The Application of Quasi-Experimental Methods in the Context of Marketing: A Literature Review

Ioannis Petrakis

The effects of marketing investments on business outcomes remain ambiguous because most of the empirical studies in the marketing literature cannot postulate causality. The two problems of causal inference are those of endogeneity and unobservability. This paper reviews 26 applications of four common techniques - Instrumental Variables (IV), Regression Discontinuity (RD), Differences-in-Differences (DID) and Propensity Score Matching (PSM) - which resolve the problems of endogeneity and unobservability by exploiting observational data in a quasi-experimental setting where causal effects can be obtained.

Wageningen University

Marketing and Consumer Behavior Group

Minor Thesis

Table of Contents

1. Introduction

1.1 The Ambiguity of Examining the Effects of Marketing Variables on Business Outcomes

Despite the fact that successful marketing is still being considered crucial nowadays, the impact of marketing investments on firm performance remains ambiguous (Srinivasan and Hanssens, 2009). Based on theoretical propositions companies spend vast amounts of resources in marketing activities to increase their performance. These propositions relate to business outcomes (e.g. sales) with marketing variables (e.g. promotion) through empirical models. Although many studies in marketing literature have tested the relationship between marketing variables and business outcomes, there are a few studies which can postulate causality.

According to Rubin (1974), the definition of a causal effect is the difference between two potential outcomes. The outcome of an individual when experiencing a treatment, such as promotion, minus the counterfactual outcome of the same individual if had not experienced that treatment. With respect to marketing, causality can provide information of what would have happened in the purchases of a group of people (treatment group) who were exposed to a promotion, if they were not exposed to that promotion. However, causal inference suffers from the evaluation problem of unobservability in potential outcomes. The source of unobservability is that only the outcome of the individual being treated can be observed (Angrist and Pischke, 2008).

In addition, most of the empirical models suffer from the problem of endogeneity in explanatory variables because researchers use observational data to conduct their analysis (Rossi, 2014). Endogeneity refers to unobserved variables which may co-vary with the treatment, but they are not included in the model. The effects of those kinds of variables may change the outcomes of the treatment group, despite the presence of the treatment. For example, individuals who have increased their purchase rate about a product after being exposed to a promotion might have done so just because of their high interest for the promoted product and not because of the promotion per se. Another example might be that during the time a food company was implementing a promotion for one of its biological products in a specific country, the government of that country was running a national educational campaign for supporting a more sustainable way of life. Even if the effect of promotion on sales appeared to be statistically significant, this effect can be invalidated by an alternative model which controls for the time of the government campaign. Endogeneity leads to biased and inconsistent estimates, making causal inference inappropriate. In turn, biased and inconsistent estimates are translated into sub-optimal, if not wrong, allocation of resources for marketing investments.

This paper reviews 26 applications within the marketing context of four common techniques - Instrumental Variables (IV), Regression Discontinuity (RD), Differences-in-Differences (DID) and Propensity Score Matching (PSM) - which resolve the problems of endogeneity and unobservability by exploiting observational data in a quasi-experimental setting where causal effects can be obtained.

1.2 Endogeneity in Econometric Models within the Marketing Field

Much of the empirical research in marketing literature is implemented via econometric models which are used to predict changes in business outcomes such as demand, customer retention, sales, price, and conversation rates. Econometric models aim to predict these changes through the estimation of certain parameters. The word 'parameter' refers to the main effect obtained by the model for a specific variable. The interpretation of an estimated parameter is used to give crucial insights for the proper allocation of resources and thus optimization of the marketing mix according to the statistical significance, sign and magnitude of that parameter.

The parameters which are used to predict business outcomes are divided across three groups of explanatory variables. The first group represents marketing treatment variables such as price, advertising, promotion or keyword position and thus the major group of interest. The second group represents additional variables of observed sample-specific characteristics like demographics, market characteristics, or time which are included as controls. Control variables are important because they control for any changes in the outcome despite the presence of the treatment. The four quasi-experimental methods which are discussed in this paper use these kinds of controls in order to estimate causal effects. But, in a more sophisticated way than just adding variables in the model.

The third group represents the variables which are not observed by the researcher such as unobserved sample-specific characteristics or unobserved demand shocks, and thus are contained in the error term. The unobserved sample-specific characteristics are these variables which are difficult to observe and/or to quantify such as product quality, style, prestige, past experience or customer information accessed only by the manager through a loyalty program database. The unobserved demand shocks are related to contemporaneous consumer trends, word of mouth, advertising or promotional shocks. Those variables may be perceived by managers but still not observed directly by the researcher (Rossi, 2014). Empirically speaking, endogeneity occurs when the marketing variables are correlated with the error term. The problem of endogeneity has three sources of bias.

According to Rossi (2014), endogeneity is mainly caused by omitted variables bias (OVB). OVB is referred to unobserved variables which are not included in the model, but they could give an alternative or additional interpretation for the relationship of interest if they were included. These kinds of variables would affect the outcome even in the absence of the treatment. Consider a model which examines the effect of the intensity of advertising on profit. The intensity of advertising is usually determined by advertising costs. If the researcher omits the variable from the model which is related to advertising cost, then the effect of advertising on profit will be biased because the omitted variable would be part of the error term.

