Experimetrics: A Survey

Other titles in Foundations and Trends® in Econometrics

Climate Econometrics: An Overview
Jennifer L. Castle and David F. Hendry
ISBN: 978-1-68083-708-7

Foundations of Stated Preference Elicitation: Consumer Behavior and Choice-based Conjoint Analysis
Moshe Ben-Akiva, Daniel McFadden and Kenneth Train
ISBN: 978-1-68083-526-7

Structural Econometrics of Auctions: A Review
Matthew L. Gentry, Timothy P. Hubbard, Denis Nekipelov and Harry J. Paarsch
ISBN: 978-1-68083-446-8

Data Visualization and Health Econometrics
Andrew M. Jones
ISBN: 978-1-68083-318-8

Experimetrics: A Survey

Peter G. Moffatt
University of East Anglia
UK
P.Moffatt@uea.ac.uk

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Econometrics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

P. J. Moffatt. *Experimetrics: A Survey.* Foundations and Trends® in Econometrics, vol. 11, no. 1–2, pp. 1–152, 2021.

ISBN: 978-1-68083-793-3

Foundations and Trends® in Econometrics
Volume 11, Issue 1–2, 2021

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Econometrics publishes survey and tutorial articles in the following topics:

- Econometric Models
- Simultaneous Equation Models
- Estimation Frameworks
- Biased Estimation
- Computational Problems
- Microeconometrics
- Treatment Modeling
- Discrete Choice Modeling
- Models for Count Data
- Duration Models
- Limited Dependent Variables
- Panel Data
- Time Series Analysis
- Latent Variable Models
- Qualitative Response Models
- Hypothesis Testing
- Econometric Theory
- Financial Econometrics
- Measurement Error in Survey Data
- Productivity Measurement and Analysis
- Semiparametric and Nonparametric Estimation
- Bootstrap Methods
- Nonstationary Time Series
- Robust Estimation

Information for Librarians

Foundations and Trends® in Econometrics, 2021, Volume 11, 4 issues. ISSN paper version 1551-3076. ISSN online version 1551-3084. Also available as a combined paper and online subscription.

Contents

Experimetrics: A Survey

Peter G. Moffatt

University of East Anglia, UK; P.Moffatt@uea.ac.uk

ABSTRACT

This monograph aims to survey a range of econometric techniques that are currently being used by experimental economists. It is likely to be of interest both to experimental economists who are keen to expand their skill sets, and also the wider econometrics community who may be interested to learn the sort of econometric techniques that are currently being used by Experimentalists. Techniques covered range from the simple to the fairly advanced. The monograph starts with an overview of treatment testing. A range of treatment tests will be illustrated using the example of a dictator-game giving experiment in which there is a communication treatment. Standard parametric and non-parametric treatment tests, tests comparing entire distributions, and bootstrap tests will all be covered. It will then be demonstrated that treatment tests can be performed in a regression framework, and the important concept of clustering will be explained. The multilevel modelling framework will also be covered, as a means of dealing with more than one level of clustering. Power analysis will be covered from both theoretical and practical perspectives, as a means of determining the sample size required to attain a given power, and also as a means of computing ex-post power for a reported test. We then progress to a discussion of different data types arising in Experimental Economics (binary, ordinal, interval, etc.), and how to deal with them. We then consider the estimation of fully structural models, with particular attention paid to

Peter G. Moffatt (2021), “Experimetrics: A Survey”, Foundations and Trends® in Econometrics: Vol. 11, No. 1–2, pp 1–152. DOI: 10.1561/0800000035.

the estimation of social preference parameters from dictator game data, and risky choice models with between-subject heterogeneity in risk aversion. The method maximum simulated likelihood (MSL) is promoted as the most suitable method for estimating models with continuous heterogeneity. We then consider 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. The application used as an example will be the level-k model, in which subject types are defined by their levels of reasoning. We then consider other models of behaviour in games, including the Quantal Response Equilibrium (QRE) Model. The final area covered is models of learning in games.

