

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

Yosef Bonaparte<sup>†</sup>and Russell Cooper<sup>‡</sup>and Mengli Sha<sup>§</sup>

April 21, 2020

## Abstract

Barber and Odean (2000) find that households who trade more have a lower net return than others and attribute this pattern to irrationality, particularly 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 pattern of households with large turnover having lower net returns. Various forms of irrationality, modeled as beliefs about income and return processes that are not data based, **do not** improve the ability of the baseline model to explain these turnover and net returns patterns.

---

\*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.

<sup>†</sup>University of Colorado Denver, Yosef.bonaparte@ucdenver.edu

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

<sup>§</sup>Department of Economics, the Pennsylvania State University, mengli.sha0124@gmail.com

# 1 Motivation

In a highly 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.*

Barber and Odean (2000) provide neither a theoretical proof nor direct evidence that the patterns they document are inconsistent with rational choice nor do they provide a model with irrational agents that matches these observations. Their statements are provocative but not convincing.

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, highlighted in Barber and Odean (2000), that turnover rates and net returns are inversely related.<sup>1</sup> The first looks at the choice of rational agents faced with costs of portfolio adjustment. The second allows for different forms of irrationality, including overconfidence.

For rational, optimizing households, 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 its portfolio. Trading costs 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.

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 stock portfolio.

We generate simulated panel data from the estimated model to study the relationship between portfolio adjustment and returns. Moments from the simulated panel are calculated in two ways. First, the data is averaged over time by household. Distributions of turnover and net returns are studied. This is called the “time series” approach. Second, the distributions of turnover and net returns are calculated for each month and then averaged over time. The first approach highlights differences across households while the second highlights more of the high frequency fluctuations in turnover and net returns. This is called the “cross section” approach.

As discussed in detail below, throughout the analysis, the focus is on the trades between stocks and bonds. Thus the analysis excludes portfolio churning (excessive rebalancing) and thus by construction misses some of the trades underlying the evidence provided by Barber and Odean (2000).<sup>2</sup> This simplification of the

<sup>1</sup>See Barber and Odean (2001) for additional evidence, including a gender breakdown, and further discussion of overconfidence.

<sup>2</sup>Accordingly, we do not rely on the moments presented in Barber and Odean (2000) but rather calculate them using the time series and cross sectional approaches directly from the same data set. Statements characterizing our results pertain to matching these moments, not those from Barber and Odean (2000).

analysis reflects the computational challenges of multi-asset models. Nonetheless we demonstrate that the inverse relationship between turnover and net returns is present even when portfolio churning is excluded. The challenge to rational models posed in Barber and Odean (2000) remains.<sup>3</sup>

For these models, turnover, measured by the exchange of stocks for bonds, is driven by liquidity needs stemming from income shocks. Due to the presence of portfolio adjustment costs, liquidity is mostly provided by variations in a more liquid bond account. But, in the presence of large enough shocks and depleted liquid resources, stock adjustment occurs. Due to the interaction of the fixed costs of adjustment and income fluctuations, the resulting turnover can be large enough to match data moments with income processes supported by the data. Further, the net return differential between high and low turnover households matches the data as well. **Thus, though Barber and Odean (2000) argue emphatically that liquidity needs alone are insufficient to generate the patterns of turnover and net returns, we find this is not the case.**

Following the suggestion of Barber and Odean (2000), we also study a series of models which relax the assumption of perfect rationality to model overconfidence. This is in line with the literature that generally models overconfidence as mis-calibration, i.e., overestimating the precision of information about the price of a financial security (e.g., Kyle and Wang (1997), Odean (1998) and Gervais and Odean (2001)). We specifically consider models in which traders over-estimate the volatility of labor income, over-estimate the serial correlation in returns or respond to signals that are not informative about returns.

To be clear, by “irrational” we mean to capture the behavior of utility maximizing households who hold beliefs about exogenous processes, such as returns, which are not consistent with data. Our evaluation of these alternatives is based upon the optimizing behavior built upon erroneous beliefs. The implied decision rules are taken to the same data as those based upon rational choice.

The introduction of irrational traders does **not** lead to any improvement in model fit when the moments are calculated either from the time series or cross sectional approaches. Evidently, the model of stochastic income with trading costs is sufficient to match the turnover and net return patterns that exclude portfolio rebalancing.

As a final exercise, we attempt to match the moments presented in Barber and Odean (2000), even though the turnover rates they calculate include portfolio rebalancing that is, by construction, absent from our model. Surprisingly, the model with rational agents can even match those patterns. But for this exercise, there are some modest improvements in fit from the presence of irrational agents.<sup>4</sup>

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 subsequently viewed as irrational.

We do not concur. 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.<sup>5</sup>

<sup>3</sup>Else, one would be led to conclude that the irrational behavior, if it exists at all, is associated solely with excessive stock turnover. This is not the case.

<sup>4</sup>The discussion makes clear the meaning of “modest” in this context.

<sup>5</sup>To be clear, there is some gain from irrational agents in matching the Barber and Odean (2000) moments that include portfolio rebalancing, as explained in detail in section 6.3.

## 2 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.<sup>6</sup> 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.

To be clear, portfolio turnover refers to trade between stocks and bonds.<sup>7</sup> The model does not include multiple stocks and thus rebalancing of the components of a stock portfolio is excluded. This implies that the model understates actual turnover and the associated costs of rebalancing. Missing these trades makes it more difficult to match the high turnover, low net return moments highlighted by Barber and Odean (2000).

### 2.1 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 bonds and stocks respectively and  $R^s$  is the stochastic return on stocks. The return on bonds,  $R^b$  is deterministic. A household chooses between (i) portfolio adjustment and (ii) no portfolio adjustment. This choice is given:

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

for all  $\Omega$ .

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

$$\begin{aligned} v^a(\Omega) &= \max_{b' \geq 0, s' \geq 0} u(c) + \beta E_{\Omega'|\Omega} v(\Omega') \\ \text{s.t.} \\ c &= \psi y + R^b b + R^s s - b' - s' - C(s, s') - F. \end{aligned} \quad (2)$$

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 are three forms of adjustment costs in the model. There is a time cost of stock adjustment represented by  $\psi \leq 1$  in (2).<sup>8</sup> Second, the model allows a fixed cost of adjustment,  $F$ . This is distinct from the time cost of  $\psi$  as there is no interaction with income.<sup>9</sup> Finally, the model includes direct trading costs, explained further below, captured by  $C(s, s')$ . In addition to the frequency of adjustment these costs also

<sup>6</sup>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.

<sup>7</sup>It is straightforward to extend the model to multiple stocks, essentially by letting  $s$  in the state be a vector of stocks. The challenge is to estimate a model with a large state space.

<sup>8</sup>Bonaparte, Cooper, and Zhu (2012) discusses other specifications of adjustment costs.

<sup>9</sup>We will study the two cases of  $\psi > 0$  and  $F > 0$  separately so that identification of the adjustment cost is not an issue.

generate a demand for bonds and thus impact the stock share.

If the household chooses **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.*

$$c = y + R^b b - b' \tag{3}$$

$$s' = R^s s. \tag{4}$$

Here we assume that if there is no portfolio rebalancing, any return on stocks is automatically put into the stock account, i.e.  $s' = R^s s$ . In the robustness analysis, we relax this assumption so that stock returns are instead deposited into the bond account.

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. The incentive for portfolio adjustment comes from large shocks to income and returns. 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. Likewise, 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. Further, a large negative shock to income will also create an incentive to rebalance, independent of current returns. Added to this is the differential between measured gross and net returns created by the presence of trading costs,  $C(\cdot)$ . These types of responses to shocks form the link from the policy functions generated by the household optimization problem to the evidence of Barber and Odean (2000).

## 2.2 Irrational Household

Here we use the model to explain the forms of irrational behavior studied in the quantitative analysis. 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, for example, time inconsistencies nor intransitive preferences.

### 2.2.1 Misperceptions

The decision rules from the dynamic optimization problem require the agent to compute conditional expectations. For rational agents, we will require that expectations are consistent with the income and return processes from the data. With irrational agents, we relax this requirement. Doing so is relatively straightforward in the model: just replace the conditional expectation in (2) and (3) with an alternative representation of beliefs.

