Patent Race Experiment Revisited: Are Differences Between Experiments Just Random Variation?

Kähkönen, Anssi

ISBN 952-458-685-1
ISSN 1458-686X
no 25

Patent Race Experiment Revisited: Are Differences Between Experiments Just Random Variation?

Anssi Kähkönen*

May 9, 2005

Abstract

This paper reports the results of a replication experiment of a patent race experiment by Zizzo (2002, *International Journal of Industrial Organization*. 20, 877–902). First, the results of the replication experiment are presented and compared to the results of the original one. Secondly, the appropriate econometric models and a bootstrap validation method are employed. Third, data sets of both experiments, the replication and the original, are pooled and analyzed jointly. In addition, a so-called mixed model is introduced to control for different type of heterogeneity observed in experiments. Basically, the replication is in many respects successful, and the results are generally in line with the original study. However, there are differences which cannot be entirely considered as random variation between samples. In contrast to the earlier study linear random effects model is inappropriate with replication data and pooled data. Also the replication experiment provides only limited support for the theoretical predictions.

Keywords: bootstrap method, panel models, patent races, replication
JEL Classification: C15, C33, C88, C91, O31

1 Introduction

The importance of replication studies is noticed within the field of experimental economics. The scientific progress calls for replications and meta-analysis studies since they can establish empirical facts on which the new ideas in single studies are founded.[1] The reporting and documentation of experiments in experimental economics leans greatly on the guidelines provided Palfrey and Porter (1991) and others. However, the information of conducted experiments is often insufficient to replicate earlier experiment(s),

*University of Joensuu, Department of Business and Economics, P.O. Box 111, FI - 80101 Joensuu, Finland. Email: *anssi.kahkonen@joensuu.fi*.

especially if the experimental design is not well-established.[2] In practice, the replication is typically done on basis of a published article (or a working paper) from which some of the documented instructions, experimental design, and results can be found. The raw data, estimation procedures and software are seldom directly available with the published work, but normally they can be requested from (or are made available by) the author.[3]

A low incentive to replicate, due to editorial policies, is another concern.[4] Frequently, so-called *conceptual replications*, studies that are not identified primarily as replications, are conducted to 'search for the missing control group' by testing hypotheses that are not yet checked. For example new elements are introduced to the design before the initial ones are soundly studied. This problem can be partly addressed with diagnostic reasoning and meta-analysis.[5] Briefly the idea in diagnostic reasoning is to systemically search for a 'faulty' component in the experimental design, suggested by negative/disconfirming evidence, and to replace it with a new one, and proceeding recursively. However, a proper realization and documentation of replications are needed to observe how results change with condition in the first place, before one can proceed and include a new component (treatment etc.) into the experimental design.

Hunter (2001), Lindsay and Ehrenberg (1993) and Posavac (2002) have categorized replications into two main categories: *the statistical [or exact] replication* and *the scientific [or close] replication*. The replication reported in this paper can be categorized as the latter one [as replications in experimental economics usually can]. Generally, the replication needs to fulfill following conditions: the same independent and dependent variables are measured, essentially the same procedures are used, and the sample is from a population that is equivalent in terms of the study objective and outcome. Notice that in case of statistical replication the word 'same' actually means *identical*, while in scientific replication it is interpreted as *equivalent*.[6] Differences between the replication and the original study should only be due to random variation, i.e. sampling variability. However, in addition to sample variability there can be errors from other sources; such as experimental design, model specification or estimation methodology. In addition, one can also search for 'the edges of the robustness' by conducting a critical *differentiated replication*.[7]

In the original study, Zizzo (2002) experimentally tested a dynamic multi-stage patent race model of Harris and Vickers [HV] (1987, Section 4). In this close replication the same experimental design, procedures and analysis methods are used to review the HV predictions. Budd et al. (1993) have reconsidered the framework with asymmetric setting in the context of stochastic evolution of a dynamic duopoly. They specified four different effects from which two, *the joint-profit effect* and *the endpoint effect*, are at work in the context of patent race theory. The former is determined by profit/payoff incentives and the latter by cost behavior. Recently Hörner

(2004) has introduced results of a more general dynamic framework identifying Markov perfect equilibria and two distinct situations or 'triggers' where a firm exerts high efforts. First, securing the position when lead is sufficiently large and second, defending (or regaining) the leadership when the lead (or the lag) is very small.

In other experiments closely related to the patent race theory, Sbriglia and Hey (1994) reported results of a series of experiments concerning R&D competition, where the experimental design was based on a search model for innovations. They investigated investment strategies, selection of search procedures, and the use of information. In their design, the effects of experience and information to the pace of innovative activity were the prime objectives. Sbriglia and Hey argued that the theory does not provide a satisfactory framework to study real-world patent races.

The current replication is mostly successful and the results are generally in line with the original study. However, the bootstrap results indicate that there are differences which cannot entirely be interpreted as random variation between the samples. Only limited support for HV predictions was found as in the original study. In the replication data there exists some symptomatic evidence, albeit not statistically significant, supporting the prediction "...the leader always makes greater efforts than the follower.". The prediction "...the race quickly approaches "monopoly" as the disparity between player's positions widens." was not supported in the replication. With the pooled data different type of heterogeneity of games and subjects was observed between experiments: unobserved effects were fixed game-specific rather than random player-specific according to specification tests. Therefore, a so-called mixed model was introduced to control for observed heterogeneity in both experiments.

The remainder of this paper is organized as follows. Section 2 describes the experimental design, procedures, and characteristics of subjects. In Section 3, the results of the replication experiment are presented and compared to the original study. The bootstrap validation method is used for comparison of parameter estimates. In addition, the estimation results of mixed model with pooled data are presented. The last section concludes with a discussion of the implications of the replication study.

