

Estimating a Parsimonious Model of Inequality Aversion in Stackelberg Duopoly Experiments^{*}

SAU-HIM PAUL LAU † and FELIX LEUNG ‡

† *School of Economics and Finance, University of Hong Kong, Pokfulam Road, Hong Kong (e-mail: laushp@hkucc.hku.hk)*

‡ *Hong Kong Institute of Economics and Business Strategy, University of Hong Kong, Pokfulam Road, Hong Kong (e-mail: leungcyf@netvigator.com)*

Abstract:

In the Stackelberg duopoly experiments in Huck *et al.* (2001), nearly half of the followers' behaviors are inconsistent with conventional prediction. We use a test in which the conventional self-interested model is nested as a special case of an inequality aversion model. Maximum likelihood methods applied to the Huck *et al.* (2001) data set reject the self-interested model. We find that almost 40% of the players have disadvantageous inequality aversion that is statistically different from zero and economically significant, but advantageous inequality aversion is relatively unimportant. These estimates provide support for a more parsimonious model with no advantageous inequality aversion.

JEL Classification numbers: C91, D63

Keywords: Stackelberg duopoly experiments, inequality aversion, maximum likelihood estimates

* We thank Vai-Lam Mui, Hans Normann, participants of the Econometric Society North American Summer Meeting (Pittsburgh) and the Econometric Society European Meeting (Milan), an anonymous referee and two Editors for helpful comments, Steffen Huck for providing the data, Philip Ng for research assistance, and Robert Canwell for editorial help. Financial support from the Hong Kong Institute of Economics and Business Strategy at the University of Hong Kong is gratefully acknowledged.

I. INTRODUCTION

The Stackelberg model of oligopolistic interaction is one of the workhorse models in industrial organization. It provides a tractable framework to analyze strategic asymmetry in which a subset of players have commitment power. In the simplest case of a duopoly, one player (the Stackelberg leader) makes a choice first, and then the other player (the Stackelberg follower) moves after observing the leader's choice. Many interesting phenomena, such as market entry and R&D races, have been analyzed through this model of sequential moves.

Motivated by Daughety's (1990) prediction that total output and welfare are higher in a Stackelberg market than in the corresponding Cournot market in which players move simultaneously, Huck *et al.* (2001) conduct one of the first experimental studies of the Stackelberg game. While their results confirm the prediction of Daughety (1990), they also find that, interestingly, the outcome predicted by the conventional self-interested model (in which each player cares only about her own material payoff) is rarely observed. Specifically, in one set of their random-matching experiments, in which anonymous subjects are randomly assigned the roles of leaders and followers in a Stackelberg game and the design ensures that each pair of players interacts once, the behavior of the Stackelberg followers is inconsistent with conventional prediction. Conventional theory predicts that a follower, after observing the leader's output level, will react according to a downward-sloping best response function with particular values of the slope and intercept parameters. However, the followers behave according to this best response function in only 50.9% of the experiments.¹ Moreover, Huck *et*

¹This number, which is not found in Huck *et al.* (2001), is calculated by us based on the data set provided by Steffen Huck. Note that the major objective of Huck *et al.* (2001) is to use experimental evidence to compare Cournot and Stackelberg quantity-setting duopoly markets. They study the random-matching and fixed-matching treatments of both markets, and they mainly present the average level of output of the leaders and the followers. However, we focus on individual behavior

al. (2001, Table 6) estimate a linear regression model of the followers' choices, and find that the estimated parameters are significantly different from the conventional predictions.

After presenting the experimental evidence, Huck *et al.* (2001) suggest that the inequality aversion model of Fehr and Schmidt (1999) may be helpful in explaining the Stackelberg followers' behavior. Following Fehr and Schmidt (1999), they assume that some players have non-standard preferences and that a player in this group cares about the absolute level of her payoff as well as how her payoff compares with her opponent's. They then argue that, compared with the inequality aversion model, the conventional self-interested model is less consistent with the Stackelberg followers' behavior in their experiments. In their specification, the null hypothesis of the conventional self-interested model is tested against an alternative model with a linear best response function different from that of the self-interested model. However, it is questionable whether the Fehr-Schmidt model actually delivers a linear best response function. A related drawback of their procedure is that when the null hypothesis is rejected, it is not clear how the estimated parameters are related to the underlying behavioral structure.

We think that conventional econometric methods, when applied appropriately, will be useful in understanding whether individuals mainly care about their own payoffs or payoff comparison matters significantly. There are two objectives in this paper: one methodological and one empirical. Methodologically, our objective is to derive a testing procedure such that the null hypothesis of the conventional self-interested model is nested as a special case of an inequality aversion model. Empirically, we in one set of their experiments—the Stackelberg games with random matching ('STACKRAND' in their notation). We do not use their fixed-matching data set (in which the same pairs of players interact over several periods), because the theoretical predictions developed in this paper focus on a situation in which the players interact only once.

apply the derived test to see which of the above two models is more consistent with the data. It turns out that the hypothesis of the conventional self-interested model is rejected, and we further use the estimates of the inequality aversion model to simplify the theoretical model. We address both of these objectives in the context of the Stackelberg duopoly game (as in Huck *et al.*, 2001).

To examine these issues, we need to first derive the likelihood function of a sample of Stackelberg followers' choices based on the inequality aversion model. We find it helpful to consider a simplified version of the Fehr-Schmidt model that retains the model's essential idea but allows meaningful estimation. In a two-player (Stackelberg leader and follower) game, Fehr and Schmidt (1999) assume that players with non-standard preferences have the following utility function:

$$U_i = U_i(z_i, z_j) = z_i - \alpha_i \max\{z_j - z_i, 0\} - \beta_i \max\{z_i - z_j, 0\}, \quad (1)$$

where $0 \leq \beta_i < 1$, $\beta_i \leq \alpha_i$, $i, j = L, F$ with $i \neq j$, L is the Stackelberg leader, F is the Stackelberg follower, z_i is player i 's material payoff, α_i is her disadvantageous inequality aversion parameter, and β_i is her advantageous inequality aversion parameter. After reviewing previous experimental studies, Fehr and Schmidt (1999, Table III) focus on cases where the disadvantageous inequality aversion parameter takes four values: 0 (with 30%), 0.5 (with 30%), 1 (with 30%) and 4 (with 10%), and the advantageous inequality aversion parameter takes three values: 0 (with 30%), 0.25 (with 30%) and 0.6 (with 40%). Huck *et al.* (2001) also use these distributions in their analysis. While Fehr and Schmidt's (1999) conjectured distributions can in principle be incorporated in the following analysis, we believe it is more interesting to estimate the parameter values than to impose them. However, because the likelihood function is highly nonlinear (see Section IV), computational problems are likely to arise when there are too many parameters to be estimated, especially if the inequality

aversion parameters of different types of players do not differ much.²

To deal with these econometric issues, our parsimonious version of the Fehr-Schmidt model follows their assumption that a subset of players are only interested in their own material payoff, and that the remaining players have the non-standard utility function given by (1). However, as in Fehr *et al.* (2007, p. 144), we simplify the Fehr-Schmidt model by assuming that players with non-standard preferences have the same disadvantageous inequality aversion parameter and the same advantageous inequality aversion parameter ($\alpha_i = a$ and $\beta_i = b$ for all players with non-standard preferences).³

