A Derivation of the Likelihood Function

We assume that, each time a subject of class $j \in \{F, M, S\}$ has to make a decision to invest or not invest, given her type and the observed history of the other subject's behavior, she makes the decision prescribed by strategy $j \in \{F, M, S\}$ with probability $1 - \varepsilon$ and makes an "error" with probability $\varepsilon \in [0, 0.5]$. Errors are assumed to be i.i.d. across types and histories, trials, and subjects.

Table 1 shows the probability that a class F subject invests, or makes a final decision never to invest, conditional on her type, on the history, and on the fact that the other subject's behavior allows the history to occur. To understand how the table is constructed, consider the following examples. The probability that an F subject of type $(0, L)$ invests after the history $\{1\}$ is $(1 - \varepsilon)^2$, since she acted according to her strategy class twice: by not investing in round 1 and then by investing after the other subject invests in round 1. The probability that an F subject of type $(0, L)$ invests after the history $\{1, 0\}$ is $\varepsilon(1 - \varepsilon)^2$, since she acted according to her class by not investing in round 1, then she made an "error" by not investing after $\{1\}$, and finally she acted according to her class by recovering from her error and investing after $\{1, 0\}$. The probability that an F subject of type $(0, L)$ ends up not investing after experiencing history $\{0\}$ is $(1-\varepsilon)^2$, since she made two type-consistent decisions: she did not invest either after $\{\}$ or after $\{0\}$.

For M subjects and S subjects, we can construct tables analogous to Table 1. An M subject behaves in the same way as an F subject, except when her type is $(1, H)$. Thus, columns 1, 2, and 4 are as in Table 1, but the nine entries in column 3 should instead be: $(1 - \varepsilon), \varepsilon^2, \varepsilon(1 - \varepsilon), (1/2)\varepsilon(1 - \varepsilon), \varepsilon^2(1 - \varepsilon), (1/4)\varepsilon(1 - \varepsilon), \varepsilon(1 - \varepsilon), \varepsilon^3$, and $(1/4)\varepsilon(1 - \varepsilon)$. An S subject behaves in the same way as an F subject when her type is $(0, H)$ or $(1, L)$. Furthermore, her behavior when her type is $(0, L)$ is identical to her behavior when her type is $(0, H)$, and her behavior when her type is $(1, H)$ is identical to her behavior when her type is $(1, L)$. Therefore, columns 1 and 2 are identical to column 1 in Table 1, and columns 3 and 4 are identical to column 4 in Table 1.

Before constructing the likelihood function, we need some more notation. We number all of a subject's trials by $t = 1, 2, ..., 24$. Let $\mathbf{B}_i^t$ denote the full behavior

History	(0,H)	(0,L)	(1,H)	(1,L)
{}	ε	ε	ε	$(1-\varepsilon)$
{0}	$\varepsilon(1-\varepsilon)$	$\varepsilon(1-\varepsilon)$	$\varepsilon(1-\varepsilon)$	$\varepsilon(1-\varepsilon)$
{1}	$\varepsilon(1-\varepsilon)$	$(1-\varepsilon)^2$	$(1-\varepsilon)^2$	$\varepsilon(1-\varepsilon)$
{0,1}	$\varepsilon(1-\varepsilon)^2$	$\frac{1}{2}(1-\varepsilon)^2$	$\frac{1}{2}(1-\varepsilon)^2$	$\varepsilon^2(1-\varepsilon)$
{1,0}	$\varepsilon(1-\varepsilon)^2$	$\varepsilon(1-\varepsilon)^2$	$\varepsilon(1-\varepsilon)^2$	$\varepsilon^2(1-\varepsilon)$
{0,1,0}	$\varepsilon(1-\varepsilon)^3$	$\frac{1}{4}(1-\varepsilon)^2$	$\frac{1}{4}(1-\varepsilon)^2$	$\varepsilon^3(1-\varepsilon)$
no {0}	$(1-\varepsilon)^2$	$(1-\varepsilon)^2$	$(1-\varepsilon)^2$	ε^2
no {1,0}	$(1-\varepsilon)^3$	$\varepsilon^2(1-\varepsilon)$	$\varepsilon^2(1-\varepsilon)$	ε^3
no {0,1,0}	$(1-\varepsilon)^4$	$\frac{1}{4}(1-\varepsilon)^2$	$\frac{1}{4}(1-\varepsilon)^2$	ε^4