2 Experimental design, subjects and software

The replication experiment was conducted in March 2004 at the University of Joensuu consisting of eight sessions. In seven sessions there were four subjects in each, and one session was conducted with eight subjects. In total 36 subjects played 598 rounds of dyads. The sessions lasted between 45 and 60 minutes.

A recruiting method was an automated web form. A slight majority of randomly chosen participants were female (59.03%) and undergraduate students (78.60%). One of the participants had a doctoral and one a licentiate degree. The age distribution of the participants was quite similar (min 20, max 51, mean 25.07, S.D. 5.30) between the studies, and the median, 23 years, was the same. One third of the participants (32.78%) had economics or business backgrounds and the second third (33.78%) had humanities background, and the remaining participants consisted of several of disciplines like social sciences and forestry. Some of the participants (21.24%) had prior experience on experiments.

The experimental design followed very closely the original experiment[8] with the following practical exceptions. First, the instructions were translated into Finnish. Secondly, the currency was euro instead of sterling implying different exchange rate for laboratory points; one point was worth 1.2 cents instead of 1 UK penny. Thirdly, the instructions consisted of a table summarizing the experimental design (see Appendix A.1.). The total winnings were 354.32 euros (without taxes and side-costs). The average of the final winnings was 9.84 euros (S.D. 6.17, min 2.40, max 17.30) including a 2.40 euros participation fee. As a summary, the experimental design and the monetary incentives were practically equal between experiments.

The experiment was run in a computer classroom using novel Java-based software.[9] It is worth mentioning that the implementation of software was distinct to the original experiment and used a different random number generator.[10] A short description of software and a screen shot during the race can be found from Appendix A.2. The instructions were provided on both paper and computer screens. In addition to the original experiment, an extra questionnaire of 10 multiple-choice question was done before the experiment to measure (and to control) each participant's initial level of understanding after reading the instructions. [11] The second automated questionnaire was similar as in the original experiment and consisted of four multiple-choice questions.

3 Results

3.1 General results

The data sets of the replication and the original experiments have similar patterns. However, in descriptive analysis there are some differences concerning merely the minimum and the maximum values of variables. Durations of dyads between experiments are similar, in the replication the mean was 16.6 rounds with standard deviation of 1.79 (min 13, max 19), while in the original study they were 16.1 with 2.42 (min 12, max 19). Portions of tied rounds were similar between studies, 17.06% compared to 18.90%. Investments were slightly, but not statistically, larger in the replication experiment; the median was 5, the mean 4.51 and standard deviation 1.61, compared to 4.5, 4.06 and 2.15 in the original study.

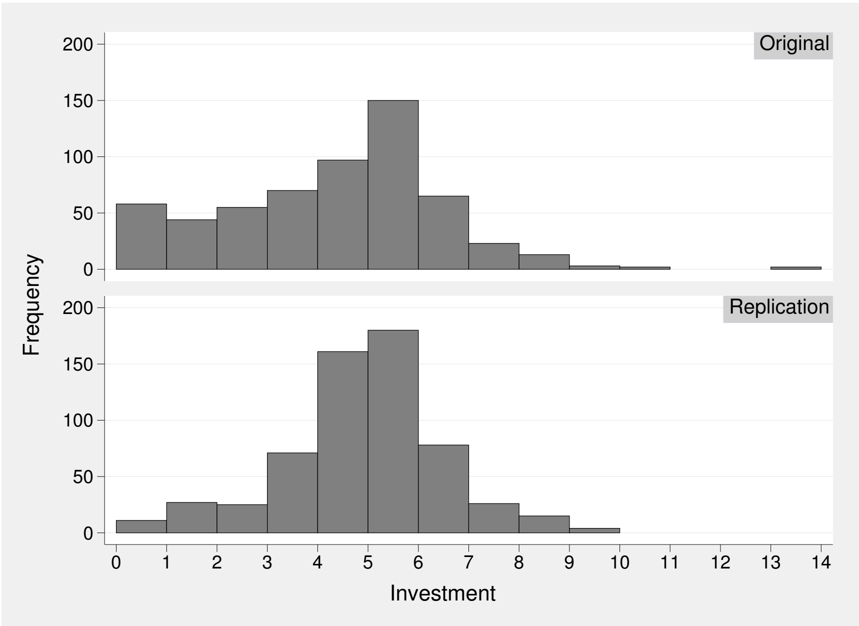

Figure 1: Distributions of investments in experiments.

The distributions of investments between experiments were dissimilar. Figure 1 illustrates that the distribution is more concentrated in the replication experiment. In the original experiment, the investments have a lower mean due to larger amount of low investments (below one) and the standard deviation is larger due to two quite large investments (13 and 13.5).

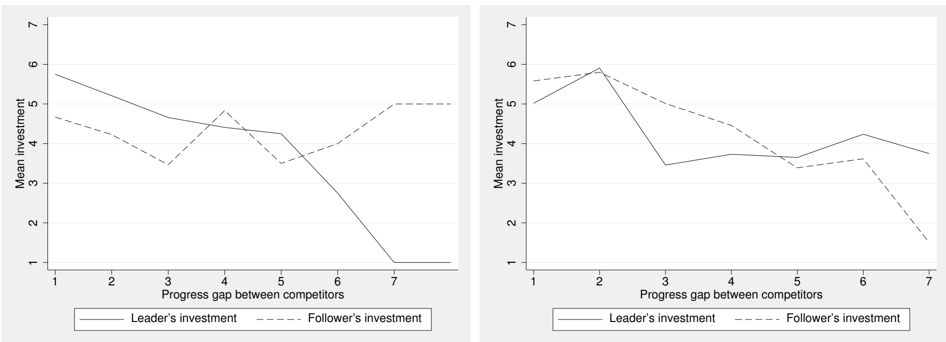

Figure 2: Reproduced Figure 4, replication (left) and original (right).

