid-ask spreads of 0.31% for purchases and 0.69% for sales are added to the linear term.**
- *port* is the total sum of equity across all securities in all accounts for each *HHN* in each month
- *totalB* is the total sum of purchases principal across all securities in all accounts for each *HHN* in each month
- *totalS* is the total sum of sales principal across all securities in all accounts for each *HHN* in each month
- *portlastm* is the total sum of equity at the end of last month *Posieq_initial* across all securities in all accounts for each *HHN* in each month
- *totalnet* is *totalB* minus *totalS*
- *sec_share* is the share of each security's equity *posi_initial* in *portlastm*, the total position for each household *HHN* at the beginning of each month
- *GRhht* is the weighted average of gross return (*ret* plus 1) by the share of equity of each security for each *HHN* total position at the beginning of each month: the weight is same as $p_{i,t}$ in Equation (2) in Barber and Odean (2000).
- *ourNRhht*: first, for each security under each AccN in each month, compute net return as $\frac{Posi_{initial} * (ret+1) - mycost}{Posi_{initial}}$. Then calculate the weighted average of the net return for each security using the same weight as in *GRhht*.

[41] This leaves us with 14,357,599 observations v.s. the 354,376 observations unmatched and deleted.

- *ourto*: our measure of turnover. It is equal to the absolute value of net trade *totalnet* across all securities for each *HHN*, divided by the total position at the beginning of the month: $\frac{|totalnet|}{Posi_{initial}}$

With all the calculations above, the data is saved as ”tohhlist_wretnew0928.dta”. This is the basis of generating cross section and time series panel.

8.3 Turnover and Return Tables

This section presents the summary tables for trade turnover and net return. One subsection looks at net trades, the other at gross. Within each sub-section, there are tables that treat reentry differently when looking at the balanced panel. Finally, within each table there are differences in the position required to be in the sample. For some the positionm labeled "threshhold" in the tables, is positive for others it must exceed $60.00.

8.3.1 Net Trades

Table 12: Net Trade Exclude Reentry

threshold	Value	T1	T2	T3	T4	T5	DQ
0	nobs B	2924	5310	2654	2655	2655	-269
0	ave_to B	0	0.002	0.013	0.03	0.118	0.118
0	med_to B	0	0	0.013	0.029	0.071	0.071
0	ave_GR B	1.013	1.014	1.014	1.014	1.015	0.001
0	ave_NR B	1.013	1.013	1.013	1.013	1.013	-0.001
0	med_NR B	1.013	1.013	1.013	1.013	1.013	0
0	ave_exit B	0	0	0	0	0	0
0	nobs UB	16073	22999	11500	11499	11500	-4573
0	ave_to UB	0	0.001	0.02	0.053	0.316	0.316
0	med_to UB	0	0	0.02	0.051	0.164	0.164
0	ave_GR UB	1.015	1.014	1.013	1.014	1.035	0.02
0	ave_NR UB	1.015	1.014	1.012	1.003	1.021	0.006
0	med_NR UB	1.012	1.012	1.012	1.012	1.017	0.004
0	ave_exit UB	0.692	0.633	0.551	0.658	0.747	0.055
60	nobs B	2915	5306	2653	2653	2653	-262
60	ave_to B	0	0.002	0.014	0.03	0.114	0.114
60	med_to B	0	0	0.013	0.029	0.072	0.072
60	ave_GR B	1.013	1.013	1.014	1.013	1.014	0.001
60	ave_NR B	1.013	1.013	1.013	1.013	1.013	0
60	med_NR B	1.013	1.013	1.013	1.013	1.013	0
60	ave_exit B	0	0	0	0	0	0
60	nobs UB	15995	22961	11481	11480	11481	-4514
60	ave_to UB	0	0.001	0.02	0.053	0.3	0.3
60	med_to UB	0	0	0.02	0.051	0.165	0.165
60	ave_GR UB	1.014	1.013	1.013	1.014	1.034	0.021
60	ave_NR UB	1.014	1.013	1.012	1.013	1.029	0.015
60	med_NR UB	1.012	1.012	1.012	1.012	1.017	0.004
60	ave_exit UB	0.69	0.631	0.55	0.655	0.741	0.051