Sometimes OVB is related to variables that affect the participation of individuals in to treatment. This is an alternative form of OVB and is called self-selection bias (SSB). Consumers may self-select themselves in receiving the marketing treatment due to their high interest for a specific product or loyalty for a specific brand. Another case related to self-selection issues is when the marketing manager targets a specific group of consumers according to information, like past purchases, processed from a loyalty program database (Busse et al., 2006). Therefore, the estimated effect of promotion on an outcome, such as sales, is biased since those consumers were intentionally selected by the manager to be treated. Self-selection bias is usually present when the marketing variable is related to targeting.

Endogeneity may also be caused by simultaneity bias (SB). SB becomes apparent when the outcome and the marketing variable are jointly determined. Consider, for instance, a demand and a supply model, which simultaneously determine price and quantity via an equilibrating mechanism. Even if omitted variables are observed and controlled by the researcher in both models, equilibrium prices and quantities would still be endogenous because are simultaneously determined within a system of equations (Verbeek, 2008).

Finally, endogeneity may be caused by measurement error (ME) in the outcome or the marketing variable. ME appears when proxies are used for an observed variable, or when it is difficult to measure this variable at all. For example, when the values of an observed variable are dominated by zeros this leads to ME. (Roberts & Whited, 2012). The problem of ME in a variable is that this variable contains a large amount of unexplained variation which, in turn, will considerably increase the value of the error term. So, it is expected that the explanatory variables in a model which suffers from ME will be endogenous. Although they are not the same, ME and SB are closely related.

1.3 Randomization and Quasi-Experimental Experimental Econometrics

The remedy for the evaluation problem of unobservability is to use the outcomes of a control group, that is a group of non-treated individuals, as the counterfactual outcomes for the treatment group. The use of the outcomes of a control group as counterfactuals is dictated by a very strong assumption - the treatment is randomly assigned across the two groups. Randomization, not only is a prerequisite for the counterfactual outcomes, but also resolves endogeneity concerns by ensuring that that the effect of the marketing variable on the outcome is not harmed by OVB/SSB, SB, or ME. Consequently, randomization of the treatment assignment yields unbiased, or at least consistent, estimates.

The root of the problems related to unobservability and endogeneity can be found in the estimation of the average effects obtained for treatment and control groups. An introduction to some of the various kinds of average effects can be helpful at this point. The Average Treatment Effect (ATE) is the expected treatment effect of an observation that is randomly picked out from the population which is represented by the sample. The Average Treatment Effect of the Treated (ATT) and the Average Treatment Effect of the Untreated (ATU) are the expected treatment effects of observations that are randomly picked out from the sub-populations of treatment and control groups, accordingly. Most of the empirical studies focus on the ATE and ATT, and specifically on the latter(Wooldridge, 2010). Yet, there is also another average effect which has received considerable attention in empirical research, the Local Average Treatment Effect (LATE). However, this quantity will be explained later.

By simply comparing ATT and ATU one cannot identify the causal effect of treatment, because the estimate of an ATT is probably confounded by unobserved factors. In essence, there may be unobserved sample-specific characteristics or demand shocks that make the treatment and control group different by creating an SSB case, and those unobserved variables would affect the outcome even in the absence of the treatment. The solution to this problem is random assignment. If the assignment is randomized, then is also independent of the potential outcomes. Thus, any kind of bias which can be captured by the error term becomes zero. Randomization makes the comparison between the treatment and control groups independent from any selection issues, so any observable (or unobservable) differences between the two groups are insignificant. Then, one may identify the causal effect of treatment by comparing the average outcomes of treatment and control groups (Roberts & Whited, 2012).

Complete randomization refers to experimental trials. Experimental research has been the cornerstone of studying causal relationships because of the use of sophisticated research designs which ensure a random selection of the treatment group which, in turn, leads to an unbiased estimate of the ATT. However, experimental research is constrained by issues which are mainly related to time, money, ethics and feasibility. For instance, marketing experimental trials are difficult to implement due to the risk of losing customers. The demanding setting of experiments has led marketing researchers to opt for quasi-experimental econometric techniques that approximate randomization under certain conditions (Angrist and Pischke, 2015).

Although econometric models cannot substitute randomization, they can be formulated in a way to rule out, at least the obvious sources of endogeneity. Quasi-experimental econometrics has been introduced in order to mimic random assignment either by controlling for the variation of the treatment variable or by controlling for the assignment mechanism itself (Meyer, 1995). During the last decade, quasi-experimental econometrics has successfully been used by marketing researchers as a tool for testing causal relationships between marketing variables and business outcomes.

Quasi-experimental research refers to natural experiments which provide marketing researchers with more opportunities for identifying causal effects. These experiments are not 'real' experiments, and that is why they are called 'quasi' experiments (Meyer, 1995). Such settings are not intentionally set up by the researcher and so the treatment group is not randomly assigned. The modeling of quasi-experiments exploits natural events such as time or arbitrary rules of thumb to control for the selection process of individuals into treatment or the variation of the treatment variable. Quasi-experiments are more similar to observational studies where the researcher cannot change the environment. In these research designs, control and treatment groups may still differ despite the presence of the treatment. So, the researcher has to take certain steps in order to control for any difference between the two groups.

According to Heckman (2000), the challenge of using quasi-experimental econometric models is related to the identification and interpretation of their parameters. The trivial issue with identification process is that many alternative models may be applied to the same data. Each model, however, has its own causal interpretation which is heavily dependent upon certain assumptions. Therefore, the validity of an estimated effect obtained by a particular model should reflect upon the assumptions required to identify the causal effect given the structure of the data.