In the replication data there was a positive correlation between investment and round (Spearman's $\rho = 0.131$, $p = 0.001$) as compared with the original study. Subjects with different backgrounds did not invest differently than others in the replication experiment. There was a positive correlation between investments in the practice stage and those in the real stage ($\rho = 0.128$, $p = 0.002$). Similar to the original study there was a positive correlation between progress and investment ($\rho = 0.131$, $p = 0.001$) as well. The average binding index discussed extensively by Zizzo (2002) is not considered in this paper.

If we compare the means and the standard deviations of investments of the leaders, when there is still two progress steps to go, 4.592 and 1.908 to the original experiment 4.477 and 2.776 we notice that the means are not equal. It is the same with the followers, 4.184 and 1.993 compared to 4.734 and 2.524. However, the means do not differ significantly according to standard t-test. Even with a limited sample with score less than 100 points (leaders 4.44 and followers 4.53) leaders appear to invest less in the original study which is clearly not the case in the replication experiment (leaders 5.33 and followers 4.61). This is shown in Figure 2, corresponding to Figure 4 in Zizzo (2002, p.887).

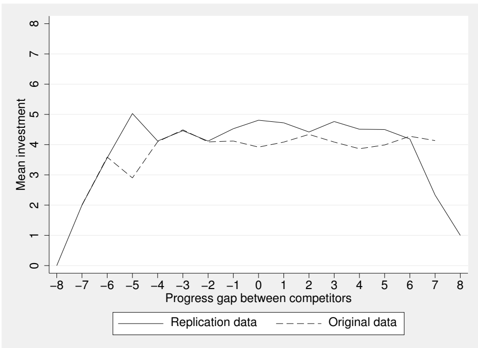

Figure 3: Reproduced Figure 5 in Zizzo (2002) with both data sets.

Investments seem to differ between experiments relative to progress gap (the gap between competitors in progress steps). With the gap of three or more the race should be characterized as 'a virtual monopoly', however, it is not. In fact, the followers invest relatively more than the leaders. Above values base on derived variable, ProgressSelfAnte[12], which describes the progress made at the beginning of each round rather than in the end of the round characterizing theory better than just a plain Progress does when there is two or less progress steps to go. In contrast to the original study, the relationship between investment and size of gap is stronger for leaders ($\rho = -0.341$, $p = 0.006$) than for followers ($\rho = -0.279$, $p = 0.025$).

In Figure 3, the reproduced Figure 5 in Zizzo (2002, p. 888), illustrates the differences between the studies in investments in relation to progress gap with the full sample. From Figure 3 we can see that even with full sample there is no monopoly by the leader, that is, the HV prediction does not get support. Instead it seems that the shape of the investments in Figure 3 exhibits two cases suggested by Hörner (2004), even though the model and the framework are more complex.

3.2 Estimation results of replication

The analysis justifications and the basis for choosing variables are from the original study.[13] The first set of variables (Progress $\times$ tie, Positive gap and Negative gap (and squared terms), and Leader) are chosen to represent the HV predictions. The second set of variables (Progress, Small score, Expany, Practice stage variables, etc.) are mostly chosen to represent the experimental design and to control certain factors. The last set of variables consists of demographical control variables (Age, Sex, and Educational background).

The HV predicts that both the Leader and the Interaction term (Progress $\times$ tie) to have positive coefficients. Only in one model the interaction term is positive and significant. It is also predicted that the correlation between gap and investment should be more negative for the followers than for the leaders. This is not the case with the full sample (correlations are -0.138 and -0.113).

The results of Models 1 to 5 on investments with the replication data are presented in Table 1, corresponding to the results in Zizzo (2002). The same random effects specification is used. In Models 1 to 4 the marginal effect of the progress of tied competitors on investment is studied through Progress variable. For Model 5 this indirect effect is removed and a 'pure' Progress $\times$ tie effect is examined. As a summary, for all five models, we may conclude that the results are generally similar to the original study. Most of the independent variables are significant and the coefficients have same signs, except for Sex, which implies that in the replication males invested significantly less than females. The HV prediction that the overall impact of negative gaps on investment should be negative and more negative than the impact of positive gaps was not supported. The estimation results raised doubts about the (multi)collinearity problem not only between Leader and Positive gap, $\rho = 0.9575$ with $p = 0.0000$, but also between Positive gap and Negative gap ($\rho = -0.6388$, $p = 0.0000$). Also educational background variables seem to be (multi)collinear. Thus some variables should be dropped from the 'candidate model'.

Specification testing implies that the random effects specification used in the original study is inappropriate for the replication data. Breusch and Pagan LM test (H_0: $Var(u_i) = 0$) for Model 1 is insignificant ($\chi^2(1) = 0.85$, $Prob > \chi^2 = 0.357$) and Hausman specification test is significant ($\chi^2(8) = 43.65(39.41)$, $Prob > \chi^2 = 0.000(0.000)$).[14]

Estimation results of Model 1 with the replication data using the fixed effects specification shows that there is a strong negative correlation (-0.378)

Variable	Coefficient	S.E.	p-value	95% Confidence Interval	
Progress	0.098	0.033	0.003**	0.033	0.163
Positive gap	-0.230	0.0672	0.001**	-0.362	-0.098
Negative gap	-0.170	0.055	0.002**	-0.278	-0.063
Small score	-1.228	0.279	0.000***	-1.777	-0.680
Round17	1.060	0.377	0.005**	0.319	1.801
Round18	1.418	0.493	0.004**	0.450	2.387
Round19	2.087	0.670	0.002**	0.771	3.404
Constant	4.494	0.145	0.000***	4.209	4.779
R^2 within	0.094				
R^2 between	0.002				
R^2 overall	0.058				

Significance levels: ***=0.001, **=0.01 and *=0.05

Table 1: Estimation results of game-specific fixed effect model.

between the explanatory variables $\mathbf{x}_{it}$ and the fixed individual specific effects u_i making the random effects estimator inconsistent with the replication data. The F-test (H_0: all $u_i = 0$) is significant ($F(35, 554) = 1.50$, $Prob > F = 0.036$), evidently justifying the use of the fixed effect specification. The results in Table 3 indicate that there exists (fixed) heterogeneity between games.

