Available online at www.sciencedirect.com
SCIENCE DIRECT®
Labour Economics 10 (2003) 399–406
LABOUR ECONOMICS
www.elsevier.com/locate/econbase

Why labour market experiments?

Armin Falk*, Ernst Fehr

Institute for Empirical Research in Economics, University of Zurich, Blümlisalpstrasse 10, CH-8006 Zurich, Switzerland

Abstract

Over the last decades, there has been a steady increase in the use of experimental methods in economics. We discuss the advantages of experiments for labour economics in this paper. Control is the most important asset behind running experiments; no other empirical method allows a similarly tight control as do experiments. Moreover, experiments produce replicable evidence and permit the implementation of truly exogenous ceteris paribus changes. We also discuss frequent objections to experiments, such as a potential subject pool bias, the stake levels used in experiments, the number of observations as well as internal and external validity. We argue that although these objections are important, careful experimentation can circumvent them. While we think that lab and field experiments offer a very valuable tool, they should not be viewed as substitutes but as complements to more traditional methods of empirical economic analysis.

Keywords: Experiments; Labour economics

1. Advantages of laboratory experiments

Scientific progress relies crucially on the testing of theories. The researcher has different data sources available for performing such testing. Theses sources can be roughly classified along two dimensions (Friedman and Sunder, 1994). We can distinguish happenstance from experimental data, and field data from laboratory data. Happenstance data is the by-product of uncontrolled, naturally occurring economic activity. In contrast, experimental data is created explicitly for scientific purposes under controlled conditions. Field data is data from natural environments while laboratory data comes from lab environments. The two distinctions allow for four combinations, and all of these are used

* Corresponding author. Tel.: +41-1-6343722; fax: +41-1-6344907.
E-mail addresses: falk@iew.unizh.ch (A. Falk), efehr@iew.unizh.ch (E. Fehr).

0927-5371/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0927-5371(03)00050-2

in economics. In labour economics, the data most commonly used is field happenstance data such as unemployment rates or data on wages, education, or income.

Given the richness of happenstance field data, why should one bother creating own data and performing experiments? Performing an experiment is, after all, time-consuming and costly, as the participants of an experiment are paid based on their decisions. Let us illustrate the potential advantages of lab experiments with the help of a concrete example, the testing of tournament theory. Tournament incentives mean that workers compete for a prize, just as in a sport tournament. The worker with the higher performance gets the prize, e.g., a higher salary or a promotion. According to tournament theory, the workers' equilibrium efforts should be chosen in such a way that marginal effort costs equal marginal gains. The latter depend on the level of the prize and inversely on the importance of chance for getting the prize (Lazear and Rosen, 1981). Given the equilibrium effort choices, the calculation of the optimal prize is straightforward.

A direct empirical test of this theory requires that the researcher knows the number of workers who compete for the prize, the effort cost functions of the workers, the exact level of the prize, and the production function including the nature of the error term. Knowledge of the payoff function of the firm and the participation constraints of the workers is further necessary for the determination of the optimal prize level. All of these features can be implemented in a lab experiment and are therefore known by the researcher. Consequently, it is possible to derive a precise prediction and to test this prediction by observing effort and prize choices of subjects who participate in the experiment. With happenstance field data, conducting such a direct test seems impossible. The researcher never knows any of the ingredients mentioned above with a sufficient level of confidence when using field data. Moreover, he cannot be sure that the environment in which the interaction takes place is similar to that assumed by the theory. For example, it is well known in theory that if the workers can engage in sabotage activities, both the optimal level of effort as well as of the prize are different from a situation where sabotage is not part of the workers' strategy set. By the same token, the researcher who uses field data knows only little about whether the interaction between workers and the firm is one-shot or repeated. In repeated games, however, theoretical predictions are usually quite different from those derived in one-shot situations. Other factors may be important too. How intensely do workers know each other? Is a lot of communication and peer pressure going on, making collusion more likely? Or is the interaction rather anonymous? These and many other details of the environment affect behaviour in the field in an uncontrolled manner. In a lab experiment, on the other hand, these environmental factors can be controlled and systematically studied. It is easy to study a treatment where subjects interact only once in an experiment and to contrast this treatment with a condition where workers interact repeatedly. Likewise, it is possible to study anonymous vs. face-to-face interactions. The study of the impact of sabotage activities on effort choices in a controlled manner is also possible.