1

Introduction

Experimetrics[1] comprises the body of econometric techniques that are customised to experimental applications. This monograph is aimed principally at two types of reader. The first is the experimental economist who is interested in expanding their skill set in econometric techniques. In particular, experimental researchers who rely heavily on straightforward treatment testing techniques will hopefully be persuaded that more sophisticated econometric techniques are more suitable in many settings. The second type of target reader is the researcher from the wider econometrics community who may be interested to discover the sort of econometric techniques that are currently being used by Experimentalists. Of particular interest to this type of reader may be the techniques that have not traditionally been part of the standard Econometrics toolkit, for example, power analysis and optimal experimental design. The best possible end result from this monograph would be the cross-fertilisation of ideas from Applied of Theoretical Econometrics resulting in new and improved techniques in Experimetrics.

[1] The word "Experimetrics" was (to the best of my knowledge) coined by Camerer (2003, p. 42). The first article containing the word in the title was Bardsley and Moffatt (2007), and a textbook bearing the title *Experimetrics* was produced by Moffatt (2015).

The most widely used approach in the experimental economics literature is the treatment testing approach. The treatment test often simply amounts to a comparison of some behavioural outcome between two samples: a control sample and a treatment sample, drawn randomly from the population of experimental subjects. Justification of this approach is usually provided in terms of subjects being assigned randomly to treatments, and all influences on behaviour other than the treatment of interest being held fixed by virtue of the experimental environment. These arguments are often used to explain away many of the types of problem that econometricians have traditionally been interested in, such as sample selection bias, measurement error, and endogeneity.

However, there are a number of compelling reasons why the level of econometrics required in the analysis of experimental data goes beyond simple treatment testing. Firstly, in planning a treatment test, issues of experimental design are important. The most basic feature of the design is the sample size, and the use of power analysis is becoming routine in the process of setting this design feature. Power analysis has underpinnings in statistical theory, and conducting a thorough power analysis is facilitated by an understanding of this underlying theory. It must be said that this interest in the use of power analysis is relatively new in econometrics and has taken hold mainly as a consequence of the rise of Experimental Economics.

Second, there are many different ways of conducting the treatment test itself. A choice must be made over parametric and non-parametric tests, between-subject and within-subject tests, and so on. Which approach is best suited in a particular situation is often far from obvious. Once again power analysis is important. Testing approaches vary widely in terms of power. However, a theme of this monograph is that there are other considerations relating to human behaviour which sometimes distort theoretical prescriptions, and the test with the highest power is not always the most suitable.

Third, the structure of experimental data is frequently such that straightforward application of treatment tests is invalid. Data is often clustered, sometimes at more than one level. It is very common for each subject to engage in a sequence of experimental tasks, and hence there is clustering (or dependence) at the subject-level. In some types of experiment, subjects interact in groups, and it is natural to expect clustering at the group level. Finally, an experiment may be divided into sessions taking place in different places or at different times, and we may expect clustering at the session level. It is well-known (Moulton, 1986) that any combination of these different levels of clustering invalidates any test that assumes independence between observations. One way of dealing with these problems is to conduct the treatment tests as tests of significance in regressions with clustered standard errors, but a superior approach is to use a multilevel model that fully incorporates all levels of clustering.

Fourth, most treatment tests are performed on the assumption that if agents respond to a treatment, they all respond to it in the same way. That is, there is an implicit assumption of homogeneity. It may be that a proportion of subjects respond to the treatment in the expected way, but the remainder are not affected by the treatment. In this case, a test that assumes homogeneity will underestimate the treatment effect for those who respond to it. A more extreme possibility is that half of the population respond positively to the treatment while the other half respond negatively, and in this case the homogeneous treatment test may completely fail to detect any effect.

The importance of subject heterogeneity extends far beyond treatment effects. In many experiments, the focus is on "home-grown" characteristics of experimental subjects, such as risk aversion, ambiguity aversion, discount rate, inequity aversion, aversion to lying, and depth of reasoning. All of these characteristics are likely to exhibit wide variation over the population of subjects. Any model of any aspect of behaviour in which such characteristics are relevant must allow for this variation. Often, the most natural way to allow for such variation is to construct a fully structural model in which the distributional parameters of the varying characteristics are estimated along with the treatment effects of interest.

A good example of this is the modelling of social preference data. In some settings, individuals face repeated tasks in each of which they decide how to divide an endowment between their self and another player, with the endowment and price of transferring varying between tasks. This sort of data set provides an opportunity to estimate social preference parameters such as aversion to inequity and preference for efficiency – parameters which are almost certain to exhibit between-subject heterogeneity.