3.3 Bootstrap results

The results of an ideal experimental design should maintain validity (both internal and external) irrespectively of subjects' characteristics and other factors if we are able to control them properly. Basically, the uncertainty, biases and errors in experimental data-analysis and modelling stem from four different sources: (i) experimental error due to design (some factors are not under control), (ii) misspecified model (omitted variable problem etc.), (iii) estimation uncertainty, and (iv) sampling variability.

An ideal unbiased experiment should minimize (or even eliminate) the experimental uncertainty and error. If we are not able to find the 'true' model that describes the design and controls for all relevant game- and subject-specific effects, the regression model results are biased due misspecification (or the omitted variable problem). Despite the fact that models in econometrics are always not-deterministic, i.e. incomplete in some sense, in the experimental setting we should be able to reduce the severity of biases prominently (or in an ideal case we should be able to make the model almost deterministic). Then the two remaining sources of model uncertainty, i.e. estimation uncertainty and sampling variability, can both be analyzed and illustrated empirically.

Generally, the methods deriving parameter estimates, like GLS in this case, involve errors due the sampling variability and non-ideal data. The bootstrap method, introduced by Efron (1979) is a re-sampling validation method whereby information in the data is 'recycled' for estimating the distributions of sample statistics, estimates in this case. It is based on an idea that the sample we have is a good representation of the underlying population. By comparing the estimates computed from different sub-samples, one can analyze how representative the estimates are. In this paper only the results of Model 4 are considered. We observe that the point and interval estimates do differ (see Table 2), as they do with all the five models. However, the result might be an outcome of sampling variability. If the bootstrap confidence intervals for parameter estimates are not overlapping with estimates from the initial and the replication data, then we are forced to argue that the results differ partly due to the modeling uncertainty or the omitted variable problem, and then we need to reconsider and re-specify the model to control the unobserved effects appropriately. An alternative explanation is that there is something wrong with the experimental design and we need to re-design the experiment.

Since parameter estimates are not directly comparable between the studies, due to distinct model specifications, we address the matter by using the improved bootstrap method BCa[15], which produces automatically confidence intervals that are second-order correct, to check how the nominal confidence intervals of parameter estimates behave. The bias-corrected and accelerated 95% confidence intervals (BCaCI), nominal estimates (Observed), bias values (Bias) and nominal 95% confidence intervals (CI) with 10000 replications are reported in Table 2. Figure 6 in Appendix B illustrates the differences in distributions of parameter estimates between experiments.

The results in Table 2 and Figure 6 in Appendix B indicate that point estimates are generally qualitative similar between studies. Most independent variables are significant and coefficients have same signs. However, the bootstrap estimates differ remarkably from the nominal ones and across the two experiments. There is hardly any overlapping for 95% nominal and BCa confidence intervals. Variable Small score, the proxy for binding budget constraint, seems to have almost equal distribution between experiments. To summarize, experiments seem to have statistically different properties suggesting alternative specification.

3.4 Results of mixed model with pooled data

The pooled data set (in total 1180 observations) consist of original and replication data sets. By pooling data sets we get a continuum of experiments and accumulate experimental data in sense of meta-analysis. With larger data set more efficient parameter estimates can be produced due to larger degrees of freedom.

In the pooled data set the correlation between investment and Sex, and some other demographic variables is insignificant, but between Age and Economics it is not, $\rho = -0.109$ with $p = 0.000$ and $\rho = 0.074$ with $p = 0.011$, respectively. These imply that in general older subjects tend to invest less and economic students tend to invest more than others.[16]

	Model 1			Model 2			Model 3			Model 4			Model 5		
Explanatory variables	Coef.	S.E.	p-value	Coef.	S.E.	p-value	Coef.	S.E.	p-value	Coef.	S.E.	p-value	Coef.	S.E.	p-value
Progress	0.101	0.032	0.002**	0.123	0.031	0.000***	0.118	0.031	0.000***	0.121	0.031	0.000***			
Progress×tie	0.054	0.056	0.337	0.006	0.044	0.884	0.020	0.039	0.621	0.040	0.039	0.301	0.089	0.037	0.016*
Positive gap	0.140	0.251	0.578	-0.215	0.069	0.002**	-0.189	0.059	0.001**	-0.157	0.058	0.007**	-0.020	0.046	0.660
Negative gap	0.093	0.189	0.624	-0.078	0.142	0.581									
Positive square gap	-0.055	0.037	0.132												
Negative square gap	-0.033	0.030	0.274	-0.010	0.025	0.677	-0.023	0.009	0.013*						
Leader	-0.043	0.393	0.914												
Small score	-0.647	0.266	0.015*	-0.691	0.266	0.009**	-0.663	0.264	0.012*	-0.730	0.262	0.005**	-0.298	0.242	0.218
Sex	-0.379	0.152	0.013*	-0.317	0.159	0.046*	-0.321	0.154	0.037*	-0.360	0.150	0.017*	-0.382	0.147	0.009**
Age	0.000	0.015	0.975	-0.006	0.015	0.676									
Economics	-0.060	0.250	0.810	0.096	0.227	0.673	0.072	0.213	0.734	0.144	0.208	0.487	0.140	0.203	0.491
Humanities	-0.072	0.209	0.729	-0.060	0.207	0.772	-0.067	0.200	0.737	0.023	0.193	0.905	0.028	0.188	0.881
Sciences	-0.294	0.284	0.300	-0.195	0.283	0.489	-0.199	0.272	0.464	-0.133	0.266	0.618	-0.094	0.260	0.717
Maths	-0.458	0.482	0.342												
Expany	0.207	0.238	0.385												
Average investments in practice stage	0.296	0.199	0.137	0.345	0.173	0.046*	0.333	0.167	0.046*	0.352	0.164	0.032*	0.258	0.158	0.103
Average investment of competitor in practice stage	-0.253	0.168	0.132												
N rounds in practice stage	0.063	0.045	0.156												
NRoundSS	0.048	0.062	0.440												
NRoundSSLoss	-0.037	0.054	0.497												
Constant	3.077	1.605	0.055	3.040	0.853	0.000***	2.900	0.795	0.000***	2.630	0.774	0.001**	3.399	0.732	0.000***
R^2 within	0.080			0.073			0.071			0.052			0.016		
R^2 between	0.276			0.155			0.151			0.163			0.235		
R^2 overall	0.085			0.067			0.066			0.058			0.035		