The existence of superior controls is, of course, not restricted to the study of tournaments. In the debate about inter-industry wage differentials, for instance, it has been notoriously difficult to judge whether the observed differentials reflect true rents or whether they are simply the result of unobserved heterogeneity (Gibbons and Katz, 1992). Likewise, it seems almost impossible on the basis of field data to judge whether unemployment is involuntary or voluntary, i.e., whether unemployed workers would in

fact be willing to work for less than the going wage. In contrast, it is possible to control workers' outside options in laboratory labour market experiments in such a way that these questions can be answered precisely (Fehr and Falk, 1999). We think that these examples show quite convincingly that the control possibilities available in laboratory experiments go substantially beyond the respective controls in the field.

Another major advantage of laboratory experiments derives from the possibility of implementing truly exogenous ceteris paribus changes. Many of the interesting variables are endogenously determined in the field, rendering causal inferences very difficult. Often, the best that can be achieved with field happenstance data is that the variables of interest are correlated. Finally, the precise replicability of experimental evidence is also of great importance. The experimenter controls the conditions under which the evidence is generated and reports all details in his study. It is therefore easy to replicate the experiment and to verify or falsify the claims made in a particular study.

2. Objections to laboratory experiments

When presenting an experimental study in a seminar or on a conference, frequently the following objections are made: first it is criticised that there is a subject pool bias because the experimental subjects are often students. This criticism is frequently justified using the argument that student subjects, unlike professionals, have no experience with the problem at hand. Second, people object that subjects do not take their decisions seriously because the stakes in the experiment are too low. Third, the results are questioned because of the small number of participants. In our view, these objections may be justified and should, therefore, be taken seriously. However, the objections are not fundamental because careful experimentation can circumvent them.

The reason why experimentalists rely primarily on students as a subject pool has to do with convenience. Students are easy to recruit, have a quick understanding of the rules of the experiment, and rather low opportunity costs. However, there is no fundamental reason for excluding other subjects, and nonstudent subject pools are used increasingly. Moreover, it is possible to find interesting behavioural differences across subject pools in a precise way with careful experimentation. For instance, Cooper et al. (1999) examined the ratchet effect arising from piece rate incentives with Chinese students and Chinese middle managers as experimental participants. List (2003) examined whether the endowment effect (Kahneman et al., 1986) is removed by market experience. To study this question, he recruited inexperienced and experienced traders who trade sports memorabilia at sports-card shows. Likewise, Fehr and List (2003) recruited CEOs for one of their experiments to study the extent to which CEOs use explicit incentives and how they respond to these incentives. They were particularly interested in the question as to what extent certain kinds of explicit incentives are counterproductive and, if so, whether students and CEOs use such counterproductive incentives.

Cooper et al. (1999) observed that experimental behaviour slowly, yet incompletely, converged both with managers and with students towards the pooling equilibrium. Thus, the behavioural differences across subject pools vanished over time. In the early periods of their experiments, they observed some subject pool differences, however. If the exper-

imental instructions were formulated in an abstract context-free language that was removed from the daily context in which the managers' decisions took place, student behaviour converged more quickly toward the equilibrium than the managers' behaviour. In contrast, if the instructions made explicit reference to the interactions between planners and managers, the managers' behaviour converged more quickly towards the equilibrium. The results of List (2003) also indicate that traders with little market experience exhibit an endowment effect whereas no endowment effect could be found in experienced traders. List was able to show that the absence of an endowment effect among experienced traders is not due to selection effects but to the market experience per se. Fehr and List (2003) also found differences in the behaviour of students and CEOs. CEOs were significantly more trusting and trustworthy than students. However, both subject pools predominantly used the available explicit incentive, even though this had strong negative side effects and decreased their earnings.

