Chapter 14
Experimental Data and Econometrics: Outline

Peter G. Moffatt, Anna Conte

Abstract The chapter charts the evolution of the range of econometric techniques that have been applied by Experimental Economists. The most commonly used technique has been treatment testing, and a good deal of attention is paid to studies using this approach, which includes standard parametric and non-parametric treatment tests, tests comparing entire distributions, and bootstrap tests. Power analysis has dramatically increased in importance in recent years, and applications of power analysis will be covered from both theoretical and practical perspectives. The decision over which econometric technique is appropriate for a given application is often determined by the data type. We therefore distinguish the data types arising in Experimental Economics (binary, ordinal, interval, etc.), and identify appropriate techniques, with key examples from the literature. We then consider the recent growth in popularity of fully structural models, with particular attention paid to studies that estimate of social preference parameters from dictator game data, and those that estimate risk aversion parameters while assuming between-subject heterogeneity in risk aversion. In these contexts, we pay close attention the method of maximum simulated likelihood (MSL) which has become a popular method for estimating models with continuous heterogeneity. We then consider studies that have applied finite mixture models as a way of capturing discrete heterogeneity; that is, when the population of subjects divides into a small number of distinct types. Most of the techniques covered in the chapter are established econometric techniques that have been adopted by experimentalists. However, a small number of techniques can claim to be unique to experimental economics. Examples of these are the Quantal Response Equilibrium (QRE) Model, and models of learning in games. Applications of these models are described in some detail.

Peter G. Moffatt ✉
University of East Anglia, Norwich, United Kingdom, e-mail: p.moffatt@uea.ac.uk

Anna Conte
Sapienza University of Rome, Rome, Italy, e-mail: anna.conte@uniroma1.it

14.1 Introduction

This chapter attempts to chart the evolution of econometric methods that have been applied to the analysis of data from Economic Experiments. This area of research is sometimes referred to as 'Experimetrics' (Moffatt, 2015).

An important point needs to be made clear at the outset. The majority of econometric methods applied in the analysis of Experimental data are methods that previously existed in the econometrics and/or statistics literatures. We are clearly interested in these applications of econometric techniques. However, some techniques have been developed exclusively for particular Experimental applications, and we will pay particularly close attention to these.

The assessment of which techniques have 'aged well' and which have not will be dealt with as far as feasible given the fairly recent development of the field of Experimetrics.

14.2 Treatment Testing

The most common approach within Experimetrics is the treatment testing approach. In certain types of experiment, there are compelling reasons for considering this to be the natural approach. Provided that subjects are assigned randomly to treatments, and that the experimental environment is such that all influences on behaviour other than the treatment of interest are held fixed, then it is fair to say that there is no need to address many of the types of problem that econometricians have traditionally been interested in, such as sample selection bias, measurement error, and endogeneity.

There are many different ways of conducting treatment tests. Siegel & Castellan (1988) provided a fairly complete practical guide to these tests. Fully parametric tests such as the independent-samples t-test, which amounts to a comparison of two sample means, may be the first choice. The analogous test in a situation of paired data is the paired-comparison t-test. These tests rely on strong distributional assumptions, most importantly normality of the decision variable, and in situations in which normality does not hold, they require large sample sizes. There are many situations in which these conditions are unlikely to hold, and for this reason, many experimental researchers have a preference for non-parametric alternatives such as the Mann-Whitney test, which uses ordinal and not cardinal information, and essentially amounts to a test of equality of medians. An alternative approach which does not rely on distributional assumptions but which makes full use of the cardinal information in the data is the bootstrap technique (Ellingsen & Johannesson, 2004). A similar approach which has received attention recently is that of permutation testing (Holt & Sullivan, 2023), the most simple example of which is Fisher's Exact test used for discrete outcomes.

Tests comparing variances are also potentially useful in some settings, for example when one treatment places subjects in a less familiar situation than the other (Moffatt, 2019). In a situation in which it is unclear which functional of the distribution (e.g. mean, median, variance) is impacted by the treatment, a test of the equality of entire distributions may be preferred, the most popular being the Kolmogorov-Smirnov test and the Epps-Singleton test (Forsythe et al., 1994; Eckel & Grossman, 2000).

Some treatment tests have been introduced explicitly for data from economic experiments. One example is the Conlisk test Conlisk (1989) which has been useful in testing Allais Paradox (Allais, 1953) and the Preference Reversal phenomenon.