Significance levels: ***=0.001, **=0.01 and *=0.05. Progress, number of progress steps made by the subject; Progress × tie, progress if subject is tied, 0 otherwise; Positive gap and Negative gap, absolute value of gap between competitors, when gap is positive or negative, 0 otherwise and square terms, squares of gaps; Leader, 1 for the leaders, 0 otherwise; Small score, 1 if the subject has less than 100 points, 0 otherwise; Sex, 1 for the males, 0 for the females; Age, age in years; Economics (similarly to other educational variables), 1 for subjects who have economics as a main subject, 0 otherwise; Expany, 1 if experimental experience, 0 otherwise; Average investment (of competitor) in the practice stage, as is; N rounds in practice stage, number of rounds in the practice stage; NRoundSS, number of rounds in the practice stage with less than 100 points; NRoundSSLoss, NRoundSS if the subject lost the practice stage, 0 otherwise.

Table 2: Estimation results of Models 1 to 5.

	Original						Replication					
Variable	Observed	Bias	95% CI		95% BCaCI		Observed	Bias	95% CI		95% BCaCI	
Progress	0.3162818	0.0029493	0.2451505	0.387413	0.2342681	0.4003379	0.1208984	0.0116258	0.0600656	0.1817313	0.0480201	0.1686339
Progress×tie	-0.0888076	-0.001819	-0.1784379	0.0008227	-0.1817214	0.0023822	0.0401079	-0.0039486	-0.0359207	0.1161364	-0.0245084	0.1133798
Positive gap	-0.5109711	-0.0125759	-0.6645611	-0.3573811	-0.6534731	-0.3534342	-0.1568643	-0.0176044	-0.2705205	-0.0432081	-0.2650695	-0.0110134
Small score	-0.8990316	-0.0374788	-1.685978	-0.1120857	-1.85419	0.1207609	-0.7297286	-0.2006314	-1.243229	-0.2162282	-1.295163	0.1503341
Sex	0.7398794	0.0054904	0.1442949	1.335464	0.4256359	1.034925	-0.3601263	-0.0089804	-0.6548603	-0.0653923	-0.6214984	-0.0807289
Economics	1.565637	0.0084692	0.1972273	2.934047	0.9951704	2.132346	0.1441832	-0.00584	-0.2626333	0.5509998	-0.2189333	0.5286103
Humanities	1.592879	0.0093188	0.1615401	3.024219	0.9994608	2.219554	0.0229472	-0.0023882	-0.3549313	0.4008256	-0.3250537	0.3880569
Sciences	1.481409	0.011849	0.2025656	2.760252	0.9577339	2.006663	-0.1328947	-0.0301109	-0.6546705	0.3888811	-0.5545493	0.3440611
Average investments in practice stage	0.7873969	0.0029188	0.5255157	1.049278	0.6619599	0.904358	0.3518434	0.0337272	0.0309183	0.6727685	0.0193935	0.5999469
Constant	-1.760801	-0.0232384	-3.54999	0.0283873	-2.505275	-1.00775	2.630355	-0.1599612	1.113531	4.147179	1.490744	4.165274

Table 3: Nominal and bootstrap parameter estimates and confidence intervals for Model 4.

For Model 1 with pooled data set both Breusch and Pagan LM test ($\chi^2(1) = 54.93$, $Prob > \chi^2 = 0.000$) and Hausman test ($\chi^2(8) = 33.59$, $Prob > \chi^2 = 0.029$) are significant. With fixed effect specification there exists some negative correlation (-0.177) between explanatory variables and game-specific effects. The null (all $u_i = 0$) is rejected by F-test ($F(35, 1124) = 4.36, Prob > F = 0.000$) justifying the use of specification.

A candidate model, a so-called mixed effects model, which consist a mixture of fixed effects (FE) and random effects (RE) is introduced.[17] The model includes a dummy variable Replication, indicating observations from the replication data set, which implies that investments were significantly larger in replication than in the original experiment. Also the round-dummies for the last rounds (between 16 and 19), which represent endpoint effects discussed in Budd et al. (1993), are included. The mixed model can be written in the following form:

$$y_{ijt} = \alpha + \mu_i + \nu_j + \mathbf{x}'_{ijt}\boldsymbol{\beta} + \varepsilon_{ijt}, \tag{3.1}$$

where i denotes individual, j denotes game and t denotes period. Exogenous variables are denoted by $\mathbf{x}_{ijt}$, $\boldsymbol{\beta}$ is matrix of coefficients, α is a general constant, μ_i is unobservable subject-specific effect treated as random (normally distributed, i.e. $\mu_i \sim N(0, \sigma^2_\mu)$, ν_j is unobservable game-specific effect which is treated as fixed and ε_{ijt} denotes remainder disturbance.