**Income** Consider the following representation of the income process.

$$y_t = (1 - \tilde{\rho}_y) \tilde{\mu}_y + \tilde{\rho}_y y_{t-1} + \varepsilon_t^y \tag{5}$$

where  $\varepsilon_t^y$  is the period  $t$  innovation and  $y_t$  is the period  $t$  log level of income. Assume that the innovation is normally distributed with mean zero and standard deviation of  $\tilde{\sigma}_y$ .

For rational agents, the parameters  $(\tilde{\mu}_y, \tilde{\rho}_y, \tilde{\sigma}_y)$  are taken from the data. Else, agents may hold beliefs about the income process that differs from those obtained from the data.

The expected income process used in the optimization problem is easily modified to reflect these alternative beliefs. The key issue, discussed below, is one of identification: are decision rules sufficiently dependent on beliefs that  $(\tilde{\mu}_y, \tilde{\rho}_y, \tilde{\sigma}_y)$  can be inferred from the moments.

**Returns** Analogous to the income process, consider the following representation of beliefs of agents about returns:

$$R_t = (1 - \tilde{\rho}_R)\tilde{\mu}_R + \tilde{\rho}_R R_{t-1} + \varepsilon_t^R \quad (6)$$

where the notation parallels that of (5). Again, 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 (6), 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”.

### 2.2.2 Noisy Advice

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 be induced to believe 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 (1) and the options of adjustment and non-adjustment given by (2) and (3) respectively with the modified state vector of  $(\Omega, z)$ .

This choice is given:

$$v(\Omega, z) = \max\{v^a(\Omega, z), v^n(\Omega, z)\} \quad (7)$$

for all  $(\Omega, z)$ . The revised options are:

$$\begin{aligned} v^a(\Omega, z) &= \max_{b' \geq 0, s' \geq 0} u(c) + \beta E_{\Omega'|\Omega, z} \int_{z'} v(\Omega', z') dG(z') \\ \text{s.t.} \\ c &= \psi y + R^b b + R^s s - b' - s' - C(s, s'). \end{aligned} \quad (8)$$

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

$$\begin{aligned} v^n(\Omega, z) &= \max_{b' \geq 0} u(c) + \beta E_{\Omega'|\Omega, z} \int_{z'} v(\Omega', z') dG(z') \\ \text{s.t.} \\ c &= y + R^b b - b' \quad (9) \\ s' &= R^s s. \quad (10) \end{aligned}$$

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  $\Omega'$ .

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,  $R^{s'}$ .

### 3 Trading Costs and Exogenous Processes

The goal of the analysis is to study the relationship between portfolio turnover and return. To do so, we must estimate the parameters of the household choice problem. The estimation uses a simulated method of moments approach based upon this model of dynamic household choice. Solving the household optimization problem requires the specification of trading costs and exogenous income and return processes.

#### 3.1 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.<sup>10</sup>

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

$$C^b(s_{-1}, s) = \nu_0^b + \nu_1^b(s - s_{-1}) + \nu_2^b(s - s_{-1})^2 \quad (11)$$

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

$$C^s(s_{-1}, s) = \nu_0^s + \nu_1^s(s_{-1} - s) + \nu_2^s(s - s_{-1})^2. \quad (12)$$

Bonaparte and Cooper (2009) use this monthly household account data to estimate these parameters.<sup>11</sup> The trading costs, measured in dollars, are estimated in a regression where the dependent 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 $\nu_0^i$	56.10 (0.05)	61.44 (0.061)
Linear $\nu_1^i$	0.0012 (1.63e-06)	0.0014 (1.93e-06)
Quadratic $\nu_2^i$	$-1.01e^{-10}$ (2.88e-13)	$-1.28e^{-10}$ (9.26e-13)
Adj. $R^2$	0.251	0.359
Number of Observations	1,746,403	1,329,394

Table 1: Estimated Trading Costs

<sup>10</sup>The following section as well as the Appendix provide additional details about the data set and the calculation of these costs and moments.

<sup>11</sup>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.

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.

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 1 when the trading costs are integrated into the household optimization problem.

### 3.2 Income and Returns

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.<sup>12</sup> 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.<sup>13</sup>

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.

Of course, it is not clear *a priori* that unemployment risk is an issue for those households with direct holdings at a brokerage account. We return to this later in the discussion of results.

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 1.0061 with a standard deviation of 0.0439. The estimated serial correlation of annual returns is not significantly different from zero. Barber and Odean (2000) find no significant difference in the returns for the average households in their sample compared to a market index.<sup>14</sup>

<sup>12</sup>This is the same process as used in Bonaparte, Cooper, and Zhu (2012) and is described in the Appendix of that paper. Variables are normalized by the mean income of 6000 dollars per month income taken from the Barber and Odean (2000) sample.

<sup>13</sup>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.

<sup>14</sup>Specifically, from Table II and related discussion, the average return of the households in the sample are not statistically different from a NYSE/AMEX/Nasdaq value weighted index.

## 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:

$$J = \min_{(\Theta)} (M^s(\Theta) - M^d)' W ((M^s(\Theta) - M^d)). \quad (13)$$

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

The parameter vector is  $\Theta = (\beta, \gamma, \psi)$  in the case of rational households. 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 the re-estimation of  $(\beta, \gamma, \psi)$  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 problem. The simulated moments are calculated from this panel, just as in the actual data.<sup>15</sup> 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 1.

### 4.2 Moments

The choice of moments is motivated by 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. Our calculation of these moments is slightly different than reported by Barber and Odean (2000) to match more closely the elements of our household choice model.<sup>16</sup> As discussed below, these moments are informative about the parameters to be estimated.

The data is a panel, indicating trades by household over their period of activity with this particular discount broker. This creates a couple of important issues in calculating the turnover and return moments. First, there are two ways to create turnover quintiles from these monthly measures of turnover.<sup>17</sup> Second, the moments depend on the treatment of exit from the sample.

The first approach to measuring turnover, hereafter called "time series", calculates the quintiles of the time series average of the turnover rates by household. 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.

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

<sup>15</sup>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.

<sup>16</sup>The procedures used to produce the moments are detailed in the Appendix.

<sup>17</sup>We are grateful to an anonymous referee for challenging our initial interpretation of the Barber and Odean (2000) methodology.

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. As we shall see, this measure captures quite well the short-run liquidity needs of investors.

The second issue is 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.<sup>18</sup> In contrast, “unbalanced” moments are from households who started in January 1991 but did not necessarily maintain an account through the entire sample.

Our baseline model does not include exit and thus we match moments created from a balanced panel. Focusing on households with an account in January 1991 is consistent with Barber and Odean (2000) though it is not clear how they account for exit and the resulting trades in their calculations. Thus we study both balanced and unbalanced panels when matching those moments.

#### 4.2.1 Turnover

For this analysis, 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. This is consistent with our model in that there is only a single stock. In contrast, as discussed below, the Barber and Odean (2000) includes rebalancing.

Specifically, turnover for household  $i$  in period  $t$  is

$$T_{i,t} \equiv \left| \left( \frac{s_{i,t} - s_{i,t-1}R_t^s}{s_{i,t-1}R_t^s} \right) \right|. \quad (14)$$

Note that the turnover rate depends on the household’s stock wealth at the start of period  $t$ ,  $s_{i,t-1}R_t^s$ , inclusive of current stock returns.<sup>19</sup>

Table 2 shows the turnover moments. The turnover patterns are quite different comparing the cross section and time series approaches. Clearly, the cross section picks up a considerable amount of inaction so that the turnover rates in the 4 lower quintiles are much higher in the time series approach. In fact, over 80% of the households do not adjust their portfolio in a given month.<sup>20</sup> But, the highest quintile has a very large average turnover rate in the cross section compared to the time series.

Evidently, inaction is common in the cross section, punctuated by infrequent large adjustments. These trading patterns are smoothed over in the time series moments. This might be interpreted as the smoothing out of the short-term liquidity needs that motivates the large trades at the monthly frequency.

#### 4.2.2 Net Return

We calculate the net return for household  $i$ , denoted  $R_i^n$ , **on the stock portfolio** as

$$R_{i,t}^n = \frac{s_{i,t-1}R_t^s - C(s_{i,t} - s_{i,t-1}R_t^s)}{s_{i,t-1}}. \quad (15)$$

Here the cost function comes from the results in Table 1. In calculating the net return costs of trade due to the bid and ask spread are included as well.