All these examples suggest that subject pool differences may be a real issue. However, the studies also show that the different subject pools do not behave in fundamentally different ways. After some time, students' and managers' behaviour was very similar in the Cooper et al. (1999) study. Likewise, in Fehr and List (2003), both the students and the managers exhibited a lot of trusting behaviour, which should have been absent if all players are assumed to be purely selfish. In addition, both subject pools exhibited a lot of non-selfish, reciprocally fair, behaviour. Thus, although there are some quantitative subject pool effects, the qualitative patterns of behaviour were rather similar across the different pools.

The role of stake levels is also an important issue. There is agreement among experimentalists that higher stake levels typically reduce the behavioural variance (Camerer and Hogarth, 1999). This suggests that at higher stake levels, subjects become more focused. However, Camerer and Hogarth, who survey dozens of studies examining the effects of stake size, find that the central behavioural tendencies are rarely overturned by changes in the stake level. If some money is at stake, variations in the stake size frequently do not change the average behaviour. Recently, an interesting and important exception (Holt and Laury, 2002) showed that subjects are significantly more risk averse when the average earnings in a lottery-choice experiment increased from roughly US $70 to $230. In contrast, in the context of fairness preferences, increases in the stake level have little or no impact on the incidence of fair responses (Cameron, 1999; Slonim and Roth, 1998; Hoffman et al., 1996; Fehr and Tougareva, 1995).

Finally, the objection can also be circumvented that subject pools are often small relative to the data sets relying on happenstance field data. First of all, it is always possible to increase the number of observations with additional experiments. Secondly, it is possible to conduct large-scale experiments that are representative for whole countries. Recently, several researchers have conducted representative experiments. Harrison, Lau and Williams (2002) elicited discount rates for a representative sample of the Danish population while Bellemare and Kröger (2003) conducted trust games, which are representative for the Netherlands. Interestingly, the results of Bellemare and Kröger (2003) indicate that no participation selection bias exists for their laboratory experiment.

A more general critique against laboratory experiments concerns their *internal and external validity*. Internal validity concerns the question whether the data gathered in an experiment permits causal inferences. Internal validity is therefore simply a matter of

proper experimental controls, a sensible experimental design, and a correct data analysis. The objection that experiments are not externally valid is more serious. Sceptics doubt that it is possible to generalise inferences from the laboratory to the field. In our view, external validity raises two questions. On a very general level, external validity holds only if the principle of induction holds. According to this principle, behavioural regularities will persist in new situations as long as the relevant underlying conditions remain substantially unchanged. It is true that we can never be sure that an experimental result will be replicable, even if we keep all conditions constant. While this is true, however, this critique is not confined to experiments. In fact *no* empirical result whatsoever can guarantee that the same result will be replicated again under identical circumstances.

The second problem concerning external validity concerns the question of whether the experiment captures the essential conditions that prevail in reality. A sceptic could argue that a particular experiment is not externally valid and not realistic because it misses decisive features. The experimentalist's response should be to try implementing the neglected conditions. However, one should also keep in mind that experiments as well as economic models are deliberately unrealistic in the sense that they abstract from reality. The simplicity of a model or an experiment is often a virtue because it enhances the understanding of the interaction of relevant variables. Moreover, whether realism is important or not depends on the purpose of the experiment. Often the purpose is to test a theory or to understand its failure. In this case, the evidence is important for theory building but not for a direct understanding of reality. Let us conclude our discussion on realism with a quote from one of the pioneers in experimental economics: "The art of posing questions rests on an ability to make the study of simple special cases relevant to an understanding of the complex. General theories and models by definition apply to all special cases. Therefore, general theories and models should be expected to work in the special cases of laboratory markets. As models fail to capture what is observed in the special cases, they can be modified or rejected in light of experience. The relevance of experimental methods is thereby established." (Plott, 1982, p. 1509) He adds: "While laboratory processes are simple in comparison to naturally occurring processes, they are real processes in the sense that real people participate for real and substantial profits and follow real rules in doing so. It is precisely because they are real that they are interesting." (Plott, 1982, p. 1482).