1.4 Quasi-Experimental Experimental Econometric Tools and Their Assumptions

The most popular quasi-experimental econometric tools are Differences-in-Differences (DID), Regression Discontinuity (RD), Instrumental Variables (IV), and the Propensity Score Matching (PSM). All of these tools provide researchers with the opportunity to tackle endogeneity, and therefore consistently estimate causal effects even in the absence of complete randomization. However, each of them is accompanied by certain assumptions. If these assumptions are not satisfied, the estimated effect would still be biased and inconsistent.

IV models are the standard textbook solution for endogeneity (Larker and Rusticus, 2010; Rossi, 2014) and used to control for the variation of the endogenous variable. The instrument is a variable which shifts the variation of the endogenous variable. The conditions of relevance and exclusion restriction should be satisfied for attaining consistent estimates of the effect under examination. Relevance, necessitates that the instrument and the endogenous variable are partially correlated. Exclusion restriction, necessitates that the instrument is not correlated with the error term and therefore is exogenous. The combination of relevance and exclusion restriction implies that the instrument should affect the outcome only through the endogenous variable. If these two conditions are met then IV approximate randomization (Heckman, 1997). The next three methods relate to the controlling of the assignment mechanism, and therefore expand on the notion of the counterfactual.

RD models use arbitrary rules of thump or heuristics to control for the selection of observations into treatment. Such designs are characterized by three ingredients - forcing variable, threshold and the treatment. Specifically, the observations receive a score, and the treatment is assigned to those observations whose score is above a predetermined threshold. The variable which describes the score of observations is called forcing variable, because 'forces' observations in to treatment. The observations whose score is below the threshold represent the control group. RD can be interpreted as local randomization, and the estimated effect under this setting is called LATE. The word 'local' stands for the differences between observations who are just above and just below the threshold. RD has two main conditions. The first condition is that the differences of observations whose score is around the threshold should be random, otherwise the LATE is confounded. The rationale of this condition is that observations which are just below the threshold must be, on average, exactly like those which are just above of it. So, any difference in the average outcomes between the treatment and control groups can be attributed to the treatment. An alternative interpretation of RD is that the effect of the treatment on the outcome is smooth as long as observations are below the threshold. Then, any discontinuity in the outcome must be caused by treatment

(Imbens and Lemieux, 2008). The second condition of RD is that the threshold is determined independently of the forcing variable, therefore is exogenous. This means that observations cannot affect their assignment into treatment since they are not aware of the threshold (Jacob and Zhu, 2012).

DID models utilize the exogeneity of external shocks to control for the selection of observations into treatment. Such designs are characterized by three ingredients- time, fixed effects, and the treatment. The treatment is an external shock, a change in the environment, which is 'naturally' applied to a specific group of observations in a given period. The control group is then characterized by observations who do not receive the treatment in that period. In such settings, the treatment assignment is considered to be random, after getting rid of all the fixed effects. The term 'fixed' refers to unobserved effects that are constant through time and may still affect the outcome despite the presence of the treatment. The necessary condition for DID is that of common trends. Specifically, the outcomes of the two groups must exhibit common trends during the pre-treatment period. This condition implies that in the absence of treatment, the average change in the outcomes would have been the same for both treatment and control groups. Then any change in the outcomes of the treatment group can be attributed to the treatment (Ashenfelter, 1978; Roberts & Whited, 2012).

PSM procedures are used to create a matched sample among the treatment and control groups. This technique is characterized by four ingredients - propensity score, matching algorithms, treatment, and a set of observed covariates. Such procedures start with the estimation of the propensity score (Rosenbaum and Rubin, 1983), which is defined as the probability of an observational unit in receiving the treatment based on certain covariates. Then, a matching algorithm is used to throw out of the sample the control units which are different to the treated ones (Rubin, 2006), based on their propensities. The control units which are not excluded can be used as valid counterfactuals. By comparing how outcomes vary for treated units with respect to observationally similar non-treated ones, one can identify the effect of the treatment. PSM is based on the assumption of conditional independence. According to this assumption, there is a set of observed covariates, and after one controls for them, the potential outcomes are independent of the treatment assignment. In other words, after controlling for these covariates, the treatment assignment is "as good as random" (Heinrich et al., 2010).

1.5 Motivation, Aim and Research Questions

There is a need in the marketing field for more quasi-experimental econometric applications to assess the effects of marketing variables on business outcomes within a robust frame of inference. The emergence of database marketing has made managers capable of customizing pricing and advertising among other components of the marketing mix, with respect to the type of consumers they want to satisfy (Rossi et al., 1996). Moreover, according to Rossi (2014), the marketing field has plenty of high quality data with a considerable variety of marketing variables for which effects on outcomes are crucial. However, most of this data is gathered by observational methods and hence the empirical models which are used to evaluate marketing effects suffer from the various sources of endogeneity.

This paper is about how marketing has been using advanced tools to identify causality, and the need for marketing researchers to become familiar with such tools. On the one hand, it might be difficult for the researcher to find a good instrument that satisfies the conditions of relevance and exclusion restriction. On the other hand, it might be easier for the researcher to find an external shock or an arbitrary selection rule for applying a DID or an RD model, accordingly. The estimation of causal effects in marketing is, therefore, trivial and requires knowledge on various methodologies. The motivation of this thesis is to help marketing researchers in acquiring knowledge about the use of quasi-experimental econometric tools within the context of marketing.

This thesis reviews 26 empirical marketing studies in a wide variety of marketing areas. This literature retrieval analyzes the most common endogenous marketing variables, the efforts that have been done to identify an appropriate model to rule out endogeneity and the necessary assumptions for a causal interpretation with respect to the structure of the data and marketing theory.