This table shows trading patterns for Net Trades, Excluding Re-entry

Table 13: Net Trade Include Re-entry

threshold	Value	T1	T2	T3	T4	T5	DQ
0	nobs B	2924	5310	2654	2655	2655	-269
0	ave_to B	0	0.002	0.013	0.03	0.118	0.118
0	med_to B	0	0	0.013	0.029	0.071	0.071
0	ave_GR B	1.013	1.014	1.014	1.014	1.015	0.001
0	ave_NR B	1.013	1.013	1.013	1.013	1.013	-0.001
0	med_NR B	1.013	1.013	1.013	1.013	1.013	0
0	ave_exit B	0	0	0	0	0	0
0	nobs UB	16550	24439	12220	12219	12220	-4330
0	ave_to UB	0	0.002	0.022	0.055	0.319	0.319
0	med_to UB	0	0	0.022	0.053	0.154	0.154
0	ave_GR UB	1.014	1.014	1.013	1.014	1.025	0.011
0	ave_NR UB	1.014	1.014	1.012	1.006	1.01	-0.004
0	med_NR UB	1.012	1.012	1.012	1.012	1.015	0.003
0	ave_exit UB	0.659	0.606	0.546	0.627	0.647	-0.013
60	nobs B	2915	5306	2653	2653	2653	-262
60	ave_to B	0	0.002	0.014	0.03	0.114	0.114
60	med_to B	0	0	0.013	0.029	0.072	0.072
60	ave_GR B	1.013	1.013	1.014	1.013	1.014	0.001
60	ave_NR B	1.013	1.013	1.013	1.013	1.013	0
60	med_NR B	1.013	1.013	1.013	1.013	1.013	0
60	ave_exit B	0	0	0	0	0	0
60	nobs UB	16477	24405	12203	12202	12203	-4274
60	ave_to UB	0	0.002	0.022	0.056	0.264	0.264
60	med_to UB	0	0	0.022	0.054	0.156	0.156
60	ave_GR UB	1.013	1.013	1.012	1.014	1.024	0.011
60	ave_NR UB	1.013	1.013	1.012	1.013	1.02	0.007
60	med_NR UB	1.012	1.012	1.012	1.012	1.015	0.003
60	ave_exit UB	0.658	0.604	0.546	0.624	0.642	-0.016

This table shows trading patterns for Net Trades, Including Re-entry

8.3.2 Gross Trades

The following two tables present moments based upon gross trades. The moments computed in Barber and Odean (2000) include rebalancing, and not just net flows between stocks and bonds. From their Table V, monthly turnover is the average of sales and purchases so that portfolio rebalancing is included as well. Sales and purchases are treated slightly differently in terms of timing.

For a sale, turnover of asset j for household i in period t is

$$T^s_{i,j,t} \equiv s_{i,j,t} - s_{i,j,t-1} R^s_{jt}.^{42} \tag{23}$$

Using this, the sales turnover is given by

$$T^s_{i,t} \equiv |(\frac{\sum_j |T^s_{i,j,t}|}{s_{i,t-1} R^s_t})|. \tag{24}$$

For purchases, the timing is a bit different. For turnover in period t associated with a sale, Barber and Odean (2000) uses lagged purchases:

$$T^p_{i,j,t} \equiv s_{i,j,t-1} - s_{i,j,t-2} R^s_{jt-1}. \tag{25}$$

The purchase turnover rate is given by

$$T^p_{i,t} \equiv |(\frac{\sum_j |T^p_{i,j,t}|}{s_{i,t-1} R^s_t})|. \tag{26}$$

With these measurements, the overall turnover in period t for trader i is:

$$T_{i,t} = .5 * (T^s_{i,t} + T^p_{i,t}). \tag{27}$$

The net return for household i, denoted R^n_i, **on the stock portfolio** is

$$R^n_{i,t} = \frac{s_{i,t-1} R^s_t - C(s_{i,t}, s_{i,t-1} R^s_t)}{s_{i,t-1}}. \tag{28}$$

Here the cost function comes from the results in Table 2. In calculating the net return costs of trade due to the bid and ask spread are included as well. Note that this is not the net return on an individual trade but rather the net return on the entire stock portfolio. In our setting, it is impossible to compute the return on a particular trade without imposing some arbitrary accounting rule to assign trading costs to net returns of a particular purchase or sale.

[42] In the model, these are beginning of period values.

Table 14: Gross trade exclude reentry

threshold	Value	T1	T2	T3	T4	T5	DQ
0	nobs B	2922	5310	2654	2655	2655	-267
0	ave_to B	0	0.001	0.007	0.015	0.073	0.073
0	med_to B	0	0	0.007	0.015	0.038	0.038
0	ave_GR B	1.013	1.013	1.014	1.014	1.014	0.001
0	ave_NR B	1.013	1.013	1.013	1.013	1.013	-0.001
0	med_NR B	1.013	1.013	1.013	1.013	1.013	0
0	ave_exit B	0	0	0	0	0	0
0	nobs UB	11349	16768	8383	8384	8384	-2965
0	ave_to UB	0	0.001	0.009	0.023	0.138	0.138
0	med_to UB	0	0	0.009	0.022	0.068	0.068
0	ave_GR UB	1.016	1.015	1.013	1.015	1.031	0.015
0	ave_NR UB	1.016	1.015	1.012	1	1.017	0.001
0	med_NR UB	1.013	1.012	1.012	1.012	1.016	0.003
0	ave_exit UB	0.705	0.638	0.556	0.644	0.705	-0.001
60	nobs B	2914	5306	2653	2653	2653	-261
60	ave_to B	0	0.001	0.007	0.015	0.073	0.073
60	med_to B	0	0	0.007	0.015	0.038	0.038
60	ave_GR B	1.013	1.013	1.014	1.013	1.014	0.001
60	ave_NR B	1.013	1.013	1.013	1.013	1.013	0
60	med_NR B	1.013	1.013	1.013	1.013	1.013	0
60	ave_exit B	0	0	0	0	0	0
60	nobs UB	11281	16736	8368	8368	8368	-2913
60	ave_to UB	0	0.001	0.009	0.023	0.128	0.128
60	med_to UB	0	0	0.009	0.023	0.069	0.069
60	ave_GR UB	1.014	1.014	1.013	1.015	1.031	0.016
60	ave_NR UB	1.014	1.014	1.012	1.014	1.026	0.012
60	med_NR UB	1.013	1.012	1.012	1.012	1.016	0.003
60	ave_exit UB	0.704	0.636	0.555	0.642	0.699	-0.005