Another example is the modelling of risky choice data. Here, data on repeated choices between lottery pairs may be used to estimate preference parameters such as risk aversion and probability weighting. Again, these are parameters for which between-subject heterogeneity is expected. This sort of heterogeneity is allowed for in estimation using the method of maximum simulated likelihood.

Some sorts of heterogeneity may be referred to as *discrete heterogeneity*, meaning that, instead being characterised by continuously varying preference parameters, it takes the form of the population subjects dividing into discrete "subject-types", with discretely different models of behaviour. One example is behaviour in public goods games, in which the population might be assumed to divide between free-riders (who follow the Nash prediction by contributing zero), reciprocators (who are willing to contribute only if they see others contributing), selfish contributors (who contribute in anticipation of reciprocity by others), and altruists (who contribute regardless). Note that these four types are each defined by a different sub-model, and the econometric objective is to use the experimental data to estimate the parameters of these four models, along with a set of mixing proportions, which reveal the proportions of the population who are of each type. The econometric framework used for this purpose is the finite mixture model. A variety of methods are available for the estimation of finite mixture models, including machine learning techniques.

Another application of the finite mixture model which is used for illustrative purposes in this monograph is to depth of reasoning models. Here, the objective is to use data from behaviour in one-shot interactive games in order to estimate the proportion of the population who act at each level of reasoning.

Also of interest is behaviour in repeated games, where the initial questions to be addressed are how closely subject choices adhere to the Nash-equilibrium prediction, and whether sequences of choices appear random. More sophisticated analysis of repeated-game data reveals whether and how decision-makers *learn* to optimise their behaviour in repeated games. Are they attracted to strategies that have proved beneficial in previous tasks (reinforcement learning), or do they go further and form beliefs about the other players' actions and then make optimal decisions based on these beliefs (belief learning)?

The purpose of this monograph is to survey these techniques. However, it aims to be more than just a survey of the literature because it also aims to explain the techniques and evaluate them with illustrations, sometimes using data sets from previously published studies.

Section 2 provides an overview of treatment testing. A range of treatment tests are illustrated using the example of a dictator-game giving experiment in which there is a communication treatment. Standard parametric and non-parametric treatment tests, and also tests comparing entire distributions, are surveyed briefly. Then bootstrap tests are proposed as a way of avoiding the disadvantages of both parametric and non-parametric tests. Then it is explained how a treatment tests can be performed in a regression framework, as the test of significance of the effect of a dummy variable representing the treatment, one considerable advantage of this approach being that it becomes possible to correct the test for clustering of the data that inevitably arises at the subject-level and/or the session level. The final part of the section covers multi-level modelling, in which subject-specific and session-specific random effects are both incorporated in estimation.

Section 3 covers power analysis, an area of growing importance in experimental economics. The question of central interest is usually: what sample is required in a treatment test to attain a benchmark level of power (e.g., 0.8)? Tailor-made routines are outlined, but it is made clear that these methods are only applicable to a fairly limited range of treatment testing problems. The Monte Carlo method is proposed as a useful method for performing power calculations in situations in which ready-made routines are not available. The method is be applied to investigate questions including: how much more powerful is a within-subject test with n subjects than a between-subject test with $2n$ subjects? Then the method is applied to consider the problem of how to choose both n and T simultaneously in order to attain a required power in a panel data context.

Section 4 covers the different types of data arising in Experimental Economics, and the reasons why they arise in the form they do. Decision-making under risk is the chosen context in which these are considered. This choice of context is convenient because different data types arise depending on the elicitation method used: binary data arises when the subject chooses between lotteries; ordinal data arises when the subject makes a choice and also reports strength of preference; interval data arises when the subject is faced with a sequence of choice problems and decides on their switch-point, that is, where in the sequence they switch from the safe to the risky choice or vice versa; exact data arises when the subject has been asked to report a certainty equivalent of a single lottery.

The remainder of the monograph is concerned mainly with the estimation of fully structural models. Section 5 considers methods for the structural estimation of social preference parameters. Section 6 considers ways of dealing with continuous heterogeneity. The method of maximum simulated likelihood (MSL) is be used for all examples, and this method is explained in reasonable detail. Models with one dimension of heterogeneity, such as the Fechner and RP models of risky choice, in which only risk aversion varies between subjects, are be considered first. We then move on to settings in which there is more than one dimension of heterogeneity, for example: models in which subjects vary in both risk aversion and probability weighting; models in which subjects vary in risk aversion and time preference; models in which subjects vary in risk aversion and inequity aversion.