To summarize, the simplified Fehr-Schmidt model considered in this paper is given by (1) and

$$\Pr(\alpha_i = a \ \& \ \beta_i = b) = p; \quad \Pr(\alpha_i = \beta_i = 0) = 1 - p, \quad (2)$$

where

$$0 \leq p < 1, \quad (3)$$

$$0 \leq b < 1, \quad (4)$$

²Fehr *et al.* (2007, footnote 17) give a four-type example based on a perfect correlation version of Fehr and Schmidt (1999). Instead of pre-specifying the inequality aversion parameters and the proportion of each type, as in Fehr *et al.* (2007), we can generalize this example by letting them be unknown constants to be estimated. In this case, it is assumed that proportion p_1 of the population have $\alpha_i = \beta_i = 0$, proportion p_2 have $\alpha_i = a_1 > 0$ and $\beta_i = b_1 > 0$, proportion p_3 have $\alpha_i = a_2 > a_1$ and $\beta_i = b_2 > b_1$, and proportion $(1 - p_1 - p_2 - p_3)$ have $\alpha_i = a_3 > a_2$ and $\beta_i = b_3 = b_2$. In this example, there are eight parameters (or nine parameters if we do not follow Fehr *et al.* (2007) to assume $b_3 = b_2$) to be estimated. The computational problems associated with the maximum likelihood estimation of this example are likely to be significant, especially because there are also the restrictions of $0 \leq \beta_i < 1$ and $\beta_i \leq \alpha_i$.

³Fehr *et al.* (2007) mention that their analysis would have been very tedious if they had used the four-type specification of Fehr and Schmidt (1999). As a result, they simplify the Fehr-Schmidt model and assume that 60% of the individuals have $\alpha_i = \beta_i = 0$, and 40% have $\alpha_i = 2$ and $\beta_i = 0.6$. While we also assume two types of individuals in this paper, we allow the parameters to be estimated.

and

$$b \leq a. \tag{5}$$

Using this model, we first derive the likelihood function of the followers' behavior. Using the derived likelihood function and the experimental results with randomly matched players in Huck *et al.* (2001),⁴ we obtain two major empirical findings. First, the null hypothesis of the self-interested model is clearly rejected against the alternative model based on inequality aversion, and the null hypothesis of no advantageous inequality aversion is not rejected. Second, we estimate that almost 40% of the Stackelberg followers have non-standard preferences, and the disadvantageous inequality aversion parameter is statistically different from zero and economically significant. These estimates are consistent with the properties of the Fehr-Schmidt model; moreover, they provide useful guidance if we want to further simplify it.

The rest of this paper is organized as follows. In Section II we summarize the behavior of the experimental Stackelberg followers in Huck *et al.* (2001) that conventional theory fails to explain, and we propose a parsimonious model based on Fehr and Schmidt (1999). In Section III we derive the followers' best response functions. In Section IV we provide estimation and hypothesis testing of the proposed model based on laboratory evidence in Huck *et al.* (2001). Section V provides the conclusions.

II. THE BEHAVIOR TO BE EXPLAINED, AND THE PROPOSED MODEL

In the experimental Stackelberg duopoly with quantity competition considered in Huck *et al.* (2001), a player—the Stackelberg leader L —commits her output level first. Knowing the choice of the leader, the other player—the Stackelberg follower

⁴Cox *et al.* (2007, Section 5) also perform statistical analysis with this data set to examine the importance of reciprocity.

F —selects her output. The profit (or material payoff) of player i is given by

$$z_i = z_i(q_i, q_j) = (\max\{d - q_i - q_j, 0\} - g) \times q_i, \quad (6)$$

where $i, j = L, F$ with $i \neq j$, q_i is the output level of player i , and $d > g \geq 0$. One interpretation of (6) is that the price of a homogenous product either depends linearly on the sum of the two players' outputs or is zero (if the sum of the outputs is too high), and g is the constant marginal cost. Note that except for the different roles of the leader and follower, the two players are symmetric with the same material payoff function.

Each player chooses the output level within a closed interval $q_i \in [0, \frac{d}{2}] \equiv [\underline{q}, \bar{q}]$. Note that within this interval, we have

$$\max\{d - q_i - q_j, 0\} - g = h - q_i - q_j, \quad (7)$$

where

$$h = d - g. \quad (8)$$

If $g > 0$, then $h - q_i - q_j$ (and thus, a player's material payoff) may be positive or negative.⁵ In Huck *et al.* (2001), $d = 30$ and $g = 6$.

We first summarize the conventional predictions regarding the above game. According to conventional theory, each player chooses output to maximize her material payoff, given the behavior of her opponent. The subgame-perfect equilibrium predicted by conventional theory is as follows. Observing the level of output q_L selected by the leader, the Stackelberg follower chooses q_F to maximize $z_F = (h - q_L - q_F) q_F$.

⁵We assume $\bar{q} = \frac{d}{2}$ in this paper, because Huck *et al.* (2001) use this value in their experiments. It can be shown that the best response functions of the Stackelberg followers are the same whether $\bar{q} = \frac{d}{2}$ or $\frac{h}{2}$, but the derivation would be simpler if we restrict $q_i \in [0, \frac{h}{2}]$, because $h - q_i - q_j$ is always non-negative in this case. In the Appendix, we discuss how the analysis of the simpler case ($\bar{q} = \frac{h}{2}$) guides our analysis of the more complicated case ($\bar{q} = \frac{d}{2}$).

It is straightforward to show that the best response function of the follower of this game is given by

$$q_F^S(q_L) = \frac{h}{2} - \frac{q_L}{2}. \quad (9)$$

We refer to the followers with ‘standard’ preferences as type S followers. The best response function of type S followers is given in the upper panel of Figure 1. Anticipating that a follower will respond according to (9), the leader chooses q_L to maximize $z_L(q_L, q_F^S(q_L))$. The optimal choice of the leader is given by $\frac{h}{2}$.

While the conventional theory may be useful for predicting behavior in some situations (such as R&D races in which history dictates who is the leader and who is the follower), it is inconsistent with the experimental evidence in Huck *et al.* (2001). We focus on the anomalies of the followers’ behavior in the Stackelberg duopoly experiments with random matching in their paper. The estimated response function of the followers is much flatter than predicted, and the slope coefficient is significantly different from the predicted value of -0.5 given in (9). It is also significantly different from the predicted value of -0.49 for the discretized game in Huck *et al.* (2001); see their footnote 9.