Table 1: Probability of an F subject's behavior

of subject i during (i's) trial t. By full behavior, we mean the round in which she invests, if at all. We formalize $\mathbf{B}_i^t$ as a four dimensional vector of zeros and ones. $\mathbf{B}_i^t = (0,0,0,0)$ signifies that the subject did not invest, the vector $\mathbf{B}_i^t = (0,0,1,0)$ signifies that she invested in round 3, and so on. Let $-i_t$ denote the subject matched with subject i during trial t, and let $\mathbf{B}_{-i}^t$ denote the full behavior of subject $-i_t$ during trial t.[1] Denote the behavior of subject i over all trials as $\mathbf{B}_i$, where we have $\mathbf{B}_i = (\mathbf{B}_i^1, ..., \mathbf{B}_i^t, ..., \mathbf{B}_i^{24})$, and denote the behavior of all the subjects matched with subject i (during the trials they are matched with i) as $\mathbf{B}_{-i}$, where we have $\mathbf{B}_{-i} = (\mathbf{B}_{-i}^1, ..., \mathbf{B}_{-i}^t, ..., \mathbf{B}_{-i}^{24})$. Finally, let $\mathbf{T}_i^t \in \{(0,H),(0,L),(1,H),(1,L),-1\}$ denote subject i's type during trial t, where type -1 means that subject i was sitting out or bankrupt, let $\mathbf{T}_i = (\mathbf{T}_i^1, ..., \mathbf{T}_i^t, ..., \mathbf{T}_i^{24})$, and let $\mathbf{T} = (\mathbf{T}_1, ..., \mathbf{T}_i, ...\mathbf{T}_n)$.

From Table 1, or the analogous tables corresponding to M subjects and S subjects, the probability of $\mathbf{B}_i^t$ is determined, given that the subject is of strategy class j, type $\mathbf{T}_i^t$, and given that the other subject's behavior is $\mathbf{B}_{-i}^t$.[2] We denote this probability, which also depends on the parameter ε, as

$$\Pr(\mathbf{B}_i^t | j, \mathbf{T}_i^t, \mathbf{B}_{-i}^t; \varepsilon). \tag{A.1}$$

For example, suppose that in trial t, subject i is type $(1,H)$ and the other subject

[1] Define $\mathbf{B}_i^t = \mathbf{B}_{-i}^t = (-1,-1,-1,-1)$ if i did not participate in trial t (either because she sat out or because she went bankrupt).

[2] If a subject is sitting out trial t or has gone bankrupt, then $\mathbf{B}_i^t = \mathbf{B}_{-i}^t = (-1,-1,-1,-1)$ with probability one.

invests in round 1, $\mathbf{T}_i^t = (1, H)$ and $\mathbf{B}_{-i}^t = (1, 0, 0, 0)$. If subject i is an F subject, then the probability of $\mathbf{B}_i^t = (1, 0, 0, 0)$ is ε (row 1, column 3 of Table 1), the probability of $\mathbf{B}_i^t = (0, 1, 0, 0)$ is $(1 - \varepsilon)^2$ (row 3, column 3 of Table 1), the probability of $\mathbf{B}_i^t = (0, 0, 1, 0)$ is $\varepsilon(1 - \varepsilon)^2$ (row 5, column 3 of Table 1), and the probability of $\mathbf{B}_i^t = (0, 0, 0, 0)$ is $(1-\varepsilon)\varepsilon^2$ (row 8, column 3 of Table 1). Given that the other subject invests in round 1, investing in round 4 is impossible, so the probability of $\mathbf{B}_i^t = (0, 0, 0, 1)$ is zero.