<sup>18</sup>Specifically, this requires 71 months of reports in the position data.

<sup>19</sup>So  $s_{i,t}$  represents holdings at the end of period  $t$ .

<sup>20</sup>By comparison, Bonaparte, Cooper, and Zhu (2012) report a 46.7% **biannual** stock adjustment rate for households in the PSID.

Table 2: Data Moments

case	Mean					Median		
	t1	t2	t3	t4	t5	DR	t5	DR
Time Series	0.00	0.0039	0.0134	0.0291	0.1155	-0.0010	0.0690	-0.0010
Cross Section	0.00	0.00	0.00	0.00	0.2721	-0.0009	0.1314	0.0001
BO	0.002	0.012	0.029	0.060	0.215	-0.0046	na	na

This table reports data moments. 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. The row labeled BO is from Table V from Barber and Odean (2000) of that paper. The row labeled “cross section” reports the time series average of the quintiles computed from the cross section for each month. The row labeled “time series” calculates the quintiles of the time series average of the turnover rates by household. The last two columns compute t5 and DR using median turnover and rates.

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.

Clearly, there is a mechanical relationship between high turnover and low net return. If the household, for example, buys stocks, then using the specification of the trading cost function,  $C(s'_i - R^s s_i)$ , the net return becomes:

$$R_i^n = R^s - \frac{\nu_0^b}{s_i} - \frac{\nu_1^b(s'_i - R^s s_i)}{s_i} - \frac{\nu_2^b(s'_i - R^s s_i)^2}{s_i}. \quad (16)$$

From this calculation, the linear part of the cost function implies that the net return is lower when trades are large and when the initial stock holdings is relatively low. 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. Below we discuss how much of the low net returns is associated with high turnover through the linear part of the cost function.

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 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.

The differential in net returns is reported in Table 2 in the “DR” column. The sign and magnitude of the return differential depends on the sample. For the two balanced samples, the return differential is negative, as reported in Barber and Odean (2000). In fact, the differential is largest in the cross section, an order of magnitude larger than in the time series.

### 4.2.3 Barber and Odean (2000) Moments

The final row of Table 2 reports the moments directly taken from Barber and Odean (2000). They are clearly closer to our time series (household averaged quintiles) moments than those from the cross section.<sup>21</sup> There are a couple of differences in calculation of the moments.

First, recall that our turnover is net, thus excluding trades that do not involve changes in stockholdings. From their Table V, Barber and Odean (2000) calculate monthly turnover as the average of sales and purchases so that portfolio rebalancing is included as well. From that perspective, turnover for household  $i$

<sup>21</sup>There is not sufficient detail in Barber and Odean (2000) to determine exactly the calculation of their moments nor sample selection. We suspect, but cannot verify, that these are time series moments from an unbalanced panel.

in period  $t$  is

$$T_{i,t} \equiv \left| \left( \frac{\sum_j |s_{i,j,t} - s_{i,j,t-1} R_{jt}^s|}{s_{i,t-1} R_t^s} \right) \right|. \quad (17)$$

In this expression,  $j$  indicates a particular asset. So the numerator is the amount of asset  $j$  traded. Total turnover is then simply the sum of all asset trades during a period of  $t$ . The denominator represents total financial wealth of household  $i$  at the start of period  $t$ .

Second, as there is no discussion of exit in their analysis, it seems their moments are based upon an unbalanced panel. So the averaging across households in Barber and Odean (2000) includes those with different sample lengths. And, perhaps more importantly, may also include the large trades made upon exit.

Table 2 makes clear the additional turnover calculated by Barber and Odean (2000) compared to our time series and cross section tabulations for unbalanced panels. Their turnover rates in the 4 lowest quintiles are higher than our moments. But for the key turnover category, t5, we obtain about the same turnover rate in the time series.

Third, we use our formula to calculate the net return so that our model and the data are consistent. The return differential in the “BO” row of Table 2 is based upon their calculation of the total trading costs.

Though our model does not include portfolio rebalancing, we will estimate the model to match these moments are well. In part this is informative about how much of the overall pattern detected by Barber and Odean (2000) can be explained by liquidity needs alone.

#### 4.2.4 Outliers

The turnover and return calculations are based upon the means by quintile. Allowing 5 quintiles certainly highlights the distribution of turnover and returns. Yet, it turns out, there remains substantial heterogeneity within the cells, particularly the high turnover group.

The last two columns in Table 2 report the DR and t5 measures based upon the median of the quintile. In this case, the turnover rates and net return differentials drop substantially.<sup>22</sup>

For the cross section, the DR measures are all slightly positive. Evidently the negative average return differentials were driven by the large turnovers in t5.

Figure 1 shows the distribution of turnover rates within the t5 quintile, using the time series approach. The panels differ in terms of the turnover rate, with the right side illustrating a small fraction of trades with turnover in excess of 30% in a given month. Clearly, there are some extremely large trades inside of the t5 group, present in both the cross section and the time series. These trades seem to generate the net return differential. The estimation returns to this below as we study results based upon median rather than mean measures of return and turnover.

## 5 Results

The estimation results are presented for the time series and cross section calculation of the turnover and return moments. The section that follows goes further to study a mixture of time series and cross sectional moments as well as the moments from Barber and Odean (2000).

A model based upon liquidity shocks on turnover and net return is likely to be able to match the cross section moments. The presence of the adjustment costs will create the monthly inaction in turnover for most

<sup>22</sup>Though not in the table, the differences between the balanced and unbalanced panels return differential disappears and the highest turnover rates are much closer.

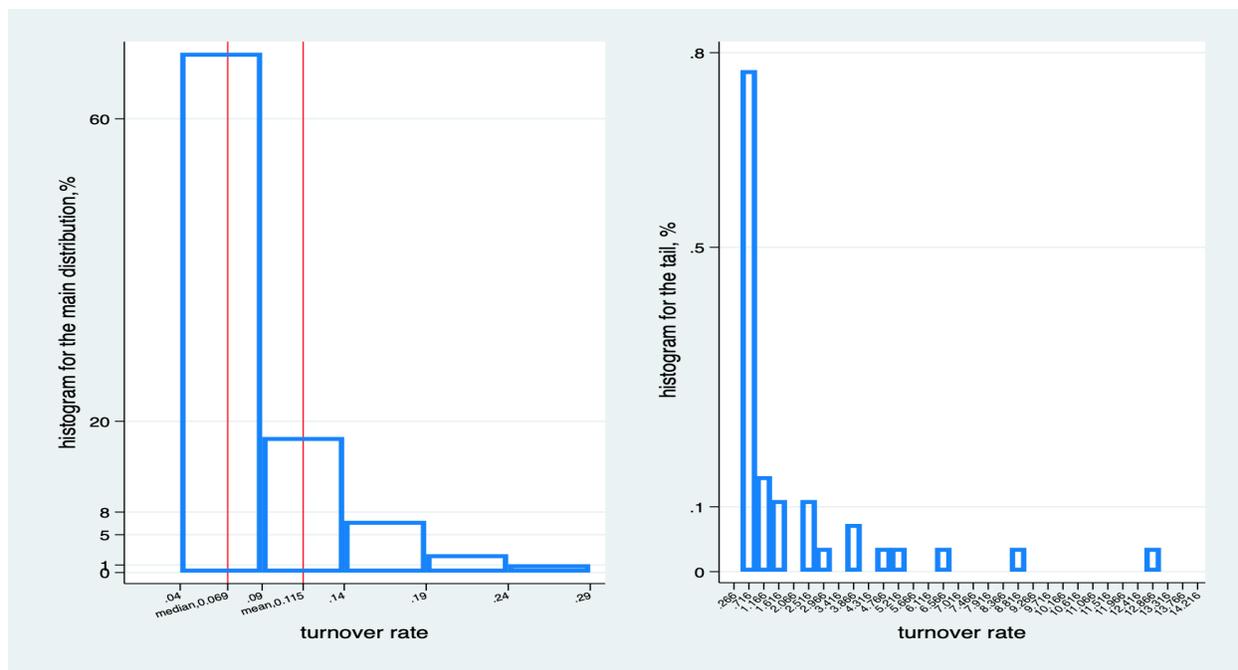


Figure 1: Turnover Distribution: t5

households, punctuated by some large adjustments by the rest. These large adjustments reflect not just the income shock but the imbalance between the stock and bond accounts created by previous inaction.