Huck *et al.* (2001) also discuss the anomalies of the Stackelberg leaders’ behavior. Conventional theory predicts that the Stackelberg leader would take the first-mover advantage and set the level of output at the profit-maximizing level ($\frac{h}{2}$). However, the leaders choose that value in less than 27.3% of the 220 trials. The prediction is even less accurate for experienced players (by looking at the outcomes of Round 9, the penultimate round), who choose the profit-maximizing level in 13.6% of the trials.⁶ While it is also interesting to examine whether the Stackelberg leaders’ behavior is consistent with the Fehr-Schmidt model, in this paper we choose to focus on the behavior of the experimental Stackelberg followers. This is because the predicted action

⁶Huck *et al.* (2001) mention that they prefer the penultimate round to the last round because of the possible end-game effects.

of the Stackelberg leader depends on her expectation of the follower’s type. Thus, any test about these predictions is a joint test of the underlying theoretical structure and the assumed expectations formation mechanism.⁷ However, the predictions of the followers’ choice (as in Section III) are purely based on the assumed behavioral model, since the follower has already observed the leader’s action. We believe it is more fruitful to first focus on clean predictions of the followers’ behavior, which are based solely on their preferences and not confounded by other auxiliary assumptions.

To explain the anomalies in Huck *et al.* (2001), we consider a simplified version of the inequality aversion model in Fehr and Schmidt (1999); see also Bolton (1991) and Bolton and Ockenfels (2000). This model, which has a relative payoff component, is particularly useful in explaining experimental evidence because the reference groups and outcomes in this context are reasonably clear. As mentioned in Fehr and Schmidt (1999, p. 822): ‘The subjects enter the laboratory as equals, they do not know anything about each other, and they are allocated to different roles in the experiment at random. Thus, it is natural to assume that the reference group is simply the set of subjects playing against each other and that the reference point, i.e., the equitable outcome, is given by the egalitarian outcome.’

In Fehr and Schmidt (1999) and Huck *et al.* (2001, Section 4), agents with non-standard preferences have the utility function given by (1). They also allow very general distribution of inequality aversion parameters (Table III of Fehr and Schmidt, 1999).

As mentioned in the Introduction, in order to provide a parsimonious explanation and to minimize computational problems associated with estimation, we simplify the Fehr-Schmidt model by assuming that proportion $1 - p$ of the players have standard preferences as defined by the material payoff function in (6) and the remaining proportion p have non-standard preferences (called type *NS* players), where all type *NS*

⁷This point has also been mentioned in, for example, Charness and Haruvy (2002).

players have the same disadvantageous inequality aversion parameter ($\alpha_i = a$) and the same advantageous inequality aversion parameter ($\beta_i = b$), with the restrictions (4) and (5).

III. BEST RESPONSE FUNCTIONS OF THE TWO TYPES OF FOLLOWERS

To examine the followers' behavior in this model, our first step is to transform (1), in which player i 's utility level depends on the players' material payoffs (z_i and z_j), into a form in which the utility level depends on the players' choice variables (q_i and q_j). Using (6) and (7), player i 's utility can be expressed in terms of the players' choice variables as follows:

$$\begin{aligned} V_i(q_i, q_j) &= U_i(z_i(q_i, q_j), z_j(q_j, q_i)) \\ &= (h - q_i - q_j) q_i - \alpha_i \max\{(h - q_i - q_j)(q_j - q_i), 0\} \\ &\quad - \beta_i \max\{(h - q_i - q_j)(q_i - q_j), 0\}. \end{aligned} \quad (10)$$

In the following analysis, we study the optimal response of a Stackelberg follower, after observing the leader's choice. As the two types of followers have different inequality aversion parameters, their best response functions are different.

The best response function of type S followers is given by (9), whereas the best response function of type NS followers is defined by

$$\begin{aligned} q_F^{NS}(q_L) &= \arg \max_{q_F} V_F(q_F, q_L) \\ &= \arg \max_{q_F} [(h - q_L - q_F) q_F - a \max\{(h - q_L - q_F)(q_L - q_F), 0\} \\ &\quad - b \max\{(h - q_L - q_F)(q_F - q_L), 0\}]. \end{aligned} \quad (11)$$

The analysis of the best response function of type NS followers is complicated by the fact that the utility loss due to disadvantageous inequality (resp. advantageous inequality) may or may not be relevant, depending on whether or not $(h - q_L - q_F)(q_L - q_F)$

is positive (resp. negative). As shown in the Appendix, the form of the best response function of type *NS* followers differs with respect to the level of q_L . For convenience in subsequent analysis, we label the intervals $[\underline{q}, \tilde{q}_{AB}]$, $[\tilde{q}_{AB}, \tilde{q}_{BC}]$, and $[\tilde{q}_{BC}, \bar{q}]$ as Interval A, B, and C, respectively,⁸ where

$$\tilde{q}_{AB} = \left(\frac{1-b}{3-2b} \right) h \quad (12)$$

and

$$\tilde{q}_{BC} = \left(\frac{1+a}{3+2a} \right) h. \quad (13)$$

If a leader chooses a low output level in Interval A, it is shown in the Appendix that it is not optimal for type *NS* followers to choose q_F in $[\underline{q}, q_L)$. As a result, disadvantageous inequality aversion is irrelevant for them. It can further be shown that if q_L is in Interval A, then the best response function of type *NS* followers is given by

$$q_F^{NS}(q_L) = \frac{h}{2} - \frac{q_L}{2(1-b)}. \quad (11a)$$

If a leader chooses a high output level outside Interval A, it can be shown (in the Appendix) that it is not optimal for a type *NS* follower to choose q_F in the interval $(q_L, \bar{q}]$. As a result, advantageous inequality aversion is irrelevant but disadvantageous inequality aversion becomes potentially relevant. There is a qualitative difference in the behavior of type *NS* followers when the leader chooses an output level outside Interval A. In the Appendix, it is shown that if q_L is in Interval B, the best response function of type *NS* followers is upward sloping, given by

$$q_F^{NS}(q_L) = q_L. \quad (11b)$$

⁸Note that \tilde{q}_{AB} in (12) is given by the intersection of (11a) and the 45-degree line ($q_F = q_L$), and \tilde{q}_{BC} in (13) is given by the intersection of (11c) and the 45-degree line; also see the lower panel of Figure 1. Note that \tilde{q}_{AB} belongs to both Intervals A and B, and \tilde{q}_{BC} belongs to both Intervals B and C. The overlapping endpoints of the different intervals do not cause problems because the functions we focus on in subsequent analysis— $q_F^{NS}(q_L)$ in (11a) to (11c)—turn out to be piecewise continuous.