Judging from the above discussion regarding the ambiguity of examining causal effects due to unobservability, endogeneity and the solutions to these issues which are offered by experimental or quasi-experimental designs, the aim of this thesis is to bring these concepts into the field of marketing by answering the following research questions: By applying quasi-experimental econometric tools to estimate causal effects in the context of marketing:

1) What kinds of problems have been solved?
2) Why the researcher should be aware of all these methods?
3) How these methods help to identify the problem, and how the problem in turn helps to identify the correct method?

1.6 Outline

The structure of the remaining Sections is unfolded within the following sequence: In Section 2 applications of the four quasi-experimental methods are reviewed in terms of data structure, marketing theory, causal mechanism, main variables, endogeneity concerns and the assumptions to be fulfilled according to the applied method. In Section 3 the discussion regards the answers of the research questions supported by the information given in Section 2, but also refers to the contributions and the limitations of the research conducted for this thesis. Section 4 provides conclusions and recommendations for further research. Finally, an appendix is provided which discusses more technical issues and gives practical examples.

2. Quasi-Experimental Experimental Econometrics in Marketing Research

2.1 Instrumental Variables

According to Rossi (2014), IV methods " *do not use all of the variation in the data to identify causal effects, but instead partition the variation into that which can be regarded as clean or as though generated via experimental methods, and that which is contaminated and could result in endogeneity bias*".

IV were initially introduced by Wright (1928) to correct for simultaneity bias. Today IV are mostly employed to correct for omitted variables bias (Angrist and Pischke, 2008). In both cases though, instruments play the same role and this effectively simplifies the explanation to a single-equation model (Larcker and Rusticus, 2010). The word 'equation' usually refers to linear regression models which estimate the effect of an explanatory variable on the outcome of interest using ordinary least squares (OLS). If the correlation between the explanatory variable and the error term is zero, then the OLS estimate is consistent. That is, the estimate asymptotically approximates the true value of the parameter under examination. The rationale of asymptotic theory is that as the sample size approaches infinity the mean estimate of the model converges to the population mean (Rossi, 2014). However, if the explanatory variable is correlated with the error term via other unobservables, then one needs an IV estimator to achieve consistency. IV improves over OLS in the sense that the effect is still biased but is consistent, while under endogeneity the effect would not only be biased but also inconsistent. Notably, all of the quasi-experimental methods discussed in this paper provide biased but consistent estimates and thus improve over OLS when endogeneity is apparent.

IV estimators are consistent only if the instrument is valid. The only way to detect a valid instrument is to realize the theory which describes the relationship between the outcome, the endogenous variable and the instrument. The question that the researcher should ask with respect to a candidate instrument is, "Is the instrument related to the outcome only through its relationship with the endogenous variable?" If the answer is no, then instrument is not valid. For example, considering once again the model which examines the effect of advertising intensity on demand. For the exclusion restriction condition, the instrument should not be correlated with the demand error. For the relevance condition, the instrument should be partially correlated with advertising intensity. In relation to the aforementioned variables and with respect to marketing theory, advertising cost can be a valid instrument. The relevance of the instrument can be empirically checked by regressing the endogenous variable on the instrument, and this is usually called the fist-stage equation (appendix) . On the contrary, the exclusion restriction cannot be empirically checked and therefore mandates the researcher to provide strong theoretical arguments based on the literature and opinion of experts.

Table 1 provides an overview of the 8 articles which are reviewed in this Section. Specifically, Table 1 describes the name of the paper, academic journal, type of data, outcomes, potential endogenous marketing variables, and proposed instrumental variables which are analyzed by the authors of the reviewed papers. The criteria used regarding the selection of papers for this Section, but also for the following ones, were the novelty of quasi-experimental design, prestige of the journal, recentness of the paper, and space given to the text regarding the discussion on endogeneity. Particulary, the papers were selected after hitting the name of the method and endogeneity-for example: “IV” and “endogeneity”-in google scholar and the websites of the cited journals, as well as discussion on assumptions at least to some extent. The papers finally were selected in terms of their application on various marketing areas, so the reader can have a broader view of these applications in the general context of marketing.

Table1: IV Applications in Marketing Research

Authors	Name of the Paper	Journal	Data	Dependent Variable(s)	Potential Endogenous Variable(s)	Proposed Instrumental Variable(s)
Villas—Boas and Winner (1999)	Endogeneity in Brand Choice Models	Management Science	Household Level Panel Data	Demand	Price	lagged prices, lagged market sha cost variables
Kuksov and Villas-Boas (2008)	Endogeneity and Individual Consumer Choice	Journal of Marketing Research	Household Level Panel Data	Demand	Price, Display, Feature	cost variables with additional lags
Chintagunta and Dube (2005)	Estimating a Stockkeeping-Unit-Level Brand Choice Model That Combines Household Panel Data and Store Data	Journal of Marketing Research	Household and Store Level Panel Data	Demand	Price	wholesale prices
Barroso and LLobet (2012)	Advertising and Consumer Awareness of New, Differentiated Products	Journal of Marketing Research	Household Level Panel Data	Demand Price, Advertising	Price, Advertising	cost functions and price differences with respect to their individual time means for price , product characteristics for advertising
Petrin and Train (2010)	A Control Function Approach to Endogeneity in Consumer Choice Models	Journal of Marketing Research	Household Level Data	Demand	Price	control function and Hausman Type instruments
Danaher and Dagger (2013)	Comparing the Relative Effectiveness of Advertising Channels : A Case Study of a Multimedia Blitz Campaign	Journal of Marketing Research	Individual Level Panel Data	Sales, Profit	Advertising	number of mail and e-mail contacts before the main campaign, previous purchase incidence, sum of visits to rival websites, membership period of a person in the loyalty program, number of times Google was employed for the investigation of rival firms in the previous month
Dinner et al. (2014)	Driving Online and Offline Sales: The Cross-Channel Effects of Digital Versus Traditional Advertising	Journal of Marketing Research	Individual Level Panel Data	Search Impressions, Click Through Rate, Online Sales, Offline Sales	Advertising	advertising expenses of cheap retailers as instruments for advertising levels of the expensive retailer of interest or weekly unit costs for television, magazine, newspapers, online display for the whole retail industry, and quarterly manufacturer price indexes for periodicals and newspapers
Rutz et al. (2012)	A latent Instrumental variables approach to modeling keyword conversion in paid search advertising	Journal of Marketing Research	Keyword Level Data	Click-Through Rate, Conversion Rate	Keyword Position	latent variables or lagged position and lagged cost per click