Section 7 considers 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. The application used as an example will be the level-k model, in which subjects are defined by their levels of reasoning, with the level of reasoning typically ranging from 0 up to 3 or 4. Section 8 considers other models of behaviour in games, including the Quantal Response Equilibrium (QRE) Model. Section 9 considers models of Learning. Section 10 concludes.

References

Allais, M. (1953). "Le Comportement de l 'Homme Rationnel devant le Risque: Critique des Postulats et axiomes de l'ecole Americaine". *Econometrica*. 21: 503–546.

Alós-Ferrer, C., Đ. Granić, J. Kern, and A. K. Wagner (2016). "Preference reversals: Time and again". *Journal of Risk and Uncertainty*. 52(1): 65–97.

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

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

Andreoni, J. (1995). "Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments". *The Quarterly Journal of Economics*. 110(1): 1–21.

Andreoni, J. and J. Miller (2002). "Giving according to GARP: An experimental test of the consistency of preferences for altruism". *Econometrica*. 70(2): 737–753.

Andreoni, J. and L. Vesterlund (2001). "Which is the fair sex? Gender differences in altruism". *The Quarterly Journal of Economics*. 116(1): 293–312.

Ansari, A., R. Montoya, and O. Netzer (2012). "Dynamic learning in behavioral games: A hidden Markov mixture of experts approach". *Quantitative Marketing and Economics*. 10(4): 475–503.

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

Arad, A. and A. Rubinstein (2012). "The 11–20 money request game: A level-k reasoning study". *American Economic Review*. 102(7): 3561–3573.

Arrow, K. J. and A. C. Enthoven (1961). "Quasi-concave programming". *Econometrica: Journal of the Econometric Society*. 29(4): 779–800.

Atkinson, A. C. (1996). "The usefulness of optimum experimental designs". *Journal of the Royal Statistical Society: Series B (Methodological)*. 58(1): 59–76.

Bacon, P. M., A. Conte, and P. G. Moffatt (2020). "A test of risk vulnerability in the wider population". *Theory and Decision*. 88(1): 37–50.

Bardsley, N. (2008). "Dictator game giving: Altruism or artefact?" *Experimental Economics*. 11(2): 122–133.

Bardsley, N., R. Cubitt, G. Loomes, P. Moffat, C. Starmer, and R. Sugden (2010). *Experimental Economics: Rethinking the Rules*. Princeton University Press.

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

Becker, G. M., M. H. DeGroot, and J. Marschak (1964). "Measuring utility by a single-response sequential method". *Systems Research and Behavioral Science*. 9(3): 226–232.

Ben-Ner, A., F. Kong, and L. Putterman (2004). "Share and share alike? Intelligence, socialization, personality, and gender-pairing as determinants of giving". *Journal of Economic Psychology*. 25(5): 581–589.

Berenson, M. L., D. Levine, and D. Rindskopf (1988). *Applied Statistics: A First Course*. New York: Prentice Hall.

Blavatskyy, P. R. (2011). "Probabilistic risk aversion with an arbitrary outcome set". *Economics Letters*. 112(1): 34–37.

Bolton, G. E. and A. Ockenfels (2000). "ERC: A theory of equity, reciprocity, and competition". *American Economic Review*. 90(1): 166–193.

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

Bosman, R. and F. Van Winden (2002). "Emotional hazard in a power-to-take experiment". *The Economic Journal*. 112(476): 147–169.

Bruhin, A., K. Janizzi, and C. Thöni (2020). "Uncovering the heterogeneity behind cross-cultural variation in antisocial punishment". *Journal of Economic Behavior and Organization*. 180: 291–308.

Butler, D., A. Isoni, G. Loomes, and D. Navarro-Martinez (2014). "On the measurement of strength of preference in units of money". *Economic Record*. 90(s1): 1–15.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Camerer, C. F., A. Dreber, E. Forsell, T. H. Ho, J. Huber, M. Johannesson, M. Kirchler, J. Almenberg, A. Altmejd, T. Chan, E. Heikensten, F. Holzmeister, T. Imai, S. Isaksson, G. Nave, T. Pfeiffer, M. Razen, and H. Wu (2016). "Evaluating replicability of laboratory experiments in economics". *Science*. 351(6280): 1433–1436.