On the other hand, if q_L is in Interval C, the best response function of type NS followers is downward sloping, given by

$$q_F^{NS}(q_L) = \frac{h}{2} - \frac{q_L}{2(1+a)}. \quad (11c)$$

The intuition of the difference in (11b) and (11c) is as follows. When advantageous inequality aversion is irrelevant, there are two components in a type NS follower's utility level according to (10)—she wants to maximize the first term (material payoff) but to minimize the second term (utility loss from disadvantageous inequality). To maximize the material payoff, the best response of the follower is given by (9). In Interval B, whether the follower uses (9) or (11b), the difference in material payoff is relatively minor (since the material payoff function is quite flat around the optimal choice) but the difference in the utility loss from inequality aversion is more important. Therefore, the choice of a type NS follower that minimizes the utility loss due to disadvantageous inequality also maximizes (10), and the follower chooses (11b). In Interval C, the reduction in material payoff if the follower uses (11b) is relatively large. As a result, minimizing the utility loss due to inequality aversion cannot compensate for the loss in the material payoff, and it is optimal for the follower to compromise between these two terms and choose according to (11c).

The best response function of a type NS follower is represented in the lower panel of Figure 1. There are a number of interesting features. First, this function is continuous but is piecewise linear instead of smooth throughout the interval $[q, \bar{q}]$. Second, the best response function of a type NS follower is upward sloping in Interval B (with a slope of 1) but is downward sloping in Interval C (and is less steep than in Interval A). Third, the best response function of type S followers is just a special case of that of type NS followers, when both parameters a and b approach 0.

IV. MAXIMUM LIKELIHOOD ESTIMATION AND TESTING

The theoretical analysis of the simplified Fehr and Schmidt (1999) model in Section III shows that the best response function of Stackelberg followers with non-standard preferences is piecewise linear, with the intervals determined endogenously by the inequality aversion parameters. On the other hand, the best response function of the Stackelberg followers with standard preferences is linear. We now provide statistical analysis of this model.

The likelihood function of a sample of Stackelberg followers' choices

We first derive the likelihood function of the followers' choices conditional on the leaders' actions in a sample of n independent experimental trials. In the statistical model, we interpret that the theoretical predictions of the simplified Fehr-Schmidt model in Section III form the systematic part of the followers' behavior, and we introduce a random error term to capture other unspecified influences on their behavior.

To simplify the notation in the following analysis, we use x and y to stand for q_L and q_F , respectively. Let x_i and y_i represent, respectively, the leader's and follower's choices for the i -th ($i = 1, \dots, n$) observation. If the follower in the i -th trial has standard preferences, then

$$y_i = q_F^S(x_i) + \varepsilon_i, \quad (14)$$

where the best response function $q_F^S(\cdot)$ is given by (9) and the random error ε_i is assumed to be independently and identically distributed according to a normal distribution $N(0, \sigma^2)$. Alternately, if the follower in the i -th trial has non-standard preferences, then

$$y_i = q_F^{NS}(x_i) + \varepsilon_i, \quad (15)$$

where the best response function $q_F^{NS}(\cdot)$ is given by (11a) to (11c), depending on the value of the leader's choice x_i .

According to the simplified Fehr-Schmidt model, the population of Stackelberg followers consists of proportion $1 - p$ of players with standard preferences, and proportion p of players with non-standard preferences. Thus, from an econometrician's perspective, the probability density of observing y_i (conditional on x_i and parameters a, b, p and σ) is given by

$$(1 - p) \times f_S(y_i | x_i; \sigma) + p \times f_{NS}(y_i | x_i; a, b, \sigma), \quad (16)$$

where $f_S(y_i | x_i; \sigma)$ is the probability density of observing y_i when the follower has standard preferences, and $f_{NS}(y_i | x_i; a, b, \sigma)$ is the probability density of observing y_i when the follower has non-standard preferences.

In the analysis below, we simplify $f_S(y_i | x_i; \sigma)$ as $f_S(y_i)$ and $f_{NS}(y_i | x_i; a, b, \sigma)$ as $f_{NS}(y_i)$. From (14), (15), (9) and (11a) to (11c), we obtain

$$f_S(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[\frac{-\left(y_i - \frac{h}{2} + \frac{x_i}{2}\right)^2}{2\sigma^2} \right], \quad (17)$$

and

$$f_{NS}(y_i) = f_A(y_i)^{1-D_B(x_i)-D_C(x_i)} \times f_B(y_i)^{D_B(x_i)} \times f_C(y_i)^{D_C(x_i)}, \quad (18)$$

where

$$f_A(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[\frac{-\left(y_i - \frac{h}{2} + \frac{x_i}{2(1-b)}\right)^2}{2\sigma^2} \right], \quad (18a)$$

$$f_B(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[\frac{-(y_i - x_i)^2}{2\sigma^2} \right], \quad (18b)$$

$$f_C(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[\frac{-\left(y_i - \frac{h}{2} + \frac{x_i}{2(1+a)}\right)^2}{2\sigma^2} \right], \quad (18c)$$

and the two indicator variables are given by

$$D_B(x_i) = \begin{cases} 1 & \text{if } \left(\frac{1-b}{3-2b}\right) h < x_i \leq \left(\frac{1+a}{3+2a}\right) h \\ 0 & \text{otherwise} \end{cases}, \quad (18d)$$

and

$$D_C(x_i) = \begin{cases} 1 & \text{if } \left(\frac{1+a}{3+2a}\right) h < x_i \\ 0 & \text{otherwise} \end{cases}. \quad (18e)$$

Therefore, the (log) likelihood function of observing a sample of n independent Stackelberg experiments is given by:

$$\begin{aligned} & \ln L(a, b, p, \sigma; (x_1, y_1), \dots, (x_n, y_n)) \\ &= \sum_{i=1}^n \ln \left\{ (1-p) f_S(y_i) + p \left[f_A(y_i)^{1-D_B(x_i)-D_C(x_i)} f_B(y_i)^{D_B(x_i)} f_C(y_i)^{D_C(x_i)} \right] \right\}. \end{aligned} \quad (19)$$

The maximum likelihood estimates of the parameters of the simplified Fehr-Schmidt model are obtained by maximizing the log likelihood function (19), and the standard errors are obtained from the Hessian matrix. Even though (19) is highly nonlinear, the estimates can be obtained by standard optimization procedures. Applying the maximum likelihood estimation method to the random-matching Stackelberg experiments of Huck *et al.* (2001) with $h = 24$,⁹ we obtain the results in Table 1.¹⁰ The maximum likelihood estimates of p , a and b are 0.388, 5.231 and 0.156, respectively.