These liquidity needs are clear at the monthly frequency. The challenge is the time series where these liquidity needs are likely to be averaged over time.

Throughout, there is the issue of exit and thus the choice between a balanced and the unbalanced panel. For a couple of reasons, we focus on the balanced panel. First, as discussed by Barber and Odean (2000), the choice of households to have an account with this broker is difficult to understand given the transactions costs and performance. Thus formally modeling this choice of participation and ultimately exit from the relationship is non-trivial. Second, our model does not include a participation decision and thus the choice between direct and indirect holdings.<sup>23</sup>

Importantly, our choice of a balanced panel **does not** mitigate the key findings in the data. As made clear in Table 2, for both the cross section and time series moments, there are high turnover rates in the fifth quintile and a lower net return compared to the first quintile.

There are two main findings. First, there is no evidence of irrational choices in the time series analysis. All of the turnover and net return differentials are driven by liquidity needs interacting with the fixed adjustment costs. Second, irrationality does improve the model fit in the cross section iff there is no unemployment risk. Once that risk is present, the gains from irrational agents disappear.

## 5.1 Time Series

In this case, the turnover moments are calculated from quintiles of the time series average of turnover by household. That is, for a particular household  $i$ , the average of the monthly turnover is calculated from the

<sup>23</sup>There does not appear to be any such dynamic choice model in the literature. This is left for future work.

Table 3: Data and Model Moments: Time Series

case	t1	t2	t3	t4	t5	DR
Data	0	0.0039	0.0134	0.0291	0.1155	-0.0010
Rational						
No UR	0.0065	0.0113	0.0168	0.0247	0.1151	-0.0017
UR	0.0065	0.0104	0.0157	0.0242	0.1159	-0.0018
No UR (M)	0.00	0.00	0.0141	0.0257	0.0710	-0.0008
UR (M)	0.00	0.00	0.0237	0.0349	0.0694	-0.0016
Irrational						
$\delta$	0.0065	0.0113	0.0168	0.0247	0.1151	-0.0017
$\sigma_R$	0.0065	0.0115	0.0174	0.0262	0.1076	-0.0018
$\rho_R$	0.0065	0.0113	0.0168	0.0248	0.1152	-0.0017
$\mu_R$	0.0066	0.0113	0.0174	0.0266	0.1137	-0.0026
good news	0.0065	0.0113	0.0168	0.0247	0.1151	-0.0017
bad news	0.0063	0.0112	0.0171	0.0259	0.1160	-0.0017

This table reports data and simulated moments by household averages over time. The panel is balanced. “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. The rows with  $M$  use the median turnover and return for t5.

period in which the household had an account. The turnover rates are then ranked to create the quintiles. For this exercise, the data and simulated moments come from a balanced panel.

Table 3 reports the findings in terms of moments. The model with rational agents, either with or without unemployment risk matches the turnover profile well and generates the large turnover in the fifth quintile. The return differential is actually too large relative to the data.

The mechanism here, as discussed earlier, is a combination of liquidity needs and adjustment costs. Rational households desire to smooth consumption relative to their stochastic income. But liquidity is costly, given the return differential between bonds and stocks. Thus, consumption smoothing arises from using the bond account to buffer income shocks supplemented by occasional stock trades. Thus the magnitude of the turnover is enhanced by the adjustment costs. Clearly this alone can account for the high turnover as well as differential in returns.

Table 4 reports the parameter estimates and fit. Across the cases, the estimates of  $\beta, \gamma, \psi$  are quite close. At the annual frequency, the discount factor is about 0.86, close to the estimate reported in Cooper and Zhu (2015).<sup>24</sup> The adjustment costs ranges from 1.3 to 2% of average monthly income.

The fit is very close for the two rational model, keeping in mind that there are 6 moments and 3 parameters. The model with unemployment risk fits slightly better.

There are six cases of irrationality reported. Specifically the misperceptions allowed are about: (i) unemployment risk  $\delta$ , (ii) volatility in asset returns  $\sigma_R$ , (iii) the serial correlation in asset returns,  $\rho_R$ , (iv) the mean of asset returns,  $\mu_R$ , (v) good news and (vi) bad news.

The first four of these exercises implements the discussion in section 2.2.1. To implement the noisy advice model, assume  $z \in \{0, 1\}$ . When  $z = 0$ , households believe that  $R^s$  is an iid process, as specified in the baseline model. Alternatively,  $z = 1$  leads the households to put all weight on the lowest (highest) realization of  $R^{s'}$ . These are the “bad” and “good” news cases respectively. The estimation includes  $p$ , the probability that  $z = 1$  for the two cases in which this realization of  $z$  is viewed as either extremely good or bad news.

For some of these estimation exercises, parameters were started in the neighborhood of the rational estimates with the additional parameter of perceptions set at the value from the data. Additional estimation

<sup>24</sup>The product of the annual discount factor and the mean return is 0.944, thus less than unity.

was started at a number of distant parameter values to study global identification. The results reported are the best fits from these various starting points.

Table 4: Parameter Estimates: Time Series

case	$\beta$	$\gamma$	$\psi$	IR Parm	DP	$J$
	Rational					
No UR	0.9876	2.1659	0.9868	na	na	1.3e-04
UR	0.9892	2.2552	0.9803	na	na	1.2e-04
No UR (M)	0.9613	2.5932	0.9389	na	na	2.7e-05
UR (M)	0.9760	0.9866.	0.9889	na	na	1.6e-04
	Irrational					
$\delta$	0.9876	2.1659	0.9867	0.0149	0.0140	1.3e-04
$\sigma_R$	0.9876	2.1658	0.9868	0.0439	0.0439	1.3e-04
$\rho_R$	0.9876	2.1659	0.9868	0.0004	0	1.3e-04
$\mu_R$	0.9883	2.2407	0.9854	1.0057	1.0061	1.3e-04
good news	0.9876	2.1659	0.9868	0.0000	0	1.3e-04
bad news	0.9877	2.1670	0.9864	0.0000	0	1.2e-04

This table reports estimated parameters for the various cases:  $\beta$  is the discount factor,  $\gamma$  is relative risk aversion and  $\psi$  is the fraction of income remaining after portfolio adjustment. For the irrational cases, “IR Parm” is the estimated beliefs and “DP” is the parameter from the data.  $J$  is the difference between model moments and data moments as described in equation (13). The rows with  $M$  use the median turnover and return for t5.

Table 4 there are some small differences in the parameters and in some experiments, such as the  $\mu_R$  case, a deviation between the estimated belief, “IR Parm”, and the statistic from the data, “DP”. But in no case does the fit improve upon the rational model with unemployment risk.

This is a key finding of the paper. **Allowing misperceptions about the stochastic processes for income and return does not improve the fit over the rational model.**

From Table 2, the behavior in the fifth quintile itself seems to be driven by outliers. The median turnover and return differential in that cell is 0.0812 and 0.0020 respectively, quite different from the mean values used in the estimation. This suggests that perhaps the largest turnovers are causing the negative return differential.

To study this, the model is re-estimated to focus on the median t5 and DR moments, labeled “(M)” in the tables of results.<sup>25</sup> We find that the model fit is not as good in this case. This is informative: there are not large trades in the tails that make it difficult for the rational model to match the moments. If anything, the opposite is true.

## 5.2 Robustness

There are a couple of robustness exercises based upon these estimates. These include estimating a fixed cost of adjustment, changing the dividend reinvestment procedure, increasing the mean level of income, reducing the cost of trade and using a weighting matrix.

For all of these cases, we see that the theme remains. The model with rational agents is able to reproduce both the return differential and the high turnover rates.

<sup>25</sup>The data moments are given in the appropriate row of Table 2,

Table 5: Parameter Estimates: Robustness

case	$\beta$	$\gamma$	$\psi$	$J$
Baseline	0.9876	2.1659	0.9868	1.3e-04
Low Cost	0.9876	2.1635	0.9868	2.0e-04
No Reinvestment	0.9881	2.3782	0.8716	3.0e-04
F	0.9854	2.8566	0.0469	1.2e-04

This table reports estimated parameters for the various cases:  $\beta$  is the discount factor,  $\gamma$  is relative risk aversion and  $\psi$  is the fraction of income remaining after portfolio adjustment.  $J$  is the difference between model moments and data moments as described in equation (13). The row labeled Baseline is the baseline estimation with no unemployment risk.

