Summary

This book is an in-depth study of the statistical properties of the Propensity Score Matching estimators method used in estimation of the average effects of impact. Asymptotic properties of estimators have been widely recognised by the researchers but there is a lack of research on their small sample properties. This book will try to fill this gap. Quantitative effects of impact, also called treatment effects, may be assessed using various empirical methods. The propensity score itself is a probability of receiving treatment. The two most commonly used measures of treatment impact are the average treatment effect and the average treatment effect on the treated. The first shows what is the effect of the programme on the average individual in the population. The second quantifies the effect of the programme impact on individuals that are programme participants. The statistical properties of the propensity score matching estimators are analysed for both estimators of the programme effect. Two methods of estimation are considered. The aforementioned methods are matching on the value of estimated propensity score and the inverse probability weighting, in which reciprocal of the estimated propensity score is used as a weight for each observation.

Some difficulties arise in front of the author from the fact that there is a lack of well-established terminology on matching methods in Polish. The only known academic publication in which matching methods are discussed is a book *Mikroekonometria* edited by Marek Gruszczyński. A few years ago, before the first edition of this handbook was printed, we had tried to establish translations of terminology. In general, it works; however, minor differences between the authors remained. The separate

comment is needed to the book title. The first part of the title is in English, because there is no well-established translation of it into Polish. The translation that I use tries to capture the essence of the method; however, even for microeconometricians it could be incomprehensible. The second part is in Polish and simply means small sample properties. The concept of a small sample is not the precisely defined statistical literature. The literature defines the concept of a large sample, meaning that the sample size is sufficient to conduct statistical inference using the asymptotic distribution of estimators. Analogously, a small sample can be defined as the sample size, which is not sufficient for the estimator to achieve its limiting probability distribution.

The book consists of three chapters. The purpose of the first chapter is to introduce the reader to the world of treatment effects and data matching. It has theoretical character and reviews the most important works in literature regarding the analysis of counterfactuals and matching. The foundations of counterfactual analysis are derived from the works of Neyman (1923) and Fisher (1925) on randomisation. Two basic estimators for treatment effects are described and their identification conditions. Based on the fundamental work of Rosenbaum and Rubin (1983), the matching method is discussed and their proposal to use the propensity score to join observation. The hereinafter chapter presents the assumptions and the statistical properties of the Propensity Score Matching method in large samples. A short fragment of this part is based on my doctoral dissertation. The added value of this chapter is to collect in one place the most important results of the literature from the past 15 years. The chapter closes with the presentation of various algorithms used to combine data, and tools used to assess the quality of the resulting match. Various variants of the nearest neighbour matching algorithm are discussed. The nearest neighbour matching algorithm has the advantage of speed and the disadvantage of ambiguity. The inverse probability weighting algorithm performance relies on the quality of the model used to estimate the probability. The chapter ends with presentation of scientific discussion that took place at the beginning of the twenty-first century and is related to issues if matching methods are capable of replication of fully controlled experiment results.

The second chapter presents simulations results of the properties of the matching method estimators in small samples. The properties of statistical methods in small samples are essential for the practical application of econometric tools. In specific applications, such as detecting an intervention effect on the labour market, the researcher usually faces the challenge of quantitative assessment of the impact based on a pilot survey. A characteristic attribute of a pilot survey is its small scale and, therefore, a small sample size. So, it is worth knowing if the properties of statistical matching methods that are proved by using the limit theorems are also retained in the samples with not so many observations. There is a lack in the literature of general work on the Propensity Score Matching estimator's properties in small samples. In this chapter the results of original simulation studies are presented. The structure of the experimental plan of simulations is similar to my previous works. The artificial datasets are designed to mimic the properties and behaviour of the real data. The experimental plan itself and a part of the presented results of the estimator's properties have already been published in my articles. The distinguishing feature of the presented results from the previous results is the purpose of the study and its much wider range. Two methods for treatment effect estimation were analysed along with the impact of different techniques that enforce the assumptions of Propensity Score Matching methods. Standard measures of quality are provided: the bias, the variance and the root mean squared error of the estimator.

The results of simulations allow to derive some important conclusions. First, one effective method of the propensity score estimation does not exist. Different estimation techniques for the average treatment effect estimation in small samples provide similar results. Only in the case of simulation, where the random error is asymmetric techniques that enforce the common support provide precise results. The best point estimates in terms of bias are obtained by propensity score matching with trimming. The parameter value suggested for trimming by Crump et al. (2009) improves the statistical properties of estimators. Second, an increase in the sample size does not necessary lead to better estimates, in term of bias and variance, for the average treatment effects. Third, for the average treatment effect on the treated, the estimation results are diversified. However, matching on the propensity score with caliper of reasonably small size improves precision of the estimates. It is also worth emphasising that the inverse probability weighting is the worst performing estimation technique in term of the root mean squared error. The novelty and additional value added of the study are the size and power considerations. The results indicated that there is no problem with the size of the propensity score matching based estimators. In simulations with symmetric distribution for error there are no problems with the power of the test procedures. They exist in simulation with asymmetric error component. In these simulations, the actual size of the test may slightly deviate from the one expected. A more serious problem is the low power of the test procedure.

The third chapter is an attempt to present applications and statistical properties of these methods used for the analysis of real datasets. There are four examples included. The first one is originally designed for this book, showing how one could use the available data derived from the Labour Force Survey (LFS) to evaluate the effectiveness of additional on-work training and their impact on the situation of the participating employees. The remaining examples are adapted from the literature. The second example is based on the article, of which I am a co-author. In this example the relative benefits for choosing the diversification strategy or the specialisation strategy for rural households are compared. The third example comes from my own article. The data from American subsidised jobs are analysed. The aim of the study is to estimate the gross effect of the programme, i.e. the effect of beneficiary without regard to the cost of the programme. The last example was adapted from an article by Vanderberghe and Robin (2004). The aim of the article is to compare the effect of placing children in private schools on their educational outcomes. In each example of attempts to estimate the average treatment effect (ATE) and the average treatment effect on the treated (ATT) the Propensity Score Matching method and the inverse probability weighting on the propensity score were used. Two important results are highlighted. First, the use of matching methods contributes to balancing distribution's characteristics in the experimental and control group. Secondly, there is no universal and effective method allowing the use of propensity score vectors in small samples.