Camerer, C. F., A. Dreber, and M. Johannesson (2019). "Replication and other practices for improving scientific quality in experimental economics". In: *Handbook of Research Methods and Applications in Experimental Economics*. Edward Elgar Publishing.

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

Camerer, C. F., T. H. Ho, and J. K. Chong (2004). "A cognitive hierarchy model of games". *The Quarterly Journal of Economics*. 119(3): 861–898.

Camerer, C., T. Ho, and K. Chong (2003). "Models of thinking, learning, and teaching in games". *The American Economic Review*. 93(2): 192–195.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). "Bootstrap-based improvements for inference with clustered errors". *The Review of Economics and Statistics*. 90(3): 414–427.

Cappelen, A. W., A. D. Hole, E. Ø. Sørensen, and B. Tungodden (2007). "The pluralism of fairness ideals: An experimental approach". *American Economic Review*. 97(3): 818–827.

Cherry, T. L., P. Frykblom, and J. F. Shogren (2002). "Hardnose the dictator". *American Economic Review*. 92(4): 1218–1221.

Cheung, Y. W. and D. Friedman (1998). "A comparison of learning and replicator dynamics using experimental data". *Journal of Economic Behavior and Organization*. 35(3): 263–280.

Chong, J. K., T. H. Ho, and C. Camerer (2016). "A generalized cognitive hierarchy model of games". *Games and Economic Behavior*. 99: 257–274.

Clarke, K. A. (2003). "Nonparametric model discrimination in international relations". *Journal of Conflict Resolution*. 47(1): 72–93.

Cohen, J. (2013). *Statistical Power Analysis for the Behavioural Sciences*. London: Routledge Academic.

Collins, L. M. and S. T. Lanza (2009). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Vol. 718. John Wiley & Sons.

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

Connolly, T. and D. Butler (2006). "Regret in economic and psychological theories of choice". *Journal of Behavioral Decision Making*. 19(2): 139–154.

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

Conte, A., M. V. Levati, and C. Nardi (2018). "Risk preferences and the role of emotions". *Economica*. 85(338): 305–328.

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

Conte, A., P. G. Moffatt, and M. Riddel (2019). "The multivariate random preference estimator for switching multiple price list data". School of Economics Discussion Paper 19-04, University of East Anglia.

Cragg, J. G. (1971). "Some statistical models for limited dependent variables with application to the demand for durable goods". *Econometrica: Journal of the Econometric Society*. 39(5): 829–844.

Crosetto, P. and A. Filippin (2016). "A theoretical and experimental appraisal of four risk elicitation methods". *Experimental Economics*. 19(3): 613–641.

Daniels, R. C. and S. Rospabé (2005). "Estimating an earnings function from coarsened data by an interval censored regression procedure". *Studies in Economics and Econometrics*. 29(1): 29–45.

Daykin, A. R. and P. G. Moffatt (2002). "Analyzing ordered responses: A review of the ordered probit model". *Understanding Statistics: Statistical Issues in Psychology Education, and the Social Sciences*. 1(3): 157–166.

Eckel, C. C. and P. J. Grossman (1998). "Are women less selfish than men?: Evidence from dictator experiments". *The Economic Journal*. 108(448): 726–735.

Efron, B. and R. J. Tibshirani (1994). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Engel, C. (2011). "Dictator games: A meta study". *Experimental Economics*. 14(4): 583–610.

El-Gamal, M. A. and D. M. Grether (1995). "Are people Bayesian? Uncovering behavioral strategies". *Journal of the American Statistical Association*. 90(432): 1137–1145.

Engel, C. (2020). "Estimating heterogeneous reactions to experimental treatments". *MPI Collective Goods Discussion Paper,* (2019/1).

Engel, C. and P. G. Moffatt (2012). "Estimation of the house money effect using hurdle models". *MPI Collective Goods Preprint*, (2012/13).

Engelmann, D. and M. Strobel (2004). "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments". *American Economic Review*. 94(4): 857–869.

Epps, T. W. and K. J. Singleton (1986). "An omnibus test for the two-sample problem using the empirical characteristic function". *Journal of Statistical Computation and Simulation*. 26(3–4): 177–203.

Erev, I. and A. E. Roth (1998). "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria". *American Economic Review*. 88(4): 848–881.