14.2.1 The Rise and Rise of Power Analysis

Closely related to treatment testing is power analysis, which is becoming increasingly important in experimental studies. The concept of power also plays a key role in the ongoing scientific quality debate (C. F. Camerer et al., 2016). Broadly, power analysis can be used for two different purposes: first, to determine the sample size that is required to attain a given, pre-determined, level of power; second, to compute the power of a test that has already been conducted. In both cases, a power of at least 0.8 has become accepted the benchmark for adequacy, although Zhang & Ortmann (2013) have presented evidence that this target has rarely been attained in practice. For most simple tests, power calculations are simple, and routines for computing power are available in popular software packages. In more complicated settings, for example, treatment tests being conducted in the framework of panel models, multi-level models or structural models, Monte Carlo simulations are useful in performing power calculations. One particularly useful application of monte carlo methods in this context is in assessing the relative benefits, in terms of power, of increasing the number of subjects versus increasing the number of tasks (Moffatt, 2021). Another design feature that determines power is the size of the monetary incentives. To take account of this would require incorporating an experimental budget constraint into the power calculation (Alekseev, 2023). Optimal design theory has also been used to design the choice problems within an experiment (Moffatt, 2007; Bland, 2023).

14.3 Theory Testing

In some areas of experimental economics, the focus is on a test of an economic theory, rather than the estimation of ‘home-grown’ preference parameters of the experimental subjects. The most common types of experiment falling into this category are market experiments (Smith, 1962) and auction experiments (Kagel & Levin, 1986; Ham et al., 2005).

14.4 Applications of Binary, Ordered, Interval and Censored Models

Many different types of data have arisen in Experimental Economics. The decision variable may be discrete or continuous. A discrete variable may be binary, for example the choice between a risky and safe alternative (Hey and Orme, 1994). Or it may be categorical, for example the choice between a number of different allocations (Engelmann and Strobel, 2004). Or it may be ordinal, for example, when the subject has been asked to reveal their emotions when participating in experimental games (Bosman and van Winden, 2002). A continuous variable arises when the subject has been asked how much to contribute to a public fund (Bardsley and Moffatt, 2007), or asked for their willingness to pay for an object (Isoni et al., 2011). Even continuous responses contain elements of discreteness, in the form of censoring, lumpiness, and focal points. Methods for allowing for these data features in estimation are covered in detail by Moffatt (2015).

14.5 Structural Estimation of Preference Parameters

A highly popular method for eliciting preference is the multiple price list (MPL), of which the most well-known is that of Holt & Laury (2002). Early applications of the method included elicitation of willingness to pay for a good (Kahneman et al., 1990), measurement of risk attitudes (Binswanger, 1980), and elicitation of individual discount rates (Coller & Williams, 1999). A variant of the MPL which has also become popular is the switching multiple price list (sMPL) (Andersen et al., 2006) which induces subjects to report a single switch-point (rather than a set of choices) when presented with a MPL. This switch-point implies range of values of the preference parameter and interval regression is the appropriate econometric approach for estimating the underlying distribution. When more than one sMPL is used, it becomes possible to estimate the parameters of the joint distribution of more than one preference parameter Tanaka et al. (2010). This presents new and interesting econometric challenges: Conte et al. (2026) have used Monte Carlo integration with importance sampling to estimate the joint distribution of risk aversion and probability weighting using data from two sMPL's.

A shortcoming of the analysis of data from switching multiple price lists, emphasised by Conte et al. (2026), is that although they are useful for estimating the distribution of preference parameters, such analysis, when correctly performed, does not allow for within-subject variation. That is, it is assumed that, conditional on the preference parameters, decisions are made without error.

In order to allow for within-subject variation, a different sort of design is required, consisting of a sequence of choice problems, presented independently, and not in an ordered sequence. The approach to choosing the problems might be based on the requirement to cover the whole of the Marschak-Machina Triangle, as is the approach of Hey & Orme (1994) and others. There are broadly two ways of modelling within-subject variation: the Random Preference model (Loomes et al., 2002) in which the preference parameters are allowed to vary from one decision to the next; and the Fechner model (Hey & Orme, 1994) in which each decision is subject to computational error.

Which of Fechner and Random Preference has been the superior approach for modelling individual decisions has been the subject of debate for almost three decades. Apesteguia & Ballester (2018) provide a theoretical basis for preferring the Random Preference approach over the Fechner approach. Both of the modelling approaches have been very useful in addressing central questions in decision theory such as whether expected utility (EU) theory holds, which non-EU theory best explains the data, which error specification best explains the data, and the ways in which behaviour under risk changes with decision making experience (Loomes et al., 2002; Conte et al., 2011).