Villas—Boas and Winner (1999) show that when one wants to predict demand via price and does not account for endogeneity, the estimates will be hampered by unobserved factors. The authors use scanner panel household data in order to predict demand for yogurt and ketchup products. The concern about endogeneity arises because price is determined by the cost of unobserved product characteristics such as input prices and unobserved demand shocks such as forward buying and stock piling.

Primarily, Villas—Boas and Winner (1999) use lagged prices and lagged market shares as instruments to tackle endogeneity. The argument for the selection of lagged variables is that they are easily available to the researcher. Rossi (2014) states though, that this is not a legitimate argument regarding the exclusion restriction due to the fact that forward buying and stock piling lagged variables would also be endogenous. Although these shocks are quite rare and would not represent a substantial part of the error term, may still be correlated through time (referring to autocorrelation). In that sense, lagged variables would be correlated with the current period shock. Villas—Boas and Winner (1999), employ cost variables as alternative from lags. Specifically, prices of milk are used for the yogurt products and prices of tomatoes are used for the ketchup products. The argument for using costs variables instead of lags is that the former is possibly more independent from demand shocks, at least for the current period. The results are improved via the use of cost instruments, which gives the hint that lagged variables may not be the optimal choice despite their availability.

Kuksov and Villas-Boas (2008) study the demand in ketchup products using scanner panel data and check for endogeneity in price, display and feature. Cost instruments similar to those of Villas—Boas and Winner (1999) are employed, but with additional lags. The authors report that only price is endogenous as compared to the other two marketing variables, and that lags do not explain price at all. The results are corroborated by an alternative model in which labor rates and energy prices are included as instruments.

Chintagunta and Dube (2005) consider unobserved brand characteristics a serious reason for price endogeneity in demand models. The authors combine household and store panel data to estimate the effect of price on demand for fabric softeners. In order to overcome endogeneity, Chintagunta and Dube (2005) use wholesale prices as instruments. Wholesale prices are expected to be correlated with retail prices, but are not expected to be correlated with in-store coupons and other retail-specific unobserved characteristics. Rossi (2014) though, provides an argument against the employment of wholesale prices as good instruments, because this is similar to the estimation of long-run price effects and not to solving endogeneity. Another interpretation of this argument is that using wholesale prices to instrument retail prices is similar to the projection of a highly variable price series on a much less variable series.

Barroso and LLobet (2012) assess the dynamic effect of advertising expenditures on consumers' product inclusion in their choice sets. The authors use household panel data to estimate demand for automobiles where the variables of demand, price and advertising are determined in an equilibrium, leading to simultaneity bias. Regarding the equations of price and demand, two instruments are employed. The first instrument represents cost functions of observed product characteristics. The argument is that the price of a specific product is correlated with the characteristics of all the products belonging to the same firm, but also with the characteristics of similar products in competing firms. However, product characteristics do not change frequently and may not be able to capture the relevant variation of the specific data. So, Barroso and LLobet (2012) employ the second instrument, which is composed of price differences with respect to their individual time means, with lags of a few periods. The argument is that if prices are expressed as deviations from their within-group average, their correlation with error can be eliminated. For the advertising equation the characteristics of the products are employed as instruments. The results show that manufacturers adjust their advertising expenditures to demand shocks. These results point to the classic problem of simultaneity bias in advertising and sales equations (Bass, 1969).

Petrin and Train (2010) use household data to estimate demand for cable and satellite television. Endogeneity in price arises due to unobserved product characteristics such as quality of programming. Petrin and Train (2010) use a control function to correct for omitted variables bias (asymptotically). This approach still requires valid instruments though, and therefore Hausman-type price instruments are employed (Hausman 1996). Control functions relate to the idea that the optimal instrumental variable is created by just adding random error to the endogenous variable, expressed by Reiss and Wolak (2001).

Danaher and Dagger (2013) assess the relative effectiveness of multiple advertising means. The authors employ individual-level single-source panel data taken from the loyalty program database of a department store. This study examines the relative effect of TV advertising, radio, newspaper, magazine, online display advertising, sponsored search, social media, catalog, direct mail, and e-mail channels on sales, profits and store visits. Most of these advertising-related variables are considered endogenous. The managers of the department store have information on prior sales history and contact details for each individual who belong to the loyalty program. Therefore, the managers may utilize this information to target specific individuals via social media, catalogs, mail and e-mail. Danaher and Dagger (2013) suggest that the number of mail and e-mail contacts before the main campaign, previous purchase incidence, sum of visits to rival websites, membership period of a person in the loyalty program, and the number of times Google was employed for the investigation of rival firms in the previous month can be used as legitimate instruments. The results suggest that this is a worthwhile use of the aforementioned instruments because at least one advertising variable is endogenous.