3. Adding realism

While the superior control possibilities of experiments are beyond doubt, the question whether the conditions implemented in the laboratory are also present in reality will probably always be subject to some uncertainty and debate. This is one reason why lab experiments should not be viewed as substitutes but as complements to more traditional methods of empirical economic analysis. It is, however, often possible to combine the empirical rigor of lab experiments with the realism of field data. An obvious route is the conduct of field experiments, which allows the researcher to study the behaviour of economic agents in their natural environment. A recent paper studies the impact of peer pressure on work behaviour in a setting that guarantees control *and* realism (Falk and Ichino, 2003). In their study, Falk and Ichino observe the output of subjects working on a

regular side-job, where in one treatment, two subjects work at the same time in the same room, while in a control treatment, each subject works alone in a room. The tasks of the subjects are technologically completely independent from each other. Among other things, they find that working in randomly selected pairs increases output significantly compared to the output of subjects who work alone by themselves. In another field experiment, Fehr and Götte (2002) study the inter-temporal substitution of working time and effort of workers in a firm. They pay 50% of the firms' employees a higher piece rate for 1 month, whereas the other 50% of the employees receive a constant piece rate. In this way, they create a control and a treatment group, allowing them to measure how the piece rate increase changes labour supply.

Another way of adding more realism in the context of *laboratory* experiments is to conduct so-called "real effort" experiments. In a typical lab experiment, the choice of work effort is represented by an increasing monetary function, i.e., instead of choosing real efforts subjects choose a costly number. This procedure has been used in tournament experiments (e.g., Bull et al., 1987 or Harbring and Irlenbusch, in press) or in efficiency wage experiments (e.g., Fehr and Falk, 1999). In real effort experiments, effort is an actual task. In Fahr and Irlenbusch (2000), e.g., subjects had to crack walnuts, in van Dijk, Sonnemans and van Winden (2001) subjects performed cognitively demanding tasks on the computer (two-variable optimisation problems) and in Gneezy (2003) subjects had to solve mazes at the computer. While "real effort" surely adds realism to the experiment, one should also note that it is realised at the cost of losing control. Since the experimenter does not know the workers' effort cost, it is not possible to derive precise quantitative predictions.

Experimental techniques are also valuable for programme evaluation. Unfortunately, data from large-scale randomised social programme experiments are rarely available, in particular in Europe where policymakers are hesitant to perform such social experiments. However, it is possible to collect own data. Falk et al. (2002), e.g., present a novel approach to programme evaluation that allows identification of the causal effect of a training programme on the likelihood of being invited to a job interview under weak assumptions. The idea is to measure the programme-effects of pre- and post-treatment data that are very close in time for the same individual. Their procedure is as follows: first, they recruited unemployed persons who participated in a particular programme. Then they sent out applications for these people before they finished their course. A second wave of applications was sent after the participants had successfully completed the course and had received a certificate. The new applications were exactly the same as before, except that the application now also contained the course certificate. The impact of the programme is measured by the firms' responses. They checked whether the probability of being invited to a job interview was different for the applications with certificate than for those without. A similar approach has been used to detect discrimination against female workers (Weichselbaumer, in press) and unemployment stigma (Oberholzer-Gee, 2000).

4. The future of experimentation

Over the last two decades, there has been a steady increase in the applications of laboratory methods on economic questions. We expect that this upwards trend will

continue. In fact, the award of the Nobel prize to Daniel Kahneman and Vernon Smith is likely to increase the speed with which economic departments may be willing to hire experimental economists. There are many areas of economics where laboratory methods can be applied fruitfully, among them many areas related to labour economics. One of the great strengths of economic theory is that it enables us to predict what will happen under institutions or parameter constellations (e.g., tax rates) which have never existed so far but which one might want to implement. One of the strengths of experimental methods is that we can empirically study the effects of different institutional environments, as defined by their rules and incentives. The precise observation of human behaviour in experiments also forces the researcher to take issues of human motivation and bounded rationality more seriously. In the long run, this is likely to generate a more realistic picture of human nature. Phenomena like fairness, reciprocity, loss aversion, over-confidence, non-exponential discounting, etc. have been studied successfully in the laboratory and, by studying their impact on incentives and contracts, on the design of organisations, on labour supply and labour demand, they may enhance our understanding of how firms, households, and labour markets function.