The probability that subject i chooses behavior $\mathbf{B}_i$, given that her strategy class is j, given her type realizations $\mathbf{T}_i$, and given the behavior of the other subjects she faces is $\mathbf{B}_{-i}$, is given by

$$\Pr(\mathbf{B}_i | j, \mathbf{T}_i, \mathbf{B}_{-i}; \varepsilon) = \prod_{t=1}^{24} \Pr(\mathbf{B}_i^t | j, \mathbf{T}_i^t, \mathbf{B}_{-i}^t; \varepsilon). \tag{A.2}$$

Thus, we can compute the probability that subject i chooses behavior $\mathbf{B}_i$, given her type realizations $\mathbf{T}_i$, and given that the behavior of the other subjects she faces is $\mathbf{B}_{-i}$,[3]

$$\Pr(\mathbf{B}_i | \mathbf{T}_i, \mathbf{B}_{-i}; p_F, p_M, p_S, \varepsilon) = \sum_{j \in \{F,M,S\}} p_j \Pr(\mathbf{B}_i | j, \mathbf{T}_i, \mathbf{B}_{-i}; \varepsilon). \tag{A.3}$$

We now show that the likelihood function is given by[4]

$$\Pr(\mathbf{B} | \mathbf{T}; p_F, p_M, p_S, \varepsilon) = \prod_{i=1}^{n} \Pr(\mathbf{B}_i | \mathbf{T}_i, \mathbf{B}_{-i}; p_F, p_M, p_S, \varepsilon). \tag{A.4}$$

Let $j_i \in \{F, M, S\}$ denote subject i's strategy class. Label all trials, $m = 1, ..., M$, and let $m(1)$ and $m(2)$ be the two subjects in trial m, where $m(1)$ is the subject with the lower identification number. The likelihood function (suppressing the dependence

[3]The other subjects' behavior does not affect the probability of being in class F, M, or S, except for the unlikely event that a subject has gone bankrupt, so we have $\mathbf{B}_{-i}^t = (-1, -1, -1, -1)$. Not only were bankruptcies exceedingly rare, but the probabilities of each strategy class (conditional on bankruptcy) would not change very much. The stronger inference is that the subject made many "errors." We ignore this complication.

[4]The likelihood function is also implicitly conditional on the realized matching of subjects.

on the realized types, realized matchings, and $\boldsymbol{\theta}$) is given by

$$\Pr(\mathbf{B}) = \sum_{j_1,\dots,j_n} p_{j_1} \cdots p_{j_n} \Pr(\mathbf{B}|j_1, ..., j_n). \tag{A.5}$$

Errors are independent, so behavior in one trial, conditional on the strategy classes of the subjects in that trial, is independent of behavior in any other trial. Therefore, we have

$$\Pr(\mathbf{B}|j_1, ..., j_n) = \prod_{m=1}^{M} \Pr(\mathbf{B}_{m(1)}, \mathbf{B}_{m(2)}|j_{m(1)}, j_{m(2)}). \tag{A.6}$$

Now, we claim that, for subjects 1 and 2 in a particular trial (this is without loss of generality), we can write

$$\Pr(\mathbf{B}_1, \mathbf{B}_2|j_1, j_2) = \Pr(\mathbf{B}_1|j_1, \mathbf{B}_2) \Pr(\mathbf{B}_2|j_2, \mathbf{B}_1). \tag{A.7}$$