Fallucchi, F., R. A. Luccasen, and T. L. Turocy (2019). "Identifying discrete behavioural types: A re-analysis of public goods game contributions by hierarchical clustering". *Journal of the Economic Science Association*. 5(2): 238–254.

Fallucchi, F., A. Mercatanti, and J. Niederreiter (2020). "Identifying types in contest experiments". *International Journal of Game Theory,* forthcoming, published online: URL: https://doi.org/10.1007/s00182-020-00738-w.

Faul, F., E. Erdfelder, A. G. Lang, and A. Buchner (2007). "G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences". *Behavior Research Methods*. 39(2): 175–191.

Fechner, G. (1860). *Elements of Psychophysics*. Vol. 1. New York: Holt, Rinehart and Winston.

Fehr, E. and K. M. Schmidt (1999). "A theory of fairness, competition, and cooperation". *The Quarterly Journal of Economics*. 114(3): 817–868.

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

Fréchette, G. R. (2012). "Session-effects in the laboratory". *Experimental Economics*. 15(3): 485–498.

Gächter, S. and E. Renner (2010). "The effects of (incentivized) belief elicitation in public goods experiments". *Experimental Economics*. 13(3): 364–377.

Georg, S. J. (2009). "Nonparametric testing of distributions – the Epps–Singleton two-sample test using the empirical characteristic function". *The Stata Journal*. 9: 454–465.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019). "The german socio-economic panel (soep)". *Jahrbücher für Nationalökonomie und Statistik*. 239(2): 345–360.

Gollier, C. and J. W. Pratt (1996). "Risk vulnerability and the tempering effect of background risk". *Econometrica: Journal of the Econometric Society*: 1109–1123.

Gonzalez, R. and G. Wu (1999). "On the shape of the probability weighting function". *Cognitive Psychology*. 38(1): 129–166.

Gourieroux, C. and A. Monfort (1996). *Simulation-Based Econometric Methods*. Oxford University Press.

Greene, W. (2011). *Econometric Analysis*. Seventh edition. London: Pearson.

Greene, W. H. and D. A. Hensher (2010). *Modeling Ordered Choices: A Primer*. Cambridge University Press.

Grether, D. M. and C. R. Plott (1979). "Economic theory of choice and the preference reversal phenomenon". *The American Economic Review*. 69(4): 623–638.

Halton, J. H. (1960). "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals". *Numerische Mathematik*. 2(1): 84–90.

Hammersley, J. M. and D. C. Handscomb (1964). *Monte Carlo Methods*. Methuen.

Harrison, G. W., E. Johnson, M. M. McInnes, and E. E. Rutström (2005). "Risk aversion and incentive effects: Comment". *American Economic Review*. 95(3): 897–901.

Harrison, G. W. and E. E. Rutström (2009). "Expected utility theory and prospect theory: One wedding and a decent funeral". *Experimental Economics*. 12(2): 133.

Hausman, J. A. (1978). "Specification tests in econometrics". *Econometrica: Journal of the Econometric Society*. 46(6): 1251–1271.

Hey, J. D. and C. Orme (1994). "Investigating generalizations of expected utility theory using experimental data". *Econometrica: Journal of the Econometric Society*. 62(6): 1291–1326.

Holt, C. A. and S. Laury (2002). "Risk aversion and incentive effects". *American Economic Review*. 92: 1644–1655.

Holt, C. A. and S. K. Laury (2005). "Risk aversion and incentive effects: New data without order effects". *The American Economic Review*. 95(3): 902–904.

Huber, J. and K. Zwerina (1996). "The importance of utility balance in efficient choice designs". *Journal of Marketing Research*. 33(3): 307–317.

Ioannidis, J. P. A., T. D. Stanley, and H. Doucouliagos (2018). "The power of bias in economics research". *The Economic Journal*. 127: F236–265.

Isoni, A., G. Loomes, and R. Sugden (2011). "The willingness to pay—Willingness to accept gap, the 'endowment effect', subject misconceptions, and experimental procedures for eliciting valuations: Comment". *The American Economic Review*. 101(2): 991–1011.

Jakiela, P. (2013). "Equity vs. efficiency vs. self-interest: On the use of dictator games to measure distributional preferences". *Experimental Economics*. 16(2): 208–221.

