Replication of

Deferred Compensation in Multiperiod Labor Contracts: An Experimental Test of Lazear's Model

by Huck, S./Seltzer, A.J./Wallace, B. (2011)
in: The American Economic Review, 101(2), pp. 819–843

Replication Authors:

Colin Camerer, Taisuke Imai and Gidi Nave

Huck et al test a (Lazear) labor model of how deferred compensation can increase effort and surplus, if future payments are committed. There are four conditions: Enforcing commitment to pay future wages (FCT), removing commitment (NCT), having computer-generated wage offers (yoked to subject offers in CFT), and providing firm histories to allow reputation-building (RT). CFT does not produce interesting effects and RT is "of a more exploratory nature", and hence is not an ideal focus of replication. There is a weak change in social surplus from FCT to NCT.

Since a core novelty of the design is whether firm subjects offer deferred compensation, we focus on the values of W_2 and W_3.

> **Hypothesis to bet on:**
>
> The ability to pay future deferred compensation increases worker earnings ($W_2 + W_3$) more when commitment is enforced (FCT) compared to non-enforcement (NCT).

Power Analysis

The effect is tested using a Mann-Whitney test comparing ranks of entire-session averages (6 FCT versus 6 NCT). Table 3 (data line 2, comparing data columns 1 and 2) reports the average deferred wages for the two conditions and notes $p < 0.01$. The exact p-value of the deferred wage average ($W_2 + W_3$) is 0.0039. The original sample size is 120 participants. To achieve 90% power the required sample size is 151 participants.

Sample

The sample for replication consists of 160 participants (8 sessions à 20 subjects) from the Nuffield CESS subject pool in Oxford. In the original sample, "subjects in each treatment were students of any field at Royal Holloway and UCL [in UK]". We will not use any special inclusion or exclusion criteria to match this sample to the original Royal Holloway/UCL sample. After the fact, we will be able to test original and replication sample matches on some demographic variables which need to be gathered to complete the replication.

Materials

We use the material of the original experiment (programmed in z-Tree) along with the original instructions, both available at the journal's webpage.

Procedure

We follow the procedure of the original article, with only slight but unavoidable deviations as outlined below. The following summary of

the experimental procedure is therefore based on the section "II. Experimental Design and Procedures" (pp. 824–825) in the original study.

For the two interactive treatments (the FCT and NCT), ten subjects (five workers and five firms) participate in each session. Subjects are assigned the role of firm or worker at the beginning of the experiment and keep their role throughout the session. Each session consists of 20 rounds, with random matching of workers and firms which, in the case of the FCT and the NCT, approximates the nature of one-shot games in each round. We run six sessions for each of the treatments, and thus there are observations for 30 worker players and 30 firm players over 20 rounds for each treatment. The parameters of the model are set as follows: $p = 0.5$, $CL = 0$, $CM = 20$, $CH = 40$, I (the outside income of the worker) $= 40$, $ZL = 50$, $ZM = 100$, $ZH = 140$. Wage offers could range between 0 and 120 for each stage of the round.

In addition, for some analyses additional variables will be collected as in the original paper: *testscore*, the score attained on a pre-experimental quiz, which perhaps captures understanding of the game or speed of calculation; and several personal controls obtained in a post-experiment questionnaire *sex*, *age*, *studyyear* (undergraduate = 1, 2, 3; masters = 4; PhD = 5, 6, 7), and *economics* (1 if majoring in economics).

After all rounds have been played, subjects will be paid privately in cash based on the same incentives and using the same show up fee (£4) as in the original study (average earnings were £17.28 + show-up fee for worker players and £16.19 + show-up fee for firm players per subject in the original study)

Analysis

The analysis will be performed exactly as in the original article. Each session will generate an average sum of worker earnings. A Mann-Whitney rank-sum test will be conducted using the session averages from the FCT and NCT conditions.

Differences from Original Study

The replication procedure is identical to that of the original study, with some unavoidable deviations. The replication will be performed at Oxford university, while the original data was gathered at Royal Holloway and UCL, UK, in 2002. The experiment will be conducted in English as in the original study.

Replication Results

The total of 160 subjects (80 in FCT and 80 in NCT) participated in the replication experiments. The average worker earnings ($w_2 + w_3$) in FCT is 55.79 compared to 41.56 in NCT. A Mann-Whitney test (following the original analysis we use Kruskal-Wallis test, which is equivalent to Mann-Whitney test in this context) comparing ranks of entire-session averages yields a χ^2-squared statistic of 2.162 (with 1 degree of freedom). This result has a *p*-value equal to 0.142.

The difference in mean $w_2 + w_3$ is $54.74 - 21.31 = 33.43$ in the original study and $55.79 - 41.56 = 14.23$ in the replication study (both FCT − NCT). The relative effect size is therefore 42.57% (14.23/33.43).

Unplanned Protocol Deviations

We conducted two sessions of 10 subjects simultaneously (that means, there were 20 subjects in the experimental lab). Those two simultaneous sessions are of the same treatment (FCT or NCT). Apart from that the replication experiment has been conducted exactly the way as described above, without any deviations from the protocol.

Discussion

Given the criteria and procedure outlined above, the hypothesis of interest has not been replicated at a significance level of $\alpha < 5\%$. The relative effect size equals 42.57% and the *p*-value of the hypothesis test is 0.142.

As an additional analysis, we estimate the treatment effect using a linear regression. Note that this is not the planned analysis and has no bearing on the outcome of the prediction markets. However, further analyses inform comparison between empirical features of the original and replicated data. In this analysis, we regress the dependent variable $(w_2 + w_3)$ on the dummy variable which codes 1 if the treatment is NCT

$$(w_2 + w_3) = \beta_0 + \beta_1 \cdot \mathbb{1}_{\{Treatment=\text{NCT}\}} + \epsilon,$$

with standard errors clustered at subject level. The results are reported in Table 1. The 95% confidence interval of β_1 is $[-47.26, -19.59]$ in the original data and $[-27.70, -0.76]$ in the replication data.

Table 1: Regression results of the original and replication experiments

	Original Study	*Replication Study*
NCT dummy (β_1)	**−33.43**	**−14.23**
	(6.91)	(6.77)
Constant (β_0)	54.74	55.79
	(6.11)	(5.42)
# Clusters	60	80
$H_0 : \beta_1 = 0$	$p < 0.001$	$p = 0.039$