The method of maximum simulated likelihood (MSL) has become a popular estimation method in this setting (Train, 2009; Moffatt, 2015). The method of MSL is particularly useful when more than one preference parameter is assumed to vary, for example risk aversion and probability weighting (Conte et al., 2011). Von Gaudecker et al. (2011) assume four dimensions of between-subject heterogeneity: risk aversion, loss aversion, preference toward the timing of uncertainty resolution, propensity to choose randomly. The same sort of estimation methods are useful in allowing for heterogeneity in dicount rates (Andersen et al., 2008), complexity aversion (Moffatt et al., 2015), Inequity aversion (Gauriot et al., 2020), and propensity to update beliefs (Henckel et al., 2022).

The joint estimation of more than one preference parameter has been found to be highly important in certain contexts. For example, it has recently been found that allowing for risk aversion instead of risk neutrality has the effect of greatly reducing estimates of individual discount rates in time preference modelling (Andersen et al., 2008), and also of reducing estimates of inequity aversion in the analysis of dictator game giving (Gauriot et al., 2020).

14.6 Finite Mixture Modelling

One way of modelling subject heterogeneity is to assume that the population of subjects divide into a small number of discrete ‘types’. There are a number of different estimation methods for categorising individuals to types. The first is the finite mixture approach, in which the types, and a decision rule defining each type, are specified a priori, and ML estimation is used to estimate the unknown parameters of the decision rules together with mixing proportions. This approach has been highly useful in the analysis of data from public goods games (El-Gamal & Grether, 1995; Bardsley & Moffatt, 2007), fairness experiments (Conte & Moffatt, 2014), and in levels-of-reasoning models (Bosch-Domènech et al., 2010).

Another approach to finite mixture modelling is is Bayesian estimation, in which Bayesian updating rules are used to estimate the probability of a given subject using each decision rule. This approach claims to avoid the problem of non-uniqueness of the MLE which arises with the finite mixture approach. The Bayesian approach has been used by Houser et al. (2004) to classify subjects solving difficult decision problems into three types: near-rational; fatalist; and confused.

14.7 Econometric Modelling of Behaviour in Games

14.7.1 Quantal Response Equilibrium

The Quantal Response Equilibrium (QRE) model was developed by McKelvey & Palfrey (1995). It is based on the idea that best responses are not played with certainty. Each player is assumed to calculate the expected value of each of her available actions given her (correct, in equilibrium) belief about her opponent’s choice probabilities, and she attempts to respond optimally, but makes random errors in the process. Thus QRE replaces the perfectly rational expectations equilibrium embodied in Nash equilibrium with an imperfect, or “noisy”, rational expectations equilibrium. The principle of an equilibrium is maintained by assuming that players estimate expected payoffs in an unbiased way.

14.7.2 Learning Models

A number of learning models have appeared in the literature, including Directional learning (Selten & Stoecker, 1986), Reinforcement Learning (Erev & Roth, 1998), Belief Learning (Cheung & Friedman, 1998) and Experience Weighted Attraction

(C. Camerer & Hua Ho, 1999).

Wilcox (2006) shows that when heterogeneity is neglected in estimation, a severe estimation bias results in such a way as to favour Reinforcement Learning over Belief Learning.

14.8 Summary and Conclusion

References

Alekseev, A. (2023). The (statistical) power of incentives. *Available at SSRN 4088605*.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: journal of the Econometric Society*, 503–546.

Andersen, S., Harrison, G. W., Lau, M. I. & Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, *9*(4), 383–405.

Andersen, S., Harrison, G. W., Lau, M. I. & Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, *76*(3), 583–618.

Apesteguia, J. & Ballester, M. A. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy*, *126*(1), 74–106.

Bardsley, N. & Moffatt, P. G. (2007). The experimetrics of public goods: Inferring motivations from contributions. *Theory and Decision*, *62*(2), 161–193.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural india. *American journal of agricultural economics*, *62*(3), 395–407.

Bland, J. R. (2023). Optimizing experiment design for estimating parametric models in economic experiments. *Available at SSRN 4566855*.

Bosch-Domènech, A., Montalvo, J. G., Nagel, R. & Satorra, A. (2010). A finite mixture analysis of beauty-contest data using generalized beta distributions. *Experimental economics*, *13*(4), 461–475.

Camerer, C. & Hua Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, *67*(4), 827–874.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., ... others (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436.

Cheung, Y.-W. & Friedman, D. (1998). A comparison of learning and replicator dynamics using experimental data. *Journal of economic behavior & organization*, *35*(3), 263–280.

Coller, M. & Williams, M. B. (1999). Eliciting individual discount rates. *Experimental economics*, *2*(2), 107–127.