### 5.2.1 Low Cost

The case of “low cost” reduces the fixed costs of buying and selling to 10% of their estimated values. The motivation of this is simply that currently trading costs are considerably lower and it is interesting to see if the patterns of turnover and return differential remain.

Not surprisingly, the lower trading costs, at the baseline parameters, leads to higher turnover rates in the lowest quintile and lower turnover rates in the highest quintile. Consequently the return differential is lower.

If the model is re-estimated, at the new parameters turnover in the highest quintile is restored. But the return differential remains very low. The fit with the new estimation is not nearly as good as the baseline. The parameters are close to the baseline though the trading cost,  $(1 - \psi)$ , is positive and thus a bit higher than the baseline to compensate for the reduction in the fixed cost of trading.

### 5.2.2 No Reinvestment

In the baseline model, it is assumed that even if there is no portfolio adjustment, stock returns are deposited in the stock account. As a consequence, the stock share increases in high return states unless the adjustment cost is paid. When there is adjustment, there are stock sales in high return states.

An alternative is to assume that stock returns are paid to the bond account. With that assumption, portfolio rebalancing will go in the other direction in response to a high return shock. The effect of this modification on the moments will depend, in part, on the non-linear nature of the decision rules.

Table 6: Data and Model Moments: Robustness

case	t1	t2	t3	t4	t5	DR
Data	0	0.0039	0.0134	0.0291	0.1155	-0.0010
Baseline	0.0065	0.0113	0.0168	0.0247	0.1151	-0.0017
Low Cost	0.0073	0.0127	0.0201	0.0297	0.1101	-0.0007
No Reinvestment	0.0102	0.0156	0.0191	0.0250	0.1152	-0.0009
F	0.0064	0.0110	0.0168	0.0253	0.1145	-0.0013

This table reports data and simulated moments by household averages over time. The panel is balanced. “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. The row labeled Baseline is the baseline estimation with no unemployment risk.

With this re-estimation, the estimated cost of portfolio adjustment is much larger to offset the increased need for the liquidation of income paid to the stock account. The fit is not as good as other specifications. Yet, the patterns of high turnover and the differential in returns is well matched.

### 5.2.3 F

For this experiment, the opportunity cost of adjustment is replaced by a fixed cost, denoted  $F$ . Here  $F$  is relative to mean income and is restricted,  $F \geq 0$ .<sup>26</sup>

The resulting moments are again quite close to the baseline. The fit of the model is a little better. The estimated value of risk version is a bit larger and the adjustment cost is 5% of average income.

### 5.2.4 Weighting Matrix

Another robustness exercise is to redo the baseline estimation using the optimal weighting matrix. As discussed in the appendix in detail, the variance-covariance matrix is obtained by bootstrapping.

Table 7: Parameter Estimates: Optimal Weighting Matrix

case	$\beta$	$\gamma$	$\psi$	$J$
No UR	0.9681	1.4531	0.9830	3.41e2
UR	0.9670	1.4506	0.9819	1.45e2

This table reports estimated parameters for the various cases:  $\beta$  is the discount factor,  $\gamma$  is relative risk aversion and  $\psi$  is the fraction of income remaining after portfolio adjustment.  $J$  is the difference between model moments and data moments as described in equation (13).

The weighting matrix puts a lot of weight on the average turnover in the first quintile. Therefore more effort is used to match the zero in the first quintile instead of the average turnover in the highest quintile as well as the difference in net return, which are of more interest for this study.

In Table 7, the estimates of the two cases are fairly close to each other. Consequently, due to more risk from income, the UR case exhibits higher turnover and thus higher net return differential, as is shown in Table 8.

Table 8: Data and Model Moments: Optimal Weighting Matrix

case	t1	t2	t3	t4	t5	DR
Data	0.00	0.0039	0.0134	0.0291	0.1155	-0.0010
No UR	0.00	0.0000	0.0024	0.0216	0.0543	-0.0010
UR	0.00	0.0037	0.0173	0.0324	0.1541	-0.0018

This table reports data and simulated moments by household averages over time. The panel is balanced. “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.

## 5.3 Identification

Table 9 reports the response of the turnover moments to variations in the three parameters. The reported elasticities are calculated at the baseline estimates and simulated moments, based upon a 1% increase in the parameters.<sup>27</sup>

There are a couple of points illustrated by these calculations. First, the moments are very sensitive to changes in the discount factor and the adjustment cost,  $(\beta, \psi)$ . The response to  $\gamma$  is smaller but not insignificant, as some of the higher turnover rate moments are sensitive to variations in the risk aversion.

<sup>26</sup>As this is a change in the model itself, there are no simulation results possible at baseline parameters.

<sup>27</sup>The model is non-linear. Thus the magnitude and in some cases the sign of the moment change can depend on both the magnitude and direction of the parameter change. This was checked as well and the basic arguments for local identification remain.

Table 9: Elasticity of Moments to Parameter Values: Baseline

parm.	t1	t2	t3	t4	t5	DR
$\beta$	95.31	45.25	11.96	-14.43	-77.97	-95.39
$\gamma$	3.95	7.32	17.46	18.01	4.44	3.57
$\psi$	6.12	9.36	17.61	18.94	-20.27	21.11

This table reports the elasticity of moments with respect to parameters for the baseline model, based upon a 1 % increase.

Second, some of the internal mechanisms of the model are revealed by these elasticities. Again, these are complicated due to the rich nature of the choice model as well as the moments being matched.

The response to an increase in  $\beta$  is large and complicated. As households become more patient, the adjustment rate rises in response to income shocks.<sup>28</sup> Consequently, the three lowest turnover rates rise as well. But the higher return rates actually fall since the "pent up" demand to adjust is lower. Further, the response to a return shock is somewhat different, in part because the gross return is paid to the stock account. Thus consuming from this flow requires turnover. Also, as  $\beta$  rises, all else the same, the level of financial wealth is higher which decreases the turnover rate since stock holdings are in the denominator of that rate.

As households become more risk averse, i.e.  $\gamma$  increases, consumption smoothing is more important. This increases the frequency for all turnover rates and also increases the return differential, as seen in Table 9. Looking at the adjustment cost, an increase in  $\psi$  reduces inactivity and thus causes the turnover rates to rise for all but the highest turnover class. As with the higher discount factor, t5 is actually lower since the increase in the small adjustments leads to less frequent large adjustments.  $DR$  increases as well.

In addition to these calculations, the estimation entailed multiple starting values searching for the best fit. This would uncover local and global identification problems.

## 6 Matching Cross Sectional and Rebalancing Moments

This section goes beyond the estimation of the time series to study both the cross section moments and those reported by Barber and Odean (2000). For these cases, models with both rational and irrational households are estimated.

The results reported for the time series estimation hold for the cross section moments. In particular, we do not find improvements in fit by the addition of irrational households.

The estimation matching the moments from Barber and Odean (2000) are different in a couple of ways, reflecting the fact that they are calculated from gross rather than net trades. Importantly, we do find some modest gains allowing agents to have irrational beliefs about unemployment risk and/or the variability of returns. These differences are not driven by beliefs alone but rather by a combination of irrational beliefs and alternative parameter values.

### 6.1 Cross Section

The presentation of results based upon moments calculated from the time averaged cross sectional moments are split into two parts. The first reports results for a balanced panel. The second introduces exit into the

<sup>28</sup>At the baseline parameters and the average stock return, a 1% increase in  $\beta$  increases the adjustment rate from about 20% to over 80%. Interestingly, these adjustment rates are lower when the stock return takes either its highest or lowest value.

model and the estimation and studies the moments from an unbalanced panel.

### 6.1.1 Balanced Panel

Table 10: Data and Model Moments: Cross Section

case	t1	t2	t3	t4	t5	DR
Data	0.0000	0.0000	0.0000	0.0000	0.2721	-0.0009
	Rational					
No UR	0.0000	0.0000	0.0000	0.0000	0.2723	-0.0017
UR	0.0000	0.0000	0.0000	0.0015	0.2723	-0.0015
	Irrational					
$\rho_R$	0.0000	0.0000	0.0000	0.0000	0.2724	-0.0016

This table reports data and simulated 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.