Dinner et al. (2014) use individual panel data to examine the cross-channel effects of online advertising such as display and search, and offline advertising such as traditional media. This study measures the direct effect of advertising on sales (offline and online) and the indirect effect of advertising on sales via impressions and click-through rates, for an expensive clothing retailer. All of these advertising-related variables are considered endogenous. Managers, plan weekly advertising to adjust to demand shocks which are unobserved to the researcher , and in paid-search advertising managers wait for the consumers to do multiple internet searches and then set their paid search budget accordingly.

Dinner et al. (2014) use marketing variables from similar but different markets as appropriate instruments. The underlying logic is that shocks in costs that create exogenous variation in marketing variables in market A will cause similar variation in market B. Advertising expenses of cheap retailers are used as instruments for advertising levels of the expensive retailer in question. Cheap retailers do not have the potential to affect the sales of the expensive retailer, thus will not be correlated with the error term, yet they will be related to the same cost function that the local retailer uses to set advertising levels. Dinner et al. (2014) run an alternative model with another set of instruments. Specifically, the weekly unit costs for television, magazine, newspapers, online display for the whole retail industry, and quarterly manufacturer price indexes for periodicals and newspapers are used. The results are quite similar to those of the first model.

Rutz et al. (2012) use keyword-level paid search data from a lodging chain to estimate the effect of keyword position on conversion and click-through-rates. Keyword position is endogenous because of omitted variable bias and measurement error. Conversion rate is defined as the ratio of number of clicks over the number of sales. If the conversion rate of a keyword is zero for a considerable number of observations this leads to measurement error. An alternative metric of the conversion rate is the click-through-rate, which is defined as the ratio of the number of users who click on a specific link over the total number of users who view that link. Yet, the click-through-rate does not provide managers with information regarding the performance of the keyword therefore both metrics are required. Moreover, the position of a keyword is determined by an auction the function of past clicking behavior of the individual and rival bids from other firms which try to promote their own keywords. The missing of competitive information leads to an omitted variables bias.

Rutz et al. (2012) discuss the candidacy of lagged variables like lagged position and lagged cost per click, as possible instruments. However, these variables still measured with substantial error and lags are prone to autocorrelation. An alternative IV estimator is employed which is the Latent instrumental Variables (LIV) estimator and does not require observed instruments. By fully utilizing the existing data, such estimators derive a latent variable with the properties of a good instrument. The results show that the LIV estimator gives the best estimates as compared to other estimators used in the study.

It has already been pointed out that in some cases finding a good instrument might be difficult. Customer related variables usually fail to satisfy the exclusion restriction and the cost related variables may not have the necessary amount of variation to shift the endogenous variable (Hartmann et al., 2011). So, one should opt for other techniques to examine the causal effect of interest. The following methods discussed in the paper are used to control for the assignment mechanism, and thus correct for self-selection bias.

2.2 Regression Discontinuity

RD was introduced by Thistlethwaite and Campbell (1960) and although has become well known in social sciences, its applications in marketing have been sparse (Hartmann et al., 2011). In such designs, all observations receive a score and this is described by a continuous variable which is called the forcing variable and the treatment is assigned only to those observations whose score is above a predetermined threshold. Hereafter, the observations whose score is below the threshold represent the control group. RD analysis can be interpreted as discontinuity at a threshold or as local randomization (Jacob and Zhu, 2012). The former interpretation focuses on a discontinuous jump in the outcomes of the observations with a score near the threshold, where the direction and the magnitude of the jump represent a direct measurement of the causal effect. The latter interpretation is dependent on the premise that observations which are just below the threshold are, on average, exactly like those which are just above it, where the difference in the average outcomes represent a direct measurement of the causal effect. In that sense, the outcomes of the control group can be used as valid counterfactuals. In both cases though, the causal effects are estimated via local linear regression, since these effects are identified in differences in the average outcomes of observations whose score is around the threshold .

RD has three conditions to be satisfied. First, the observations whose value is around the threshold must not differ from their treatment status. Second, the threshold should be determined by exogenous sources, so the selection of observations into treatment is completely dependent on their scores and the threshold. This condition ensures that the selection mechanism is random and cannot be manipulated. Third, the treatment status is discontinuous, meaning that there are no other modes in which observations on either side of the threshold are treated separately. If these conditions are satisfied, RD will consistently estimate the LATE. In contrast to exclusion restriction, the conditions related to RD can be checked. Hence, the 7 papers which are revised in Table 2 test these conditions or provide solid arguments where there is no need for testing because of the design.

RD is employed to identify causal parameters in marketing models mainly related to targeting. The estimation of causal effects of targeted marketing can be trivial because the outcome variable is possibly correlated with the marketing variable via a targeting rule which results to self-selection bias (Hartmann et al., 2011). The use of heuristics or arbitrary rules, based on consumer characteristics or past behavior, employed by marketing managers for targeting specific segments allow RD designs to identify the causal effects of marketing variables in question. The existence of such assignment rules has created an abundance of discontinuity settings which have so far been underutilized so far in marketing. In the appendix of Hartmann et al. (2011), there is a specific table (C.1) which describes examples of potential RD applications in marketing.