⁹There are two caveats in applying the likelihood function (19) to the Huck *et al.* (2001) data. First, this likelihood function is derived for the material payoff functions given by (6), but the payoff structure in Huck *et al.*'s (2001) experiments has been modified slightly to assure uniqueness of the equilibrium; see p. 753 of their paper. Second, the derivation of (19) is based on n independent trials ($n = 220$ in Table 1). However, the random-matching Stackelberg experiment data set in Huck *et al.* (2001) consists of 22 pairs of players, each with 10 trials. If we had the coding of individual subjects, it might be better to assume that the same follower would use either (14) or (15) in all 10 trials. Unfortunately, such detailed information for this data set has been discarded (private communication). The best we can do in this situation is to use (19), and this can still be justified if we make the perhaps more controversial assumption that a follower may use (15) with probability p and (14) with probability $1-p$ in different trials.

¹⁰Note that restrictions (3) to (5) are not imposed in the estimation. The validity of these restrictions will be examined, as discussed in the next paragraph.

The estimated best response function of type NS followers is represented by the piecewise linear function ABDE in Figure 2, whereas the best response function of type S followers is given by the straight line ACF. Among the 108 followers whose choice variables are not on the best response function of type S followers, 92 (85.1%) are above this line and 16 (14.8%) are below it.

Before providing interpretation of the estimated coefficients, in the next sub-section we will examine restrictions (4) and (5) on the inequality aversion parameters of the simplified Fehr-Schmidt model, under the maintained assumption that some players have non-standard preferences (i.e., $0 < p < 1$). We will then test the null hypothesis of the self-interested model (i.e., $p = 0$) against the alternative hypothesis of the simplified Fehr-Schmidt model.

Testing the restrictions on the inequality aversion parameters

It can be concluded from Table 1 that the restrictions (4) and (5) suggested by Fehr and Schmidt (1999) are not violated.¹¹ We focus our analysis in this sub-section to see whether or not these restrictions can be further simplified.

To examine whether the disadvantageous inequality aversion parameter is equal to or strictly larger than the advantageous inequality aversion parameter, we use a one-sided t -test to test the null hypothesis $H_0 : a - b = 0$ against the alternative hypothesis $H_1 : a - b > 0$. We construct the usual test statistic $t = \frac{\hat{a} - \hat{b}}{se(\hat{a} - \hat{b})}$, where \hat{a} and \hat{b} are, respectively, the maximum likelihood estimate of a and b , and $se(\hat{a} - \hat{b})$ is the standard error of $\hat{a} - \hat{b}$. Under the null hypothesis $a - b = 0$, this test statistic has a standard normal distribution asymptotically. Based on the estimation results in Table 1, the value of the t -statistic is calculated to be $t = 3.84$ and thus the null

¹¹For example, to test the null hypothesis $H_0 : b = 1$ against the alternative hypothesis $H_1 : b > 1$, the t -statistic is given by $t = \frac{\hat{b} - 1}{se(\hat{b})} = -7.40$. Therefore, the null hypothesis is not rejected.

hypothesis $a - b = 0$ is rejected at, say, a 5% significance level.

To examine whether the advantageous inequality aversion parameter is equal to or strictly larger than 0, we test $H_0 : b = 0$ against $H_1 : b > 0$. The value of the t -statistic is given by $t = 1.37$. Thus, the null hypothesis that advantageous inequality aversion is unimportant ($b = 0$) cannot be rejected at a 5% significance level.

In the following analysis we further simplify the Fehr-Schmidt model by using the restriction

$$b = 0, \tag{4a}$$

instead of (4). Following the same steps as above, it can be shown that the log likelihood function corresponding to the simplified Fehr-Schmidt model with no advantageous inequality aversion is given by

$$\ln L = \sum_{i=1}^n \ln \left\{ (1-p) f_S(y_i) + p \left[f_S(y_i)^{1-D_B(x_i)-D_C(x_i)} f_B(y_i)^{D_B(x_i)} f_C(y_i)^{D_C(x_i)} \right] \right\}, \tag{20}$$

where

$$D_B(x_i) = \begin{cases} 1 & \text{if } \frac{h}{3} < x_i \leq \left(\frac{1+a}{3+2a}\right) h \\ 0 & \text{otherwise} \end{cases}, \tag{20a}$$

and $f_S(y_i)$, $f_B(y_i)$, $f_C(y_i)$ and $D_C(x_i)$ are the same as before. Note that when $b = 0$, (18a) becomes (17) and (18d) becomes (20a).

The maximum likelihood estimates of this model are given in Table 1. It is observed that the estimated values of p and a are very close to those of the previous model. The estimated best response function of type NS followers (with no advantageous inequality aversion) is represented by the piecewise linear function ACDE in Figure 2.

Testing the self-interested model against the simplified Fehr-Schmidt model

Huck *et al.* (2001) examine the conventional self-interested model by using a linear regression model of the followers' choices. In their Table 6, the null hypothesis of the self-interested model with a linear best response function is tested against an alternative model with another linear best response function (i.e., the parameters are different from those of the self-interested model). However, it is not clear what behavioral model corresponds to their alternative hypothesis.

Guided by the theoretical analysis in Section III, we believe that a better way to examine the validity of the self-interested model is to use the simplified Fehr-Schmidt model (instead of an unspecified model with a linear best response function) to nest the self-interested model.

To test the null hypothesis of the self-interested model against the alternative hypothesis of the simplified Fehr-Schmidt model, we obtain the likelihood function under the null hypothesis as

$$\ln L(\sigma; (x_1, y_1), \dots, (x_n, y_n)) = \sum_{i=1}^n \ln [f_S(y_i)], \quad (21)$$

where $f_S(y_i)$ is given by (17). The maximum likelihood estimates of this model for the Huck *et al.* (2001) data are given in Table 1.

To test the null hypothesis of the self-interested model against the simplified Fehr-Schmidt model, we construct the likelihood ratio test statistic

$$LR = 2(\ln L_u - \ln L_r), \quad (22)$$

where the unrestricted log likelihood $\ln L_u$ is given by (20) and the restricted log likelihood $\ln L_r$ is given by (21). Under the null hypothesis, the test statistic of (22) has a χ^2 -distribution of 2 degrees of freedom. Based on the estimation results in Table 1, $LR = 184.88$. Therefore, the null hypothesis is strongly rejected.¹²

¹²The same conclusion is obtained if $\ln L_u$ is calculated according to (19) when advantageous

Summary

To summarize, the self-interested model is decisively rejected against the simplified Fehr-Schmidt model, using the Huck *et al.* (2001) experimental data. Moreover, the estimates of both p and a are significantly different from zero. According to the estimation results of the simplified Fehr-Schmidt model in Table 1, there is a representative group of followers (39%) with non-standard preferences, and the estimated disadvantageous inequality aversion parameter is very high (at a value of 5.2). These conclusions are robust whether advantageous inequality aversion is included in the model or not.