In this case, the focus is on monthly turnover rates across households. These rates are ranked to create the monthly quintiles. The moments are calculated from a time series average of the means of the monthly quintiles of turnover. So, in a particular month, there will be some households with relatively low (high) turnover and that same household might have high (low) turnover in another month. The return differential is the time series average of the monthly differential in net return between the high and low turnover quintiles. For this exercise, as in the calculation of the time series moments, the data and simulated moments come from a balanced panel.

From the data moments reported in Table 10, the inaction in portfolio adjustment is apparent. The inaction is much higher than in the time series moments. But when there is turnover, it is large: the mean of the highest turnover quintile is over 27% in the cross section, compared to around 11% in the time series. Still, the return differential is about the same in the cross section as in the time series moments.

Table 11 reports the parameter estimates. Compared to the time series parameter estimates without unemployment risk, for the rational model, the discount factor is slighter lower, the risk aversion is higher and the portfolio adjustment costs are higher. With these parameter values, there is more portfolio adjustment inaction and, when adjustment occurs, the turnover rate is quite large.

For the cross section, the fit is better without unemployment risk. In this case, the estimated discount rate, risk aversion and adjustment cost fall relative to the cross section estimates without unemployment risk. The improvement in the model fit comes through the reduced average turnover in t4. For both of the rational models, the return differential is much **higher** than the data moment. Again, the problem matching the data does not arise from the excess turnover in t5 nor the return differential.

As conjectured earlier, the fit is considerably better with the cross sectional model. Again, this perspective highlights stochastic liquidity needs rather than smoothing them over time as in the time series analysis. Still, both types of variations can be explained through the lens of this model.

The introduction of irrationality in the form of a slight belief in serially correlated returns (0.0004 rather than 0) does improve model fit relative to the no unemployment risk case.<sup>29</sup> But relative to the case with unemployment risk, the fit is not better for any form of irrationality.

<sup>29</sup>Other forms of irrational beliefs did not lead to any improvement in fit.

Table 11: Parameter Estimates: Cross Section

case	$\beta$	$\gamma$	$\psi$	IR	DP	$J$
Rational						
No UR	0.9795	1.0666	0.9729	na	na	6.30e-07
UR	0.9727	1.1050	0.9531	na	na	3.57e-07
Irrational						
$\rho_R$	0.9790	1.0487	0.9626	0.0004	0	6.03e-07

This table reports estimated parameters for the various cases:  $\beta$  is the discount factor,  $\gamma$  is relative risk aversion and  $\psi$  is the fraction of income remaining after portfolio adjustment. For the irrational cases, “IR Parm” is the estimated beliefs and “DP” is the parameter from the data.  $J$  is the difference between model moments and data moments as described in equation (13).

## 6.2 Time Series and Cross Section Moments

The cross section moments highlight high frequency liquidity needs while the time series smooth over these variations. Naturally, it is of interest to see if both the time series and cross sectional moments can be matched with the same set of parameter values. From the estimates presented above, the estimated risk aversion is much lower and the discount factor lower when the cross sectional moments are matched.

Table 12: Parameter Estimates: Time Series and Cross Section

case	$\beta$	$\gamma$	$\psi$	$J$
Rational				
NO UR	0.9701	2.4518	0.9919	4.4e-03
UR	0.9486	2.2756	0.9842	4.1e-03

This table reports estimated parameters for the various cases:  $\beta$  is the discount factor,  $\gamma$  is relative risk aversion and  $\psi$  is the fraction of income remaining after portfolio adjustment.  $J$  is the difference between model moments and data moments as described in equation (13).

Table 13: Data and Model Moments: Time Series and Cross Section

case	Cross Section						Time Series					
	t1	t2	t3	t4	t5	DR	t1	t2	t3	t4	t5	DR
Data	0.0000	0.0000	0.0000	0.0000	0.2721	-0.0009	0	0.0039	0.0134	0.0291	0.1155	-0.0010
Rational												
No UR	0.0000	0.0003	0.0071	0.0235	0.2370	-0.0047	0.0156	0.0260	0.0342	0.0449	0.1502	-0.0020
UR	0.0000	0.0000	0.0014	0.0127	0.2340	-0.0046	0.0061	0.0178	0.0305	0.0428	0.1573	-0.0026

This table reports data and simulated 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.

Tables 12 and 13 report the parameter estimates and moments for this exercise. In this case we are fitting 12 moments with only three parameters.

Not surprisingly the fit is not as good as it is when the time series and cross section moments are matched separately. The estimated discount factor is close to the cross section estimate while the degree of risk aversion is much higher than the other estimates. The adjustment cost is about the same.

## 6.3 Barber and Odean (2000) Moments

The results have been presented using the moments calculated in a way that is consistent with our model with a single stock. For the most part, the turnover and net return moments are closely matched without

irrational agents, with **slight improvement** for some irrational beliefs. This is the case for both the time series and cross sectional moments.

Table 14: Parameter Estimates: BO

case	$\beta$	$\gamma$	$\psi$	$\Gamma$	IR Parm	Parm	$J$
Rational: Time Series, Balanced							
No UR	0.9830	1.7351	0.9947	na	na	na	4.36e-04
UR	0.9767	1.6029	0.9917	na	na	na	2.62e-04
Rational: Time Series, Unbalanced							
No UR	0.9886	2.5459	0.9683	0.0612	na	na	0.0015
UR	0.9927	1.3656	0.9987	0.0120	na	na	0.0018
Rational: Cross-Section							
No UR	0.9469	2.0056	0.9996	na	na	na	1.26e-05
UR	0.9797	1.5313	0.9986	na	na	na	6.67e-04
Irrational: Time Series							
$\tilde{\delta}$	0.9828	1.4122	0.9951	na	0.0368	0.0140	1.44e-04
$\tilde{\sigma}_R$	0.9832	1.5123	0.9898	na	0.0958	0.0439	1.80e-04

This table reports estimated parameters for the various cases:  $\beta$  is the discount factor,  $\gamma$  is relative risk aversion and  $\psi$  is the fraction of income remaining after portfolio adjustment.  $J$  is the difference between model moments and data moments as described in equation (13).

We now turn to the matching the moments presented in Barber and Odean (2000). This is worth pursuing for at least two reasons. First, these moments are a natural challenge given the strong conclusions in their study. Second, since the moments in Barber and Odean (2000) are calculated for all trades, including rebalancing, it is insightful to determine how well a model without rebalancing can match their moments. Doing so, provides a metric for determining the importance of portfolio rebalancing in generating patterns of turnover and net return.

That said, it should be clear that this exercise is off the domain of our model. In effect, by using the turnover moments from Barber and Odean (2000) we treat all trades, including rebalancing, as adjustments between stock and bond accounts. It is straightforward to supplement the dynamic optimization model to include multiple stocks. The problem with the estimation of such a model comes from the expansion of the state space making computation and estimation essentially impossible.

### 6.3.1 Exit

Besides the inclusion of rebalancing, there is a second deviation from our baseline analysis that is potentially important. As noted a number of time, it is not clear how exit was treated in the Barber and Odean (2000) set of moments. In our analysis, a household is said to exit when the record ends. Of the 53,394 households with accounts at the start of the sample in January 1991, only about 25% had active accounts at the end of the sample, 71 months later. The median time in the sample is 41 months. To the extent that traders were induced into high turnover with low net returns, exit is an important form of adjustment. Further, the turnover rate at the time of exit is, by construction, 100%. Thus exit surely contributes to the measures of turnover.

The optimization problem is modified to include a fixed flow cost of stock market participation, presented as a fraction of average income, denoted  $\Gamma$ . Thus the budget constraint for a participating, adjusting household is given by:

$$c = \psi y + R^b b + R^s s - b' - s' - C(s, s') - \Gamma \bar{y}. \quad (18)$$

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

$$c = y + R^b b - b' - \Gamma \bar{y}. \quad (19)$$

A household can choose to sell all of its stocks and thus avoids this cost. All assets are converted into bonds. There is an option to reenter upon payment of the fixed cost.

The results reported below include moments calculated from the baseline model without exit as well as from an unbalanced panel with exit. For the calculation of the latter moments, all trades, including those that effectively close an account, are included.<sup>30</sup>

### 6.3.2 Results

Tables 14 and 15 present the parameter estimates and moments for models with rational and irrational households where the target moments are from Barber and Odean (2000). For these estimates, the simulated moments are calculated in the two ways described: the time series and cross sectional approaches. For the simulated time series moments, results are presented for both balanced and unbalanced panels since the treatment of exit is not clear from the discussion of their moments.