Conlisk, J. (1989). Three variants on the allais example. *The American Economic Review*, 392–407.

Conte, A., Hey, J. D. & Moffatt, P. G. (2011). Mixture models of choice under risk. *Journal of Econometrics*, *162*(1), 79–88.

Conte, A. & Moffatt, P. G. (2014). The econometric modelling of social preferences. *Theory and decision*, *76*(1), 119–145.

Conte, A., Moffatt, P. G. & Riddel, M. (2026). The multivariate random preference estimator for switching pultiple price list data. *Experimantal Economics*, *X*(X), X.

Eckel, C. C. & Grossman, P. J. (2000). Volunteers and pseudo-volunteers: The effect of recruitment method in dictator experiments. *Experimental Economics*, *3*(2), 107–120.

El-Gamal, M. A. & Grether, D. M. (1995). Are people bayesian? uncovering behavioral strategies. *Journal of the American statistical Association*, *90*(432), 1137–1145.

Ellingsen, T. & Johannesson, M. (2004). Promises, threats and fairness. *The Economic Journal*, *114*(495), 397–420.

Erev, I. & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, 848–881.

Forsythe, R., Horowitz, J. L., Savin, N. E. & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic behavior*, *6*(3), 347–369.

Gauriot, R., Heger, S. A. & Slonim, R. (2020). Altruism or diminishing marginal utility? *Journal of Economic Behavior & Organization*, *180*, 24–48.

Ham, J. C., Kagel, J. H. & Lehrer, S. F. (2005). Randomization, endogeneity and laboratory experiments: The role of cash balances in private value auctions. *Journal of Econometrics*, *125*(1-2), 175–205.

Henckel, T., Menzies, G. D., Moffatt, P. G. & Zizzo, D. J. (2022). Belief adjustment: A double hurdle model and experimental evidence. *Experimental Economics*, *25*(1), 26–67.

Hey, J. D. & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica: Journal of the Econometric Society*, 1291–1326.

Holt, C. A. & Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, *92*(5), 1644–1655.

Holt, C. A. & Sullivan, S. P. (2023). Permutation tests for experimental data. *Experimental economics*, *26*(4), 775–812.

Houser, D., Keane, M. & McCabe, K. (2004). Behavior in a dynamic decision problem: An analysis of experimental evidence using a bayesian type classification algorithm. *Econometrica*, *72*(3), 781–822.

Kagel, J. H. & Levin, D. (1986). The winner's curse and public information in common value auctions. *The American economic review*, 894–920.

Kahneman, D., Knetsch, J. L. & Thaler, R. H. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of political Economy*, *98*(6), 1325–1348.

Loomes, G., Moffatt, P. G. & Sugden, R. (2002). A microeconometric test of alternative stochastic theories of risky choice. *Journal of risk and Uncertainty*, *24*(2), 103–130.

McKelvey, R. D. & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and economic behavior*, *10*(1), 6–38.

Moffatt, P. G. (2007). Optimal experimental design in models of decision and choice. *MEASUREMENT IN ECONOMICS*, *357*.

Moffatt, P. G. (2015). *Experimetrics: Econometrics for experimental economics*. Macmillan International Higher Education.

Moffatt, P. G. (2019). Data analysis. In *Handbook of research methods and applications in experimental economics* (pp. 57–82). Edward Elgar Publishing.

Moffatt, P. G. (2021). Experimetrics: a survey. *Foundations and Trends in Econometrics*, *11*(1-2), 1–152.

Moffatt, P. G., Sitzia, S. & Zizzo, D. J. (2015). Heterogeneity in preferences towards complexity. *Journal of Risk and Uncertainty*, *51*(2), 147–170.

Selten, R. & Stoecker, R. (1986). End behavior in sequences of finite.

Siegel, S. & Castellan. (1988). *Nonparametric statistics for the social sciences.* New York: McGraw-Hill.

Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of political economy*, *70*(2), 111–137.

Tanaka, T., Camerer, C. F. & Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from vietnam. *American economic review*, *100*(1), 557–571.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Von Gaudecker, H.-M., Van Soest, A. & Wengström, E. (2011). Heterogeneity in risky choice behavior in a broad population. *American Economic Review*, *101*(2), 664–694.

Wilcox, N. T. (2006). Theories of learning in games and heterogeneity bias. *Econometrica*, *74*(5), 1271–1292.

Zhang, L. & Ortmann, A. (2013). Exploring the meaning of significance in experimental economics. *UNSW Australian School of Business Research Paper*(2013-32).