Our empirical results are generally consistent with those in previous studies, but there are also new findings. First, the estimation results in Table 1 are consistent with the experimental evidence in Messick and Sentis (1985) and Loewenstein *et al.* (1989) that the subjects exhibit a significantly weaker aversion to advantageous inequality than to the aversion to disadvantageous inequality. Second, Goeree and Holt (2000) use the maximum likelihood estimation method to study inequality aversion in experimental alternating-offer bargaining games. They assume that all proposers have the same utility parameters, as do the responders. According to their findings, the disadvantageous inequality aversion estimate is 0.84, and the advantageous inequality aversion estimate is 0.66 for the proposers and 0.12 for the responders. Since they do not assume that a proportion of players have the standard self-interested preferences, the comparison with our estimation results is not direct. Nevertheless, it appears reasonable to conclude that, when compared with their results, our disadvantageous inequality aversion estimate is higher and advantageous inequality aversion estimate is lower. Overall, the empirical evidence on inequality aversion based on the Huck *et al.* (2001) data set is qualitatively similar to previous studies, but our results of

inequality aversion is included.

higher disadvantageous inequality aversion estimate and lower (and insignificant) advantageous inequality aversion estimate also reveal some quantitative disagreement with previous studies. Further research will be required to sort out the differences regarding the ranges of inequality aversion parameters found in various studies.

V. CONCLUSION

In nearly half of the random-matching Stackelberg duopoly experiments in Huck *et al.* (2001), the followers do not behave as predicted by the conventional theory. Their behavior raises the questions of why the conventional theory fails by such an extent, and necessitates a better explanation.

In the Stackelberg duopoly experiments (as well as many other game-theoretic situations), testing the predictions of conventional theory is really a matter of testing the joint hypothesis of the players' (sometimes rather complicated) strategic behavior and the standard self-interested preferences. In the current case, the structure of the game is very simple (especially for the followers), and it is reasonable to assume that the players are intelligent enough to understand the game and behave strategically. As a result, it is natural to seek explanations based on non-standard preferences.

While there are well-known models of social preferences in the literature, there are also critics of this approach who worry that altering the unobservable utility function would allow one to explain any phenomenon. In his book on behavioral economics, Camerer (2003, p. 101) defends the study of social preferences convincingly: 'The goal is *not* to explain every different finding by adjusting the utility function just so; the goal is to find parsimonious utility functions, supported by psychological intuition, that are general enough to explain many phenomena in one fell swoop, and also make new predictions.'

In this paper, we follow this idea and consider a parsimonious model of social preferences that nests the standard preferences as a special case to explain experimental

Stackelberg duopoly. Our paper makes methodological and substantive contributions. Methodologically, using the study of inequality aversion in the Stackelberg game, we illustrate that existing econometric methods can be applied to test a conventional self-interested model against a relatively new model of social preferences. We show that the best response function of the Stackelberg followers with non-standard preferences is piecewise linear, with the intervals depending on the inequality aversion parameters. Moreover, the results about the theoretical model lead naturally to a maximum likelihood framework for statistical analysis. Through careful econometric analyses, our paper also makes substantive contributions by showing that the conventional self-interested model is rejected decisively against the alternative hypothesis of the simplified Fehr-Schmidt model, and that a significant proportion (close to 40%) of the players are significantly averse to disadvantageous inequality. Given that the behavior of a high proportion of individuals in the experiments conducted by Huck *et al.* (2001) is inconsistent with the conventional self-interested model, one should not apply this model indiscriminately, but rather consider whether other factors (such as inequality aversion) may be useful in specific contexts. Furthermore, our estimates of the magnitude of inequality aversion parameters provide support for the use of a more parsimonious model with no advantageous inequality aversion in future applied work.

Finally, we should point out that, while we take the Fehr-Schmidt model seriously and apply a rigorous statistical framework to examine whether its implications are consistent with the data, inequality aversion is not necessarily the only theory that may explain the Huck *et al.* (2001) data set. In recent contributions to the literature on fairness, Charness and Haruvy (2002), Falk and Fischbacher (2006) and Falk *et al.* (2008) suggest that both the intention-oriented theory of reciprocity (such as Rabin, 1993; Dufwenberg and Kirchsteiger, 2004) and the outcome-oriented theory of inequality aversion (such as Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000)

may be relevant. In the context of experimental Stackelberg games, a possible way to discriminate between intention-oriented and outcome-oriented preferences is to consider two games: one which allows the leaders to choose (as in Huck *et al.*, 2001) while the other generates the leaders' actions randomly. It is likely that the results of this paper, such as the best response functions of different types of Stackelberg followers and the associated likelihood function, will be useful in studies further examining these types of questions. More generally, we find that there is reason to believe that systematic application of econometric techniques to the study of social preferences can deepen our understanding of human behavior and motivation.

APPENDIX

Derivation of the best response function of type NS followers.¹³ As the utility level of a type NS player depends on whether z_F is larger or smaller than z_L , we analyze these two cases separately (for each of the three intervals). Note that when either (i) $h - q_L - q_F > 0$ and $q_F < q_L$ (i.e., area ODEG in Figure 1) or (ii) $h - q_L - q_F < 0$ and $q_F > q_L$ (i.e., area FGH in Figure 1), we have $z_F < z_L$ and therefore,

$$U_F = z_F - a(z_L - z_F) = (h - q_L - q_F) [(1 + a)q_F - aq_L]. \quad (\text{A1})$$

On the other hand, when either (i) $h - q_L - q_F < 0$ and $q_F < q_L$ (i.e., area EFG in Figure 1) or (ii) $h - q_L - q_F > 0$ and $q_F > q_L$ (i.e., area Oghi in Figure 1), we have

$$U_F = z_F - b(z_F - z_L) = (h - q_L - q_F) [(1 - b)q_F + bq_L]. \quad (\text{A2})$$

Consider Interval A when $q_L \in [\underline{q}, \tilde{q}_{AB}]$. Using (A1), we show that for any q_L in Interval A, U_F is increasing in q_F for $q_F \in [\underline{q}, q_L]$. Using (A2), we show that for any q_L in Interval A, U_F is first increasing in q_F for $q_F \in \left[q_L, \frac{h}{2} - \frac{q_L}{2(1-b)} \right]$ and then decreasing

¹³In the following analysis we make use of well-known properties of quadratic functions.

in q_F for $q_F \in \left[\frac{h}{2} - \frac{q_L}{2(1-b)}, \bar{q} \right]$. Thus, the best response of a type NS follower to q_L in Interval A is given by (11a).