Jiang, T., J. Potters, and Y. Funaki (2016). "Eye-tracking social preferences". *Journal of Behavioral Decision Making*. 29(2–3): 157–168.

Johnson, C., A. Baillon, Z. Li, D. van Dolder, and P. P. Wakker (2021). "Prince: An improved method for measuring incentivized preferences". *Journal of Risk and Uncertainty,* forthcoming.

Kahneman, D. and A. Tversky (1979). "Prospect theory: An analysis of decisions under risk". *Econometrica*. 47(2): 263–291.

Kanninen, B. J. (1993). "Optimal experimental design for double-bounded dichotomous choice contingent valuation". *Land Economics*. 69(2): 138–146.

Keasey, K. and P. Moon (1996). "Gambling with the house money in capital expenditure decisions: An experimental analysis". *Economics Letters*. 50(1): 105–110.

Keynes, J. M. (1936). *The General Theory of Employment Interest and Money*. Macmillan.

Kruschke, J. K. (2011). "Bayesian assessment of null values via parameter estimation and model comparison". *Perspectives on Psychological Science*. 6(3): 299–312.

Likert, R. (1932). "A technique for the measurement of attitudes". *Archives of Psychology*. 22(140): 1–55.

List, J. A. (2007). "On the interpretation of giving in dictator games". *Journal of Political Economy*. 115(3): 482–493.

List, J., S. Sadoff, and M. Wagner (2011). "So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design". *Experimental Economics*. 14: 439–457.

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

Loomes, G. and R. Sugden (1998). "Testing different stochastic specifications of risky choice". *Economica*. 65(260): 581–598.

Luce, R. D. (1959). "On the possible psychophysical laws". *Psychological Review*. 66(2): 81.

MacKinnon, J. G. (2002). "Bootstrap inference in econometrics". *Canadian Journal of Economics/Revue canadienne d'économique*. 35(4): 615–645.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

McKelvey, R. D. and T. R. Palfrey (1995). "Quantal response equilibria for normal form games". *Games and Economic Behavior*. 10(1): 6–38.

McLachlan, G. J. and D. Peel (2000). *Finite Mixture Models*. John Wiley & Sons.

Moffatt, P. G. (2005). "Stochastic choice and the allocation of cognitive effort". *Experimental Economics*. 8(4): 369–388.

Moffatt, P. G. (2007). "Optimal experimental design in models of decision and choice". In: *Measurement in Economics: A Handbook*. Ed. by M. Boumans. London: Academic Press. 357–375.

Moffatt, P. G. (2015). *Experimetrics: Econometrics for Experimental Economics*. Palgrave Macmillan.

Moffatt, P. G. (2020). "The experimetrics of depth of reasoning models". In: *Handbook of Experimental Game Theory*. Ed. by M. Capra, R. Croson, M. Rigdon, and T. Rosenblat. Edward Elgar.

Moffatt, P. G. and S. A. Peters (2001). "Testing for the presence of a tremble in economic experiments". *Experimental Economics*. 4(3): 221–228.

Moffatt, P. G., G. Pogrebna, and G. Zevallos-Porles (2020). *Depth of Reasoning Models with Sophisticated Agents*. Discussion Paper (No. 20-06). School of Economics, University of East Anglia, Norwich, UK.

Moffatt, P. G. and G. Zevallos-Porles (2020). *A Kuhn-Tucker Model for Behaviour in Dictator Games*. Discussion Paper (No. 20-03). School of Economics, University of East Anglia Norwich, UK.

Moulton, B. R. (1986). "Random group effects and the precision of regression estimates". *Journal of Econometrics*. 32(3): 385–397.

Nagel, R. (1995). "Unraveling in guessing games: An experimental study". *The American Economic Review*. 85(5): 1313–1326.

Nagel, R., C. Bühren, and B. Frank (2017). "Inspired and inspiring: Hervé Moulin and the discovery of the beauty contest game". *Mathematical Social Sciences*. 90: 191–207.

Nelson, F. D. (1976). "On a general computer algorithm for the analysis of models with limited dependent variables". *Annals of Economic and Social Measurement*. 5: 493–509.

Oehlert, G. W. (1992). "A note on the delta method". *The American Statistician*. 46(1): 27–29.

Penczynski, S. P. (2019). "Using machine learning for communication classification". *Experimental Economics*. 22(4): 1002–1029.