This table shows trading patterns for Gross Trades, Excluding Re-entry

Table 15: Gross Trade Include Reentry

threshold	Value	T1	T2	T3	T4	T5	DQ
0	nobs B	2922	5310	2654	2655	2655	-267
0	ave_to B	0	0.001	0.007	0.015	0.073	0.073
0	med_to B	0	0	0.007	0.015	0.038	0.038
0	ave_GR B	1.013	1.013	1.014	1.014	1.014	0.001
0	ave_NR B	1.013	1.013	1.013	1.013	1.013	-0.001
0	med_NR B	1.013	1.013	1.013	1.013	1.013	0
0	ave_exit B	0	0	0	0	0	0
0	nobs UB	11724	17760	8880	8880	8880	-2844
0	ave_to UB	0	0.001	0.01	0.025	0.136	0.136
0	med_to UB	0	0	0.01	0.024	0.067	0.067
0	ave_GR UB	1.015	1.014	1.013	1.014	1.024	0.009
0	ave_NR UB	1.015	1.014	1.012	1.001	1.011	-0.005
0	med_NR UB	1.013	1.012	1.012	1.012	1.015	0.002
0	ave_exit UB	0.698	0.628	0.541	0.599	0.631	-0.067
60	nobs B	2914	5306	2653	2653	2653	-261
60	ave_to B	0	0.001	0.007	0.015	0.073	0.073
60	med_to B	0	0	0.007	0.015	0.038	0.038
60	ave_GR B	1.013	1.013	1.014	1.013	1.014	0.001
60	ave_NR B	1.013	1.013	1.013	1.013	1.013	0
60	med_NR B	1.013	1.013	1.013	1.013	1.013	0
60	ave_exit B	0	0	0	0	0	0
60	nobs UB	11665	17733	8866	8866	8867	-2798
60	ave_to UB	0	0.001	0.01	0.025	0.126	0.126
60	med_to UB	0	0	0.01	0.024	0.068	0.068
60	ave_GR UB	1.014	1.013	1.013	1.014	1.023	0.01
60	ave_NR UB	1.014	1.013	1.012	1.013	1.02	0.006
60	med_NR UB	1.012	1.012	1.012	1.012	1.015	0.002
60	ave_exit UB	0.696	0.626	0.54	0.598	0.624	-0.072

This table shows trading patterns for Gross Trades, Including Re-entry

8.3.3 Cross Sectional Moments

Table 16 presents the cross sectional moments for both the balanced and unbalanced panels. For each month households are placed into a quintile based upon their turnover rates. These are averaged over the months by cell. This approach would capture relatively infrequent liquidity needs spread across households.

Notice that in the table the first 4 quintiles are grouped. The turnover rates are zero for at least 80% of the sample each month. Evidently, inaction is common in the cross section, punctuated by infrequent large adjustments. These trading patterns are smoothed over in the time series moments. Of the remaining 20% in the balanced panel, the mean (27.2%) and median (13.1%) turnover rates are higher than their time series counterparts. The same point holds for the unbalanced panel. And, as in the time series results, the turnover rates in the highest quintile are much larger for the unbalanced panels.

Table 16: Cross Section: Net Trades

	T1-T4	T5	DR
	Balanced Panel		
# of Obs	704,786	93,118	n.a.
mean Position	53114.54	163759.3	n.a
mean TO	0	.2721021	n.a
median TO	0	.1314092	n.a
mean GR	1.01097	1.015673	0.004703
median GR			
mean NR	1.010969	1.010044	-0.000925
median NR	1.011111	1.011203	0.000092
	Unbalanced Panel		
# of Obs	1,744,799	224,705	n.a.
mean Position	41970.59	124268.2	n.a
mean TO	0	.4020133	n.a
median TO	0	.1693548	n.a
mean GR	1.009732	1.017087	0.007355
median GR			
mean NR	1.009731	.977579	-0.032152
median NR	1.009036	1.009168	0.000132

This table reports data moments using the cross section approach, for both balanced and unbalanced panels. These are the time series averages of the monthly statistics. Here "Ti" is the turnover rate for quintile "i" and "DR" is the difference in the net return between the highest and lowest turnover rate quintiles.