For rational households, the estimates are presented in the top panel of Table 14. The panel labeled “Rational: Time Series” reports our main parameter estimates. As before, the row labeled “No UR” assumes no monthly unemployment risk while the “UR” row allows unemployment shocks. The next block presents the cross section estimates.

For the times series moments and the balanced panel, in the “No UR” case, the estimated value of  $\beta = 0.9830$  and the risk version estimate is 1.7351. The cost of adjustment  $\psi$  is slightly below one. Relative to the baseline, the estimated risk aversion and the discount rate are both lower. With unemployment risk, the fit is better with a lower discount factor and less risk aversion.

From Table 15, the estimated model captures quite well the distribution of turnover rates, including the highest quintile. This includes the near 21.5% turnover rate from the data. Thus the rational model has no problem matching the large turnover rates based upon liquidity needs alone. The estimated model captures about 63% of the net return differential.

The block of time series results with the unbalanced panel allows exit. For this case, the exit rate from the data is added as a moment and the flow cost  $\Gamma$  is estimated as well. The fit is not as good as the baseline model. The flow cost is estimated at about 3% of income. The estimated model matches turnover in the highest quintile and creates a substantial return differential, much larger than in the data. But, the exit rate is not matched well by the estimated model.

For the cross section moments, the best fit is obtained without unemployment risk. The discount factor is below the baseline estimate and the risk aversion is slightly lower as well. Note that  $\psi$  is quite close to one so that the only operative adjustment costs are associated with direct trading costs. From the moments, the estimated model matches the highest quintile of turnover as well as the return differential as well.

The estimation results allow for irrational households following section 2.2. The results are reported in the bottom panel of Table (14) with the moments in Table (15). The fit of the moments was improved,

<sup>30</sup>We do not include new entry.

Table 15: Data and Model Moments: BO

case	t1	t2	t3	t4	t5	DR	exit rate
Data	0.0019	0.0124	0.0289	0.0598	0.2149	-0.0046	0.0120
Rational: Time Series,balanced							
No UR	0.0129	0.0248	0.0364	0.0496	0.2148	-0.0032	na
UR	0.0047	0.0184	0.0311	0.0454	0.2128	-0.0028	na
Rational: Time Series,unbalanced							
No UR	0.0061	0.0105	0.0157	0.0249	0.2176	-0.0060	0.0010
UR	0.0048	0.0083	0.0143	0.0220	0.2133	-0.0079	0.0005
Rational: Cross Section							
No UR	0.0022	0.0112	0.0268	0.0609	0.2126	-0.0049	na
UR	0.0000	0.0022	0.0161	0.0401	0.2120	-0.0052	na
Irrational: Time Series							
$\tilde{\delta}$	0.0085	0.0211	0.0328	0.0568	0.2160	-0.0042	na
$\tilde{\sigma}_R$	0.0079	0.0225	0.0334	0.0557	0.2136	-0.0028	na

This table reports data and simulated 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.

relative to the rational time series case with unemployment risk, for two specifications. This improvements arose only when both parameters and beliefs deviated from the baseline. That is, if parameters are held fixed at their rational estimates, then allowing beliefs alone to vary did not improve the fit. The introduction of irrational beliefs did not improve the fit relative to the cross section case with rational households.

First, in the case labeled “ $\tilde{\delta}$ ” the household’s beliefs about unemployment risk was estimated to be 0.0368 compared to the baseline model assumption of 0.0140. With this form of irrationality, the estimated risk aversion is lower as is the adjustment cost. From the moments, this belief about unemployment risk brings the turnover rates in the fourth and fifth quintiles closer to the data and bring the return differential much closer to the data.

In the case of a more variable return process,  $\tilde{\sigma}_R = 0.0958$  compared to 0.0439 in the data, the model fit improved as well. In this case as well, the estimated risk aversion and adjustment costs are slightly lower than in the no UR baseline. The fit improved because of the larger differential in returns and increased turnover in the highest quintiles.

One way to evaluate the significance of the irrational beliefs is to compare the allocation under the resulting decision rule with the consumption profile of a rational agent, all else the same.<sup>31</sup> For this exercise, we focused on the case in which irrational agents overstate the probability of unemployment. This leads them to have higher precautionary savings relative to rational agents.

Specifically, first, using the decision rules of rational agents, a 4000x2000 panel is generated under the assumption of a constant stock return at its mean. The length of time is long enough to ensure the economy reaches the ergodic joint distribution of income and asset holdings. Second, 5000 random draws are taken from the ergodic distribution as initial states. Agents with positive stock holdings are selected (1904 of 5000) since these agents would have a stock account with a broker, as in the data. For each of these 1904 agents, 50 different paths of income are simulated for 60 months. These draws are data consistent, not consistent with the beliefs of the irrational agents. The return process is fixed from the actual data over the sample period of December 1991 to November 1996. Under these return and income processes, 60 months consumption profiles are generated from the rational model (CR) and the irrational model (CIR). At the end

<sup>31</sup>Evaluation of statistical significance is possible only with the proper weighting matrix. To us, the economic significance is more revealing.

of the 60 months, the asset positions are recorded from these two decision models. Finally, the consumption compensation is calculated for each of the simulated paths.

The goal was to calculate the compensating differential given to rational agents if they used the irrational decision rules. Compensation of consumption is calculated by solving for  $\chi$  that satisfies

$$\sum_{t=1}^{60} \beta^{t-1} U(C_t^R * \chi) + \beta^{60} \mathbb{E}_{y', R' | y, R} V(S_{61}^R) = \sum_{t=1}^{60} \beta^{t-1} U(C_t^{IR}) + \beta^{60} \mathbb{E}_{y', R' | y, R} V(S_{61}^{IR}) \quad (20)$$

where  $U(\cdot)$  is the utility function for the rational agent, rational expectation is taken over income and return realizations, and  $V(\cdot)$  is the value function of the rational agent at the end of the 60 month window evaluated in the terminal state for type  $j = R, IR$ , denoted  $S_{61}^j$ .

Averaging over all agents with initial stock holdings,  $\chi$  is around 99.4%, which means a 0.6% compensation of consumption for the rational agent to use the decision rules of the irrational agents.<sup>32</sup> With monthly income of about \$6,000, this is about \$36 or less than the fixed cost of a single trade.

## 7 Conclusion

The goal of this paper was to assess the claim made by Barber and Odean (2000) that the patterns of returns as a function of portfolio turnover was consistent with overconfident agents and inconsistent with rational traders. The approach uses a household dynamic optimization problem, emphasizing both the extensive and intensive margins of portfolio choice, as a framework for the analysis. In our model, portfolio adjustment costs created both inaction and large turnover and also drive a wedge between gross and net returns.

Parameters are estimated to match moments associated with turnover rates and net return patterns. We estimated models with both rational and irrational agents. For the latter, we study specifications in which agents either hold beliefs about income or returns that are not consistent with the data. We also study cases in which agents receive signals of future returns that are false.

We reach two conclusions. First, models with rational agents can match the turnover moments emphasized in Barber and Odean (2000) and calculated in a model consistent manner. Second, introducing various forms of irrationality do influence household choices, but does not improve the fit of the model with respect to the turnover and net return moments.

There is one case though where this conclusion must be qualified. If we match the moments taken directly from the Barber and Odean (2000) study, there are two cases in which some irrationality in beliefs creates a better fit. This occurs only when the moments are taken from a time series.

Does this finding support their conclusion that:

These results are consistent with models where trading emanates from investor overconfidence,  
but are inconsistent with models where trading results from rational expectations.

We think not for a couple of reasons. First, as our analysis has shown, the patterns of turnover and net return associated with transactions between stock and bond accounts are completely consistent with the optimizing behavior of rational households. Thus any irrationality lies in portfolio rebalancing that is excluded from our model but included in their moments. Second, the deviations from rational beliefs are minor, as measured by consumption compensation to rational agents for adopting the decision rules of the irrational agents. Third

<sup>32</sup>To avoid selection bias, here we are averaging over all agents including those who maintained a stock account through the sample as well as those who did not.

these deviations amount to increasing the volatility of labor and/or financial income and thus point to some missing elements of risk excluded by our simple model of portfolio choice. Fourth, there is no evidence of overconfidence in the form of excessive beliefs in the serial correlation of returns or the news provided by a broker.