Pratt, J. (1964). "Risk aversion in the small and in the large". *Econometrica*. 32: 122–136.

Prelec, D. (1998). "The probability weighting function". *Econometrica*. 66(3): 497–527.

Quiggin, J. (1982). "A theory of anticipated utility". *Journal of Economic Behavior and Organization*. 3(4): 323–343.

Rabe-Hesketh, S. and A. Skrondal (2008). *Multilevel and Longitudinal Modeling Using Stata*. STATA Press.

Rosenthal, R. W., J. Shachat, and M. Walker (2003). "Hide and seek in Arizona". *International Journal of Game Theory*. 32(2): 273–293.

Roth, A. E., V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir (1991). "Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study". *The American Economic Review*. 81(5): 1068–1095.

Rubinstein, R. Y. and D. P. Kroese (2016). *Simulation and the Monte Carlo Method*. Third edition. John Wiley & Sons.

Runco, M. (2013). "Estimating depth of reasoning in a repeated guessing game with no feedback". *Experimental Economics*. 16(3): 402–413.

Saha, A. (1993). "Expo-power utility: A 'flexible' form for absolute and relative risk aversion". *American Journal of Agricultural Economics*. 75(4): 905–913.

Selten, R. and R. Stoecker (1986). "End behavior in sequences of finite Prisoner's Dilemma supergames. A learning theory approach". *Journal of Economic Behavior and Organization*. 7(1): 47–70.

Shachat, J., J. T. Swarthout, and L. Wei (2015). "A hidden Markov model for the detection of pure and mixed strategy play in games". *Econometric Theory*. 31(4): 729–752.

Siegel, S. and N. J. Castellan (1988). *Non-Parametric Statistics for the Behavioral Sciences*. Second edition. New York: McGraw Hill.

Smith, V. L. (1982). "Microeconomic systems as an experimental science". *The American Economic Review*. 72(5): 923–955.

Stahl, D. O. and P. W. Wilson (1995). "On players' models of other players: Theory and experimental evidence". *Games and Economic Behavior*. 10(1): 218–254.

Starmer, C. (2000). "Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk". *Journal of Economic Literature*. 38(2): 332–382.

StataCorp (2019). *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC.

Stewart, M. B. (1983). "On least squares estimation when the dependent variable is grouped". *The Review of Economic Studies*. 50(4): 737–753.

Stott, H. P. (2006). "Cumulative prospect theory's functional menagerie". *Journal of Risk and Uncertainty*. 32(2): 101–130.

Su, L., Z. Shi, and P. C. Phillips (2016). "Identifying latent structures in panel data". *Econometrica*. 84(6): 2215–2264.

Sun, T. and Y. Ding (2019). "Copula-based semiparametric regression method for bivariate data under general interval censoring". *Biostatistics*. Available online at: URL: https://doi.org/10.1093/biostatistics/kxz032.

Tanaka, T., C. F. Camerer, and Q. Nguyen (2010). "Risk and time preferences: Linking experimental and household survey data from Vietnam". *American Economic Review*. 100: 557–571.

Thaler, R. H. and E. J. Johnson (1990). "Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice". *Management Science*. 36(6): 643–660.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Tversky, A. and C. R. Fox (1995). "Weighing risk and uncertainty". *Psychological Review*. 102: 269–283.

Tversky, A. and D. Kahneman (1992). "Advances in prospect theory: Cumulative representation of uncertainty". *Journal of Risk and uncertainty*. 5(4): 297–323.

Tversky, A., P. Slovic, and D. Kahneman (1990). "The causes of preference reversal". *The American Economic Review*. 80(1): 204–217.

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

Vuong, Q. H. (1989). "Likelihood ratio tests for model selection and non-nested hypotheses". *Econometrica: Journal of the Econometric Society*. 57(2): 307–333.

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

Wilcox, N. T. (2008). "Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison". In: *Risk Aversion in Experiments*. Emerald Group Publishing Limited. 197–292.

Wilcox, N. T. (2011). "'Stochastically more risk averse:' A contextual theory of stochastic discrete choice under risk". *Journal of Econometrics*. 162(1): 89–104.

Yates, F. (1934). "Contingency tables involving small numbers and the χ^2 test". *Supplement to the Journal of the Royal Statistical Society*. 1(2): 217–235.

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