Consider Interval B when $q_L \in [\tilde{q}_{AB}, \tilde{q}_{BC}]$. Using (A1), we show that for any q_L in Interval B, U_F is increasing in q_F for $q_F \in [q, q_L]$. It can also be shown that for any q_L in Interval B, U_F is decreasing in q_F for $q_F \in [q_L, \bar{q}]$.¹⁴ Thus, the best response of a type NS follower to q_L in Interval B is given by (11b).

Consider Interval C when $q_L \in [\tilde{q}_{BC}, \bar{q}]$. It can be shown that for any q_L in Interval C, U_F is increasing in q_F for $q_F \in \left[q, \frac{h}{2} - \frac{q_L}{2(1+a)} \right]$, and decreasing in q_F for $q_F \in \left[\frac{h}{2} - \frac{q_L}{2(1+a)}, \bar{q} \right]$.¹⁵ Thus, the best response of a type NS follower to q_L in Interval C is given by (11c). ■

REFERENCES

- [1] Bolton, G. E. (1991). ‘A comparative model of bargaining: Theory and evidence’, *American Economic Review*, Vol. 81, pp. 1096-1136.
- [2] Bolton, G. E. and Ockenfels, A. (2000). ‘ERC: A theory of equity, reciprocity, and competition’, *American Economic Review*, Vol. 90, pp. 166-193.
- [3] Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton University Press, Princeton, NJ.

¹⁴It turns out that the followers’ best response function is the same whether $\bar{q} = \frac{d}{2}$ or $\frac{h}{2}$, but the derivation is easier to see if $\bar{q} = \frac{h}{2}$. When $\bar{q} = \frac{h}{2}$, $h - q_L - q_F$ is always non-negative and we can use (A2) to show that for any q_L in Interval B, U_F is decreasing in q_F for $q_F \in [q_L, \bar{q}]$. On the other hand, when $\bar{q} = \frac{d}{2}$, we need to use both (A1) and (A2) to obtain the same result if $a > 0.5$ but we need only (A2) to obtain this result if $a \leq 0.5$.

¹⁵As in the previous footnote, the derivation is less complicated if $\bar{q} = \frac{h}{2}$ instead of $\frac{d}{2}$. In this case, we can use (A1) to show that for any q_L in Interval C, U_F is first increasing in q_F for $q_F \in \left[q, \frac{h}{2} - \frac{q_L}{2(1+a)} \right]$ and then decreasing in q_F for $q_F \in \left[\frac{h}{2} - \frac{q_L}{2(1+a)}, q_L \right]$. Using (A2), we can show that for any q_L in Interval C, U_F is decreasing in q_F for $q_F \in [q_L, \bar{q}]$.

- [4] Charness, G. and Haruvy, E. (2002). ‘Altruism, equity, and reciprocity in a gift-exchange experiment: An encompassing approach’, *Games and Economic Behavior*, Vol. 40, pp. 203-231.
- [5] Cox, J. C., Friedman, D. and Gjerstad, S. (2007). ‘A tractable model of reciprocity and fairness’, *Games and Economic Behavior*, Vol. 59, pp. 17-45.
- [6] Daughety, A. F. (1990). ‘Beneficial concentration’, *American Economic Review*, Vol. 80, pp. 1231-1237.
- [7] Dufwenberg, M. and Kirchsteiger, G. (2004). ‘A theory of sequential reciprocity’, *Games and Economic Behavior*, Vol. 47, pp. 268-298.
- [8] Falk, A. and Fischbacher, U. (2006). ‘A theory of reciprocity’, *Games and Economic Behavior*, Vol. 54, pp. 293-315.
- [9] Falk, A., E. Fehr and Fischbacher, U. (2008). ‘Testing theories of fairness—Intentions matter’, *Games and Economic Behavior*, Vol. 62, pp. 287-303.
- [10] Fehr, E., Klein, A. and Schmidt, K. M. (2007). ‘Fairness and contract design’, *Econometrica*, Vol. 75, pp. 121-154.
- [11] Fehr, E. and Schmidt, K. M. (1999). ‘A theory of fairness, competition, and cooperation’, *Quarterly Journal of Economics*, Vol. 114, pp. 817-868.
- [12] Goeree, J. K. and Holt, C. A. (2000). ‘Asymmetric inequality aversion and noisy behavior in alternating-offer bargaining games’, *European Economic Review*, Vol. 44, pp. 1079-1089.
- [13] Huck, S., Müller, W. and Normann, H.-T. (2001). ‘Stackelberg beats Cournot—On collusion and efficiency in experimental markets’, *Economic Journal*, Vol. 111, pp. 749-765.

- [14] Loewenstein, G. F., Thompson, L. and Bazerman, M. H. (1989). 'Social utility and decision making in interpersonal contexts', *Journal of Personality and Social Psychology*, Vol. 57, pp. 426-441.
- [15] Messick, D. M. and Sentis, K. P. (1985). 'Estimating social and nonsocial utility functions from ordinal data', *European Journal of Social Psychology*, Vol. 15, pp. 389-399.
- [16] Rabin, M. (1993). 'Incorporating fairness into game theory and economics', *American Economic Review*, Vol. 83, pp. 1281-1302.