The analysis has focused entirely on the intensive margin with respect to portfolio adjustment, as well as the choice of whether to adjust or not. There is an important omitted margin: to open an account with direct holdings or not.

Further, and perhaps more importantly, it might be that overconfidence plays a role in the decision regarding direct holding itself. As noted by Barber and Odean (2000), the actual return process in their data set is not that much different from that obtained through indirect holdings. This naturally leads to the question of why agents hold stocks directly at all.

## 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.
- BARBER, B. M., AND T. ODEAN (2001): “Boys will be boys: Gender, overconfidence, and common stock investment,” *The Quarterly Journal of Economics*, 116(1), 261–292.
- 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.
- KYLE, A. S., AND F. A. WANG (1997): “Speculation duopoly with agreement to disagree: Can overconfidence survive the market test?,” *The Journal of Finance*, 52(5), 2073–2090.
- ODEAN, T. (1998): “Volume, volatility, price, and profit when all traders are above average,” *The Journal of Finance*, 53(6), 1887–1934.

## 8 Appendix

This appendix provides additional information about the calculation of the income process and other moments.

## 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 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

$$\tilde{y}_t^a = \rho^a \tilde{y}_{t-1}^a + \epsilon_t^a. \quad (21)$$

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  $\frac{\text{flow}(e \rightarrow u + e \rightarrow n)}{\text{employmentlevel}}$ .

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.<sup>33</sup> In logs,  $y_{ue}^m = \log(6)$ . 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  $y_{uu}^m = \log(0.4 \times 6)$ .

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

$$\tilde{y}_t^m = \rho^m \tilde{y}_{t-1}^m + \epsilon_t^m. \quad (22)$$

In levels, this corresponds to

$$y_t^m = \mu_m(1 - \rho^m) + \rho^m y_{t-1}^m + \epsilon_t^m \quad (23)$$

Since we assume that the innovation  $\epsilon_t$  follows normal distribution of mean 0 and standard deviation  $\sigma_\epsilon$ ,  $Y^m$  follows log normal distribution of parameters  $(\mu_m, \sigma_y^m)$ , where  $\sigma_y^2 = \frac{\sigma_\epsilon^2}{1 - \rho^2}$ . Hence,

$$6 \equiv \mathbb{E}Y^m = e^{\mu_m + \frac{\sigma_y^2}{2}} \rightarrow \mu^m = \log(6) - \frac{\sigma_y^2}{2} \quad (24)$$

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

<sup>33</sup>This corresponds to the monthly average income of \$6000, annual \$72000 for stock holders.

### 8.1.1 Estimation

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

1. guess a vector of parameters  $\rho^m, \sigma_\epsilon^m$
2. simulate a panel of level of monthly incomes, aggregate to a panel of annual incomes (level)

$$y^a = \log\left(\sum_{i=1}^{12} \exp(y_i^m)\right) \quad (25)$$

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

$$y_t^a = \mu_a(1 - \rho^a) + \rho^a y_{t-1}^a + \epsilon_t^a \quad (26)$$

and calculate the variance of the residual.

4. compare the  $\rho^a$  and  $\sigma_\epsilon^a$  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

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.

### 8.2.1 Data Sets 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.

- 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 Cusip 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.<sup>34</sup> 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.<sup>35</sup> 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.

<sup>34</sup>This leaves us with 13,668,775 observations.

<sup>35</sup>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 purchases (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.<sup>36</sup> 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*.

<sup>36</sup>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.

### 8.2.3 The Time Series Panel

We start with "tohhlist\_wretnew0928.dta". First, drop all observations with non-positive positions (both current month and last month positions). Second, only keep households who opened all accounts before February 1991 using "keeplist.dta". Last, only keep households who always have reports of positions from January 1991 to November 1996: 71 months using "nm0927.dta"

#### Time-Series Averages

For each household, generate time-averages of: turnover, beginning of the month position, net return and gross return. Generate 4 percentiles (p20,p40 p60 and p80). Classify households into Quintile 1 if the household average turnover is in  $[0, p20]$ , into Quintile 2 if it is in  $(p20, p40]$ , into Quintile 3 if it is in  $(p40, p60]$ , into Quintile 4 if it is in  $(p60, p80]$  and into Quintile 5 if it is above p80. Compute the mean and median of average turnover in each quintile, as well as positions and net returns. <sup>37</sup>

Table 16: the Balanced Panel Time Series

	Q1	Q2	Q3	Q4	Q5	DR
N.	2903	2529	2716	2716	2716	n.a.
mean Position	25993.79	97974.06	68047.91	65173.57	73513.23	n.a
mean TO	0	.0039409	.0133699	.0290816	.1154604	n.a
median TO	0	.0038648	.013247	.0283326	.0689786	n.a
mean GR	1.010276	1.012712	1.011621	1.010451	1.011404	0.001128
mean NR	1.010276	1.012573	1.011277	1.009708	1.009258	-0.001018
median NR	1.012228	1.012817	1.01244	1.012154	1.011238	-0.00099

We also report the unbalanced one here since the weighting as changed compared to before. The unbalance panel still only keep people who opened account before 1991 February but doesn't require households to have reports for positions throughout the sample period.

Table 17: the Unbalanced Panel Time Series

	Q1	Q2	Q3	Q4	Q5	DR
N.	18,918	3,627	11,272	11,272	11,272	n.a.
mean Position	19563.53	98007.75	59175.8	48063.26	50127.36	n.a
mean TO	0	.0025307	.0129559	.0357216	.2171676	n.a
median TO	0	.0025367	.0127094	.0345575	.0929772	n.a
mean GR	1.010136	1.011324	1.010707	1.010467	1.01383	0.003694
mean NR	1.010136	1.011164	1.01032	1.003104	1.003719	-0.006417
median NR	1.011759	1.01184	1.011291	1.01119	1.010631	-0.001128

<sup>37</sup>Save the data as "TSbaldata.dta" later used in producing the var-cov matrix

### The Variance Covariance Matrix of the Balanced Panel of Time Series

We use bootstrap. Fixing seeds for each bootstrap exercise, we draw a random sub-sample (one tenth of the whole sample) without replacement from the balanced panel, TSbalddata.dta. Then we treat this sub-sample the same way as in the original one in calculating the 5 quintiles' average turnover and mean net return differential between Q1 and Q5. After repeating 100 exercises, we calculate the standard deviation for the average turnover and mean net return differential, as well as the correlation between each pair of them. The covariance is computed as the product of correlation and the standard deviation of two variables for each pair. In the end, the diagonal of the matrix are the variances of each variable, and off diagonals are the covariances between each pairs of variable. The matrix is

$$\begin{bmatrix} 2.82e-12 & 1.93e-10 & 2.40e-10 & 3.00e-10 & 2.39e-09 & 5.02e-10 \\ 1.93e-10 & 9.67e-08 & 1.72e-07 & 1.84e-07 & 1.15e-06 & 9.30e-08 \\ 2.40e-10 & 1.72e-07 & 5.02e-07 & 6.75e-07 & 2.39e-06 & 2.37e-07 \\ 3.00e-10 & 1.84e-07 & 6.75e-07 & 1.55e-06 & 7.58e-06 & 7.46e-07 \\ 2.39e-09 & 1.15e-06 & 2.39e-06 & 7.58e-06 & 5.60e-04 & -1.12e-06 \\ 5.02e-10 & 9.30e-08 & 2.37e-07 & 7.46e-07 & -1.12e-06 & 1.44e-05 \end{bmatrix}$$

#### 8.2.4 The Cross Section Panel

For the balanced panel, do the sample selection in the same way as the time series panel.

For each month, put households in quintiles as in the time series panel. Then take time-averages or medians of gross and net returns, turnovers, and positions for each quintile.

Table 18: the Balanced Panel Cross Section

	Q1	Q5	DR
N.	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
mean NR	1.010969	1.010044	-0.000925
median NR	1.011111	1.011203	0.000092

Correspondingly, the unbalanced one is

Table 19: the Unbalanced Panel Cross Section

	Q1	Q5	DR
N.	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
mean NR	1.009731	.977579	-0.032152
median NR	1.009036	1.009168	0.000132