References

Bellemare, C., Kröger, S., 2003. On representative trust. Working Paper. Tilburg University.

Bull, C., Schotter, A., Weigelt, K., 1987. Tournaments and piece rates: an experimental study. Journal of Political Economy 95, 1–33.

Camerer, C.F., Hogarth, a.R.M., 1999. The effects of financial incentives in experiments: a review and capital–labor–production framework. Journal of Risk and Uncertainty, 7–42.

Cameron, L.A., 1999. Raising the stakes in the ultimatum game: experimental evidence from Indonesia. Economic Inquiry 37, 47–59.

Cooper, D.J., Kagel, J.H., Lo, W., Gu, Q.L., 1999. Gamint against managers in incentive systems: experimental results with Chinese students and Chinese managers. American Economic Review 89, 781–804.

Fahr, R., Irlenbusch, B., 2000. Fairness as a constraint on trust and reciprocity: earned property rights in a reciprocal exchange experiment. Economic Letters 66, 275–282.

Falk, A., Ichino, A., 2003. Clean Evidence on Peer Pressure. CEPR Working Paper DP3834.

Falk, A., Lalive, A., Zweimüller, J., 2002. The Success of Job Applications: A New Approach to Program Evaluation. Working Paper No. 131. Institute for Empirical Research in Economics, University of Zurich.

Fehr, E., Falk, A., 1999. Wage rigidity in a competitive incomplete contract market. Journal of Political Economy 107, 106–134.

Fehr, E., Götte, L., 2002. Do Workers Work More if Wages are High? Evidence from a Randomized Field Experiment. WP 125. Institute for Empirical Research in Economics, University of Zurich, Zurich.

Fehr, E., List, J., 2003. The Hidden Costs and Returns of Incentives—Trust and Trustworthiness among CEOs. WP 134. Institute for Empirical Research in Economics, University of Zurich, Zurich.

Fehr, E., Tougareva, E., 1995. Do Competitive Markets with High Stakes Remove Reciprocal Fairness?—Evidence from Russia. Discussion Paper. University of Zurich.

Friedman, D., Sunder, S., 1994. Experimental Methods. A Primer for Economists. Cambridge Univ. Press, Cambridge.

Gibbons, R., Katz, L., 1992. Does unmeasured ability explain inter-industry wage differentials. Review of Economic Studies 59, 515–535.

Gneezy, U., 2003. Do High Wages Lead to High Profits? An Experimental Study of Reciprocity Using Real Effort. The University of Chicago GSB, Chicago.

Harbring, C., Irlenbusch, B., 2003. An experimental study on tournament design. Labour Economics 10, 443–464 (this issue).

Harrison, G.W., Lau, M.I., Williams, M.B., 2002. Estimating individual discount rates in Denmark: a field experiment. American Economic Review, 1606–1617.

Hoffman, E., McCabe, K., Smith, V., 1996. On expectations and the monetary stakes in ultimatum games. International Journal of Game Theory 25, 289–301.

Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. American Economic Review, 1644–1655.

Kahneman, D., Knetsch, J.L., Thaler, R., 1986. Fairness as a constraint on profit seeking: entitlements in the market. American Economic Review 76, 728–741.

Lazear, E., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. Journal of Political Economy 89, 841–864.

List, J.A., 2003. Does market experience eliminate market anomalies. Quarterly Journal of Economics 118, 41–71.

Oberholzer-Gee, F., 2000. Unemployment Stigma and Rational Herding: Hiring Practices in Europe and in the United States, mimeo. University of Pennsylvania.

Plott, C.R., 1982. Industrial organization theory and experimental economics. Journal of Economic Literature 20, 1485–1527.

Slonim, R., Roth, A.E., 1998. Learning in high stakes ultimatum games: an experiment in the Slovak Republic. Econometrica 66, 569–596.

van Dijk, F., Sonnemans, J., van Winden, F., 2001. Incentive systems in a real effort experiment. European Economic Review, 187–214.

Weichselbaumer, D., 2002. Is it sex or personality? The impact of sex-stereotypes on discrimination in applicant selection. Eastern Economic Journal (in press).