Figure 1. Best response functions of type S and type NS followers

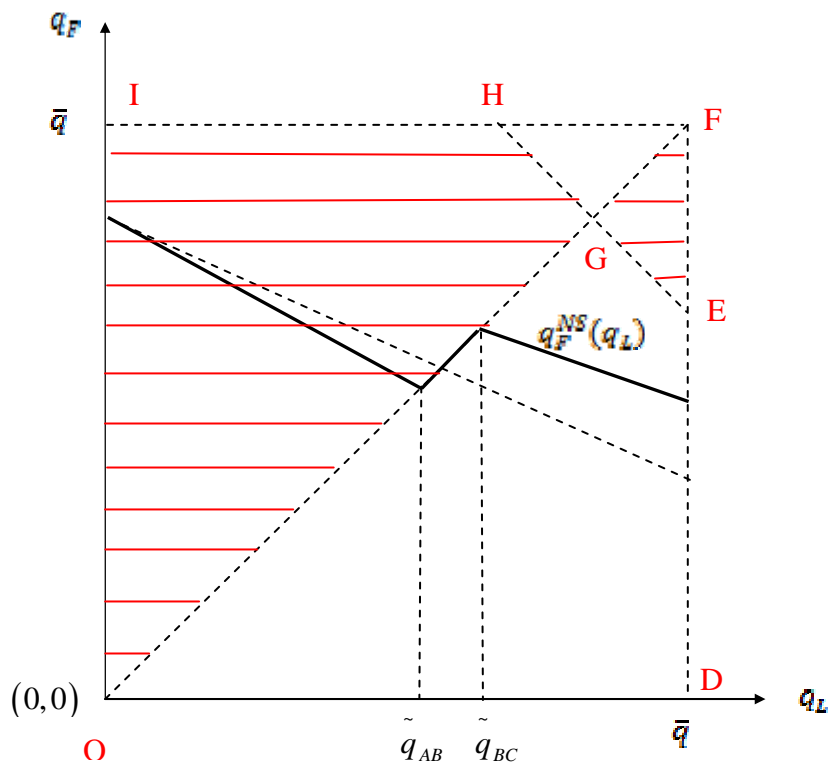
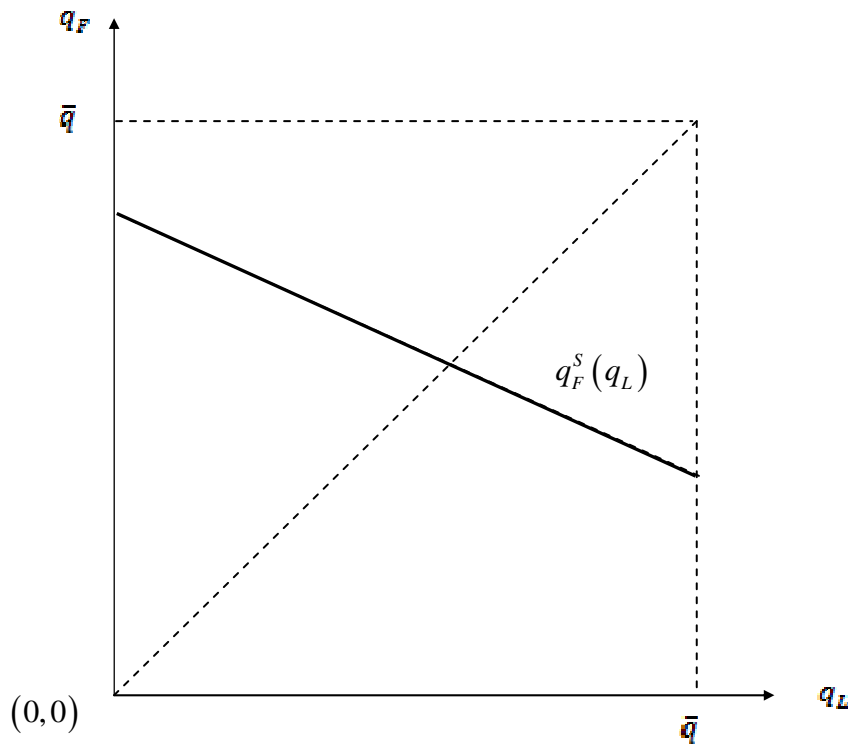


Figure 2. Estimated best response functions

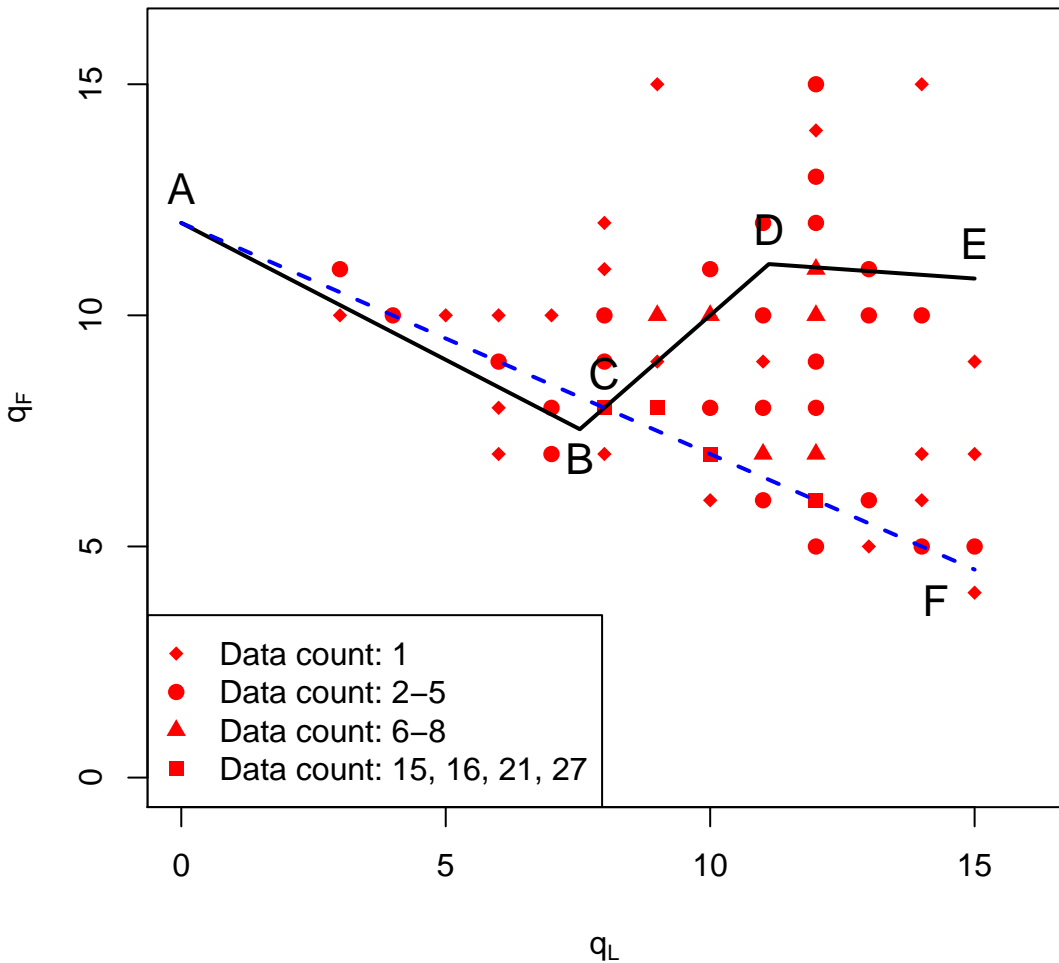


TABLE 1

Estimation results

	<i>Simplified Fehr-Schmidt model</i>	<i>Simplified Fehr- Schmidt model with no advantageous inequality aversion</i>	<i>Self-interested model</i>
p	0.388 (0.0427)	0.386 (0.0427)	-
a	5.231 (1.316)	5.231 (1.323)	-
b	0.156 (0.114)	-	-
σ	1.158 (0.0617)	1.164 (0.0619)	2.591 (0.124)
Estimate of $\text{cov}\left(\hat{a}, \hat{b}\right)$	0.000503	-	-
ln L	-428.52	-429.14	-521.58

Notes: For each of parameters p , a , b and σ , the number not in parentheses is the maximum likelihood estimate and the number in parentheses is the standard error. Note that the sample size (n) is 220 and $h = 24$ in Huck *et al.* (2001).