Table 2 provides an overview of the 7 articles which are reviewed in this Section. In particular, Table 2 describes the name of the paper, academic journal, type data, outcomes, treatment and forcing variables.

Table 2: RD Applications in Marketing Research

Author(s)	**Name of the Paper**	**Journal**	**Data**	**Dependent Variable(s)**	**Potential Endogenous Variable(s)**	**Proposed Forcing Variable**
Busse et al. (2006)	$1,000 cash back: The pass-through of auto manufacturer promotions	The American Economic Review	Car-Level Data from Automobile Transactions Data and Promotions Listings	Price	Dealer Cash and Customer Cash Promotion	Time
Hartmann et al. (2011)	Nonparametric Estimation of Marketing-Mix Effects Using a Regression Discontinuity Design	Marketing Science	Customer Level Data from Casino Database, Zip-Code level Data from Geographic Marketing Database	Profit. Customer Response	Casino e-mail Promotions, Direct-mail Promotions	Average Daily Win, Probability of Response
Yuan (2008)	Estimating the Efficiency Improvement of the Resource Allocation in the Yahoo! Keyword Auction	International Journal of Humanities and Social Science	Keyword-Level Data form Yahoo! Keyword Auctions	Bidding Values	Use of GSP	Time
Luca (2011)	Reviews, reputation, and revenue: The case of Yelp.com	Harvard Business School NOM Unit Working Paper	Customer-Level Data from Yelp.com Review Data and Revenue Data from Washington State Department of Revenue	Revenues	Rounded Average Ratings	Unrounded Average Ratings
Narayanan and Kalyanam (2015)	Position Effects in Search Advertising and their Moderators: A Regression Discontinuity Approach	Marketing Science	Keyword-Level Data from Google Keyword Auctions	Click Through Rate and Sales	Higher Keyword Position	Difference in Rank Between Keywords of Higher and Lower Ranking Scores
Chen et al. (2009)	Learning from a service guarantee quasi experiment	Journal of marketing research	Telephone Survey on Customer Evaluations Varied by Time and Location	Customer Service Evaluation Scores	Application of Service Guarantee Programs	Time
Snider and Williams (2014)	Barriers to entry in the airline industry: A multi-dimensional regression-discontinuity analysis of AIR-21	Review of Economics and Statistics	Passenger-level Data from the U.S. Department of Transformation's Origin and Destination Survey, Enplanement Data from the Federal Aviation Administration, and Survey Data about Carrier-Airport Specific Leasing Agreements	Price	Application of a Congressional Legislation	Airport Concentration Level

Busse et al. (2006) combine automobile transactions data and promotion listings data to estimate the effect of customer cash and dealer cash promotion treatments implemented by automobile manufacturers on the transaction prices, that buyers pay. Customer cash promotions relate to a subsidy that a customer receives by the manufacturer to buy a car. Dealer cash promotions relate to a bonus that a dealer acquires from the manufacturer for every car that is sold within the promotion period. Concerns about selection issues arise because manufacturers are expected to apply promotions during periods of slow sales, and when consumers push for low prices. Skipping this information will lead to a biased estimation of the effect of promotion since this effect is confounded by unobserved demand shocks.

The basic demand conditions though, which shape the rate of sales and consumers' willingness to pay, cannot change so much within a short period of time. This allows for an RD design, where the local average treatment effect of promotion can be identified just before and just after the period when a promotion is applied. Busse et al. (2006) examine the effect of a temporal treatment discontinuity, where the amount of the promoted cash sharply changes within that period. Despite the fact demand conditions may fluctuate within this short period, provided that there is no discontinuous change apart from the changes in the cash being offered, RD will consistently estimate the effect of promotion. Busse et al. (2006) focus on transactions which take place within a period of one week. The forcing variable is time and the threshold is the given week when promotions are being applied. Any discontinuities in transaction prices can then be attributed to the promotion offered that week.

The willingness to pay of customers who buy right before and right after the promotion change should be equivalent. However, if there is a substantial amount of deal-prone customers in the sample with bargaining skills who anticipate the promotion, then there are customers who would pay higher prices and customers who would pay lower prices despite the presence promotion. In that case, promotion effects should be attributed to unobserved sample characteristics other than the treatment status. After empirically checking for this assumption Busse et al. (2006) find that results do not change with respect to promotion effects although there are some deal-prone customers in the sample.

Hartmann et al. (2011) examine the effect of e-mail promotions on future profitability and the effect of direct mail promotions on response rates, while controlling for self-selection bias arising by targeting rules. The first study exploits the database marketing data of a casino’s loyalty program and the second exploits geographic marketing data of a business-to-business company. Concerning the direct mail application, the forcing variable represents the probability of a customer's response at a zip-code level, which is determined as a function of various zip code characteristics. The direct mail is then sent to the zip codes with probability of response above a given threshold. Any discontinuity in the response rate of customers with a probability of response around the threshold can be attributed to the effect of the direct mail. For the casino e-mail application the forcing variable represents an index of the money a consumer has gambled in the past. The casino then either rewards or does not reward the customers according to their gambling activity. If a customer is expected to be profitable, then receives an e-mail which indicates the relevant reward. In that case, any discontinuity in the expected profitability from customers whose past gambling activity is above the threshold (which has been defined by the managers of the casino) must be caused by the e-mail promotion.