Variable	Coefficient	S.E.	p-value	95% Confidence Interval	
Progress	0.187	0.027	0.000***	0.134	0.239
Positive gap	-0.274	0.048	0.000***	-0.369	-0.180
Negative gap	-0.087	0.038	0.023*	-0.163	-0.012
Small score	-1.145	0.239	0.000***	-1.614	-0.676
Average investment in practice stage	0.430	0.082	0.000***	0.270	0.590
Replication	1.395	0.215	0.000***	0.975	1.816
Gamed20	2.189	0.360	0.000***	1.483	2.895
Gamed21	1.612	0.353	0.000***	0.920	2.304
Gamed22	1.506	0.407	0.000***	0.708	2.303
Gamed23	0.901	0.357	0.012*	0.202	1.600
Gamed24	2.308	0.377	0.000***	1.569	3.046
Gamed27	0.979	0.376	0.009**	0.242	1.715
Gamed28	1.598	0.362	0.000***	0.888	2.307
Gamed29	2.276	0.400	0.000***	1.492	3.060
Gamed31	1.189	0.363	0.001**	0.479	1.900
Gamed32	1.431	0.372	0.000***	0.703	2.160
Gamed33	1.682	0.368	0.000***	0.960	2.404
Gamed34	1.031	0.368	0.005**	0.309	1.753
Gamed36	2.599	0.415	0.000***	1.787	3.412
Round16	-0.503	0.266	0.058	-1.024	0.017
Round17	0.968	0.307	0.002**	0.367	1.569
Round18	1.056	0.375	0.005**	0.322	1.791
Round19	0.895	0.544	0.100	-0.171	1.962
Constant	0.703	0.307	0.022*	0.102	1.304
R^2 within	0.115				
R^2 between	0.741				
R^2 overall	0.248				

Significance levels: ***=0.001, **=0.01 and *=0.05

Table 4: Estimation results of mixed model (GLS regression).

Mixed model is very intuitive for pooled data since in previous experiment subject-specific random effects were significant, but on the other hand in the replication experiment game-specific fixed effects were significant. Model makes it possible to control both statistical different properties and heterogeneity of experiments at the same time. Estimation results are presented in Table 4. As we can see R^2 between is large due to both controlled fixed game-specific effects and random variation between subjects. Also the overall coefficient of determination is larger than in previous models giving us some prediction power.

The heterogeneity between both games and subjects is controlled by using the mixed model. Also other estimation methods, i.e. ML and adaptive (also ordinary) quadrature for multilevel generalized linear mixed models (`gllamm` in Stata), were used to evaluate the efficiency and consistency of mixed effects model.[18] Loosely speaking, the game-specific fixed effects absorb both observed and unobserved subject-specific fixed effects which cannot be controlled by using plain RE- or FE-specifications. The subject-specific random effects of the mixed model are illustrated below. There exists an endpoint effect according to round dummies (Round16-Round19). Investments tend to increase in the end of the races dropping again in the very last round.

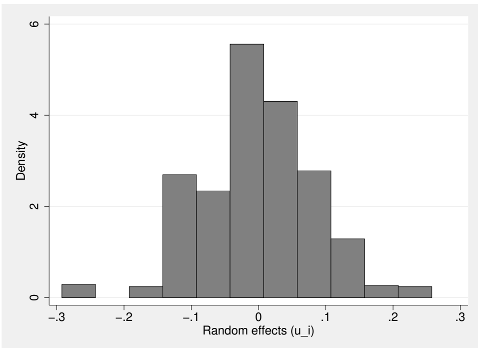

Figure 4: Distribution of random effects in mixed model.

4 Conclusion

To conclude, the replication was mostly successful and the results of the replication experiment are generally in line with the original one. However, the data sets of the experiments have statistically different properties according to the specification tests indicating that the unobserved effects are fixed game-specific rather than random subject-specific in replication data. Also the bootstrap results suggest that there are differences which cannot entirely be interpreted as random variation between samples. From the demographic variables only Age and Economics were significant in both experiments implying that the subject-specific variables are not universal and need to be controlled.

To summarize, the replication provides also limited support for HV predictions as the original experiment. However, there is suggestive evidence that the replication experiment is more favorable to the HV prediction than the original experiment in sense that the leader always makes greater effort than the followers. In contrast, the follower invested significantly more when the gap was five and less when the gap was two. Efforts were not significantly larger in a neck-to-neck situation.

Clearly we cannot control all the factors in the current experimental design even with appropriate econometric models. Thus more research is needed and more experiments are needed to be conducted. In the words of Spanos (1999) "...model specification and the design of the experiment are the two sides of the same coin. The design purports to isolate the causal relationship between inputs and outputs and what is beyond the control of the experimenter should be non-systematic; often a white noise error.". If this is not feasible, then there is a chance that the experimental design or estimation methods (or both) have ignored certain factors, which cannot be assumed away. Next logical step is to redesign the experiment in order to achieve control over the experimental setting, or at least neutralize these systematic unobserved effects. However, in experimental economics it is rarely possible to achieve total control over the experiment and identify factors and effects. Thus, we should consider appropriate methods and replications jointly as a controlling device alongside diagnostic reasoning and meta-analysis.

In the current multivariate analysis an important determinant, dynamics (i.e. investment lags), have been omitted and left uncontrolled. Preliminary results of VAR(2) models and impulse response analysis in Kähkönen (2004) suggest that subjects' decisions are characterized by actions two periods backwards giving us some insights to dynamic behavioral aspect. This observation supports the results of Hörner (2004), and points out the importance of 'decision history' in the current experimental setting, where only previous round's values were displayed to the subjects which can partly explain the low order AR process. Therefore, more advanced methods (for example panel VAR models) are needed to model dynamics and to take the underlying heterogeneity into account. Next possibly step to ease analysis could be idea of abstracting other human player out of design, and 'making uncertainty more certain' to the subjects which would certainly be a stress test for the theory. Future studies should also consider closely related issues like the persistence of monopolies, asymmetric information, different market structures or sequential innovations.

Appendix A. Additional instructions and software

A.1. Additional instructions

The experimental design and instructions followed very closely to the original experiment. In addition information in Table 5 was included in instructions (translated naturally in Finnish) to summarize the experimental design to the subjects. Other instructions which are available from the author are shortened (three pages) translated version of Zizzo's (2002) instructions.

Concept	Player X	Player Y
Investment	x	y
Cost of investment	$c_x = x^2$	$c_y = y^2$
Winning probability	$p_x = x/(x+y)$	$p_y = y/(x+y)$
"Winning rule"	X wins if $random(0,1) < p_x$	Y wins if $random(0,1) > p_x (= 1 - p_y)$

Table 5: A summary of key concepts of design provided to subjects.

A.2. A brief description of software

The experimental software was programmed by using Java programming language in co-operation with the Department of Computer Science at the University of Joensuu. The software runs over the Internet (or LAN) in browser with a Java plug-in (JRE version: 1.4.1 or higher). See figure below for a screen shot during the race. For further information contact the author.

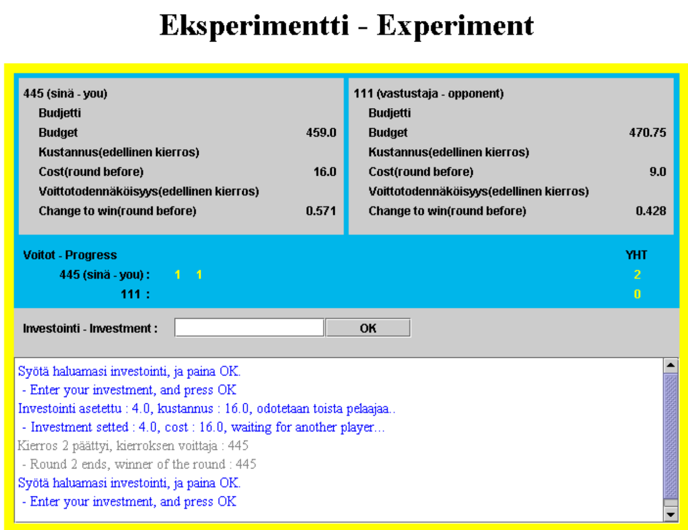

Figure 5: Screen shot of subject's view during the race.

Appendix B. Distributions of bootstrap estimates of Model 4

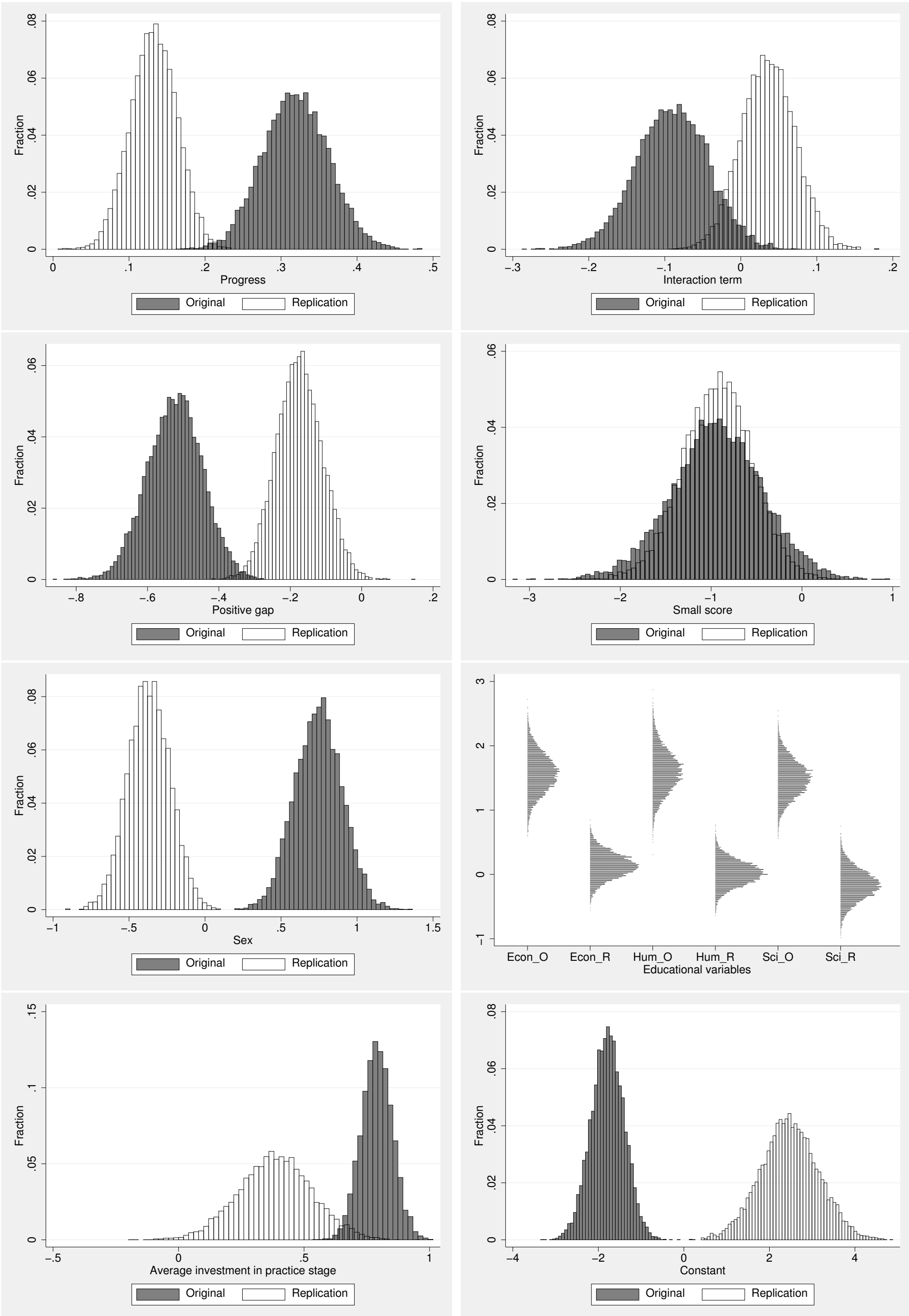

Figure 6: Distributions of bootstrap estimates with 10000 replications.

Acknowledgements

I am indebted to Daniel Zizzo for specific and constructive comments, and for providing the raw data, procedures and detailed information of the original experiment. I thank Steffen Andersen, Mikael Linden, and Timo Tammi for comments and valuable suggestions, and Kari Surakka for Java programming. I am also grateful to the participants of ESA 2004 International Meeting and EAEPE 2003 conference for helpful comments. Errors are entirely mine. The financial support of the Foundation for Economic Education and the Finnish Cultural Foundation (Northern Karelian Foundation) are gratefully acknowledged.

Notes

[1]See for example Hunter (2001).

[2]This fault is also pointed out by Zelmer (2003, p.307).

[3]The data, estimation procedures and other information about the replication experiment can be found from author's home page, URL: *http://cc.joensuu.fi/∼atkahkon/*.

[4]The problem is discussed for example in Hunter (2001) and Lindsay and Ehrenberg (1993). The Experimental Economics is a pleasant exception providing an outlet for such studies.

[5]See Tammi (2004) for extent discussion on coordination and diagnostic reasoning, and Zelmer (2003) for meta-analysis on linear public good experiments.

[6]See for example to Hunter (2001).

[7]See Hunter (2001), Lindsay and Ehrenberg (1993) and Posavac (2002) for discussion on topic and different conditions of replications.

[8]See Zizzo (2002) for details on the experimental design.

[9]Parts of the Java program and source code are available from the author on request.

[10]Instead of embedded MS Visual Basic Randomize calls used in initial study, a Java random function *(float)Math.random()* was used.

[11]Results of first questionnaire were used in econometric analysis, but they were found to be insignificant. However, a dummy variable Correct for subjects who answered correct to all ten questions was correlated (0.415) and significant (0.000) between subjects with experimental experience (Expany).

[12]I thank Daniel Zizzo for providing description of variable with instructions to compute it as well as ready computed values with original data.

[13]See legend of Table 2 for a short description of variables, and Zizzo (2002) for a more specific description.

[14]Results of more general Hausman's (1978) specification test for the random-effects are reported in parenthesis, difference bases on distinct test-statistics, however, the interpretation does not change.

[15]See Efron (1987) for introduction.

[16]However, if we divide subjects to two age-groups, over 26 years old and 26 and under, we observe that the former age-group invests significantly less ($\rho = -0.095$, $p = 0.001$). We also observe that Sex and Age are positively correlated ($\rho = 0.126$, $p = 0.000$). Yet another problem in analysis is due to zero investments (8 observations in replication and 10 observations in original experiment) which could be solved by using truncated regression or Tobit model. In addition, one player ran out of points in the replication experiment (4 observations). Possible explanation for zero investments in experiments is a simple rule of thumb or myopic heuristic decision of saving points for later rounds or trying to

drive competitor out of points. However, the zero investment is not *rationalizable strategy* since by investing nothing subjects lose round for sure. Strategy of playing zero is strictly dominated, and is never a best response implying that subjects are not perfectly rational which can partly explain the failure of HV predictions in experimental setting.

[17]See e.g. Searle (1987, Chapter 13) for a survey on mixed models.

[18]Estimation results available on request.

References

[1] Budd, C., C. Harris and J. Vickers (1993). "A Model of the Evolution of Duopoly: Does the Asymmetry between Firms Tend to Increase or Decrease?." *Review of Economic Studies*. 60, 3, 543–573.

[2] Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics*. 7, 1, 1–26.

[3] Efron, B. (1987). "Better Bootstrap Confidence Intervals." *Journal of American Statistical Association*. 82, 397, 171–185.

[4] Harris, C. and J. Vickers (1987). "Racing with Uncertainty." *Review of Economics Studies*. 54, 1, 1–21.

[5] Hausman, J.A. (1978). "Specification Tests in Econometrics." *Econometrica*. 46, 6, 1251–1271.

[6] Hunter, J.E. (2001). "The Desperate Need for Replications." *Journal of Consumer Research*. 28, 1, 149–158.

[7] Hörner, J. (2004). "A Perpetual Race to Stay Ahead." *Review of Economic Studies*. 71, 1065–1088.

[8] Kähkönen, A.T. (2004). "Patent Race Experiment Revisited." mimeo, paper presented at the ESA International Meeting 2004, Amsterdam, The Netherlands, June 2004.

[9] Lindsay, R.M and A.S.C. Ehrenberg (1993). "The Design of Replicated Studies." *American Statistican*. 47, 3, 217–228.

[10] Palfrey, T. and R. Porter (1991). "Guidelines for Submission of Manuscripts on Experimental Economics." *Econometrica*. 59, 4, 1197–1198.

[11] Posavac, E.J. (2002). "Using p Values to Estimate the Probability of a Statistically Significant Replication." *Understanding Statistics*. 1, 2, 101–112.

[12] Sbriglia, P. and J.D. Hey (1994). "Experiments in Multi-Stage R&D Competition." *Empirical Economics*. 19, 291–316.

[13] Searle, S.R. (1987). *Linear Models for Unbalanced Data.* New York: John Wiley & Sons, Inc.

[14] Spanos, A. (1999). *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data.* Cambridge, UK: Cambridge University Press.

[15] Tammi, T. (2004). "Good Practises in Experiment: Diagnostics and Coordination." mimeo, paper presented at the ESA International Meeting 2004, Amsterdam, The Netherlands, June 2004.

[16] Zelmer, J. (2003). "Linear Public Goods Experiments: A Meta Analysis." *Experimental Economics.* 6, 299–310.

[17] Zizzo, D.J. (2002). "Racing with Uncertainty: A Patent Race Experiment." *International Journal of Industrial Organization.* 20, 6, 877–902.