To verify the claim, let $\mathbf{B}_i^r$ denote the behavior of subject i during round r (one if the subject invests during that round, zero otherwise). Then we have (sometimes suppressing the dependence on j_1 and j_2)

$$\begin{aligned}
\Pr(\mathbf{B}_1, \mathbf{B}_2|j_1, j_2) &= \Pr(\mathbf{B}_1^1, \mathbf{B}_2^1) \Pr(\mathbf{B}_1^2, \mathbf{B}_2^2|\mathbf{B}_1^1, \mathbf{B}_2^1) \Pr(\mathbf{B}_1^3, \mathbf{B}_2^3|\mathbf{B}_1^1, \mathbf{B}_2^1, \mathbf{B}_1^2, \mathbf{B}_2^2) \cdot \\
&\quad \Pr(\mathbf{B}_1^4, \mathbf{B}_2^4|\mathbf{B}_1^1, \mathbf{B}_2^1, \mathbf{B}_1^2, \mathbf{B}_2^2, \mathbf{B}_1^3, \mathbf{B}_2^3) \\
&= \Pr(\mathbf{B}_1^1) \Pr(\mathbf{B}_2^1) \Pr(\mathbf{B}_1^2|\mathbf{B}_1^1, \mathbf{B}_2^1) \Pr(\mathbf{B}_2^2|\mathbf{B}_1^1, \mathbf{B}_2^1) \Pr(\mathbf{B}_1^3|\mathbf{B}_1^1, \mathbf{B}_2^1, \mathbf{B}_1^2, \mathbf{B}_2^2) \cdot \\
&\quad \Pr(\mathbf{B}_2^3|\mathbf{B}_1^1, \mathbf{B}_2^1, \mathbf{B}_1^2, \mathbf{B}_2^2) \Pr(\mathbf{B}_1^4|\mathbf{B}_1^1, \mathbf{B}_2^1, \mathbf{B}_1^2, \mathbf{B}_2^2, \mathbf{B}_1^3, \mathbf{B}_2^3) \cdot \\
&\quad \Pr(\mathbf{B}_2^4|\mathbf{B}_1^1, \mathbf{B}_2^1, \mathbf{B}_1^2, \mathbf{B}_2^2, \mathbf{B}_1^3, \mathbf{B}_2^3) \\
&= \Pr(\mathbf{B}_1|j_1, j_2, \mathbf{B}_2) \Pr(\mathbf{B}_2|j_1, j_2, \mathbf{B}_1) \qquad \text{(A.8)}
\end{aligned}$$

The behavior of one subject in a trial depends on the other subject's behavior but not on the other subject's strategy class (given the other's behavior), so the claim follows.

From (A.6) and (A.7), we have

$$\begin{aligned}
\Pr(\mathbf{B}|j_1,...,j_n) &= \prod_{m=1}^{M} \Pr(\mathbf{B}_{m(1)}|j_{m(1)},\mathbf{B}_{m(2)}) \Pr(\mathbf{B}_{m(2)}|j_{m(2)},\mathbf{B}_{m(1)}) \\
&= \prod_{i=1}^{n} \Pr(\mathbf{B}_i|j_i,\mathbf{B}_{-i}). \qquad \text{(A.9)}
\end{aligned}$$

Substituting (A.9) into (A.5), we have

$$\begin{aligned}
\Pr(\mathbf{B}) &= \sum_{j_1,...,j_n} \left[p_{j_1} \cdots p_{j_n} \prod_{i=1}^{n} \Pr(\mathbf{B}_i|j_i,\mathbf{B}_{-i}) \right] \\
&= \sum_{j_1,...,j_n} \left[\prod_{i=1}^{n} p_{j_i} \Pr(\mathbf{B}_i|j_i,\mathbf{B}_{-i}) \right] \\
&= \prod_{i=1}^{n} \left[\sum_{j \in F,M,S} p_j \Pr(\mathbf{B}_i|j,\mathbf{B}_{-i}) \right] \\
&= \prod_{i=1}^{n} \Pr(\mathbf{B}_i|\mathbf{B}_{-i}), \qquad \text{(A.10)}
\end{aligned}$$

which is what we wanted to show.