Hartmann et al. (2011) explain that the RD estimates, in both studies, are valid because the customers cannot affect their selection into the treatments. In the casino study customers are classified into specific tiers, where each tier provides special promotional offers and hence different treatments, according to their gambling activity. Despite the fact that customers are aware that more bet will advance them into higher tiers with more lucrative rewards, they do know the precise thresholds that cluster them into specific tiers neither how their past gambling activity is determined. In the direct mail study customers are classified with respect to an unknown function. Even if these customers are aware of how the probability of response is estimated, the promotional benefits from receiving the direct mail promotion are insignificant with respect to the moving costs in order to get a zip code with higher probability of response.

Yuan (2008) evaluates the effect of using GFP (Generalized First Price) auction instead of GSP (Generalized Second Price) auction on bidding values offered by firms in their effort to achieve a desirable keyword position for their products. The paper examines the efficiency of an auction with respect to how well the monetary value of the bid determines the desirable position of the keyword. In a fully efficient auction the bidders with higher values should always outperform the bidders with lower values. The author uses bid data gathered from Yahoo! keyword auctions. The selection bias comes from the intention of bidders who are self-selected into the treatment of using GSP as a new and potentially more efficient auction type.

In the RD analysis of Yuan (2008) the forcing variable is time, the treatment variable is the use of GSP over GFP, and the threshold is the introduction day of GFP. Then, any discontinuities in the bidding behavior can be attributed to GSP. The identification challenge is that the bidders should act in GFP as they would have done in GSP. In other words, bidders should behave in GFP as if there were no auction change. This assumption is violated in the specific paper and the author reports that bidders act more strategically in GSP by frequently changing their bids and thus 'gaming' the system. In this case, any discontinuity in the bidding values among firms which use GFP and those which use GSP, might be caused by strategic behavior and not by the selection of GSP.

Luca (2011) measures the effect of Yelp ratings (a large website for food restaurants in the US) on fast food restaurant revenues where the author combines the review data and the revenue data. Yelp exhibits the average rating for each restaurant rounded to half-star. The treatment represents the sorting of a restaurant right after a certain star (or a half-star). The forcing variable describes the unrounded average rating as given by the reviewer. The threshold is the specific star rating which is expected to create a discontinuity in the revenues of restaurants whose average ratings are sorted after it. Restaurants may fake the ratings by positively evaluating themselves. However, if gaming seriously affects the revenues, then there should be a considerable amount of ratings sorted just right after the star rating in question. This is not the case for the specific sample as reported by the author. Moreover, the rounding process relates to institutional features of Yelp and therefore is independent of unobserved factors such as restaurant quality. Thus, the threshold is determined by exogenous sources and the assignment of restaurants into treatment is completely dependent on their average ratings and the star rating which represents the threshold.

Narayanan and Kalyanam (2015) examine the effect of keyword position on click through rates and sales orders resulting from Google's GSP. The authors use a dataset consisted of daily observations which include information on a focal bidder as well as its closest (but not all) rivals. The determination of keyword position in search engine auctions contains serious selection issues. Consumers are targeted with respect to their propensities to view the link that the ad leads to or even to make a purchase of the product being sponsored by that link. Omitted variables bias arises because one cannot obtain complete information regarding competitive behavior (although this is not the case for Yahoo! auctions). According to Rutz et al. (2012), the availability of appropriate instruments in such contexts is extremely difficult.

Narayanan and Kalyanam (2015) propose an RD design by considering higher keyword position as the treatment variable. The forcing variable is the difference of the rank (quality score of the keywords which is used as an input for the search engine auction) between the keywords of higher and lower ranking scores. The threshold is the keyword position which is expected to create a discontinuity in the two aforementioned outcomes. The identification of the keyword position effect is dependent on bid selection which would position the keyword just above the threshold. In contrast to Yahoo! GSP, the Google GSP is much more efficient in the sense that advertisers' bidding behavior reflects their motivation in achieving a specific position for their keyword, instead of setting bids only for placing the keyword in a higher position. Google GSP appears to be almost fully efficient and therefore the behavior of the bidders cannot invalidate the results.

Chen et al. (2009) investigate the effect of a service guarantee program applied by a middle-priced hotel chain on customer evaluation scores. Because the participation of customers to the program is promoted via lobbies and cards placed in guest rooms, the authors apply an RD design in order to overcome selection issues. Telephone surveys were conducted by a third party marketing research firm before and after the implementation of the service guarantee program period. Time is the forcing variable and the period in which the service program was applied is the threshold. This study examines any discontinuities in customer evaluation scores caused by the application of the customer service program. The identification of this effect can be invalidated if there are loyal customers in the sample who would positively evaluate the service program. The authors report that the number of these customers is insignificant and even when excluded the results do not change.

Snider and Williams (2011) applied an RD analysis to examine the impact of a Congressional legislation (AIR-21) on competition at major US airports which are considered to be market-saturated. AIR-21 aims at increasing competition by supporting new entries by smaller low-cost airline firms. The mandate of this legislation is that airports above a certain threshold of concentration (more than 50% of customers in the airport are being serviced from two major carriers) must proceed with certain actions to ensure the access of new entrants to airport facilities. The analysis examines whether the application of AIR-21 creates a discontinuous decrease in the airline fares.

Snider and Williams (2014) combine passenger level data from the U.S. Department of Transformation’s Origin, Destination Survey enplanement data from the Federal Aviation Administration, and survey data referring carrier-airport specific leasing agreements. Selection issues arise because the level of concentration is expected to be highly associated with unobserved airport characteristics. The causal effect of ARI-21 can be invalidated if airports are able to lower their concentration levels in order to be excluded from the treatment. Although this would not be feasible, mainly due to extremely large costs in adjusting the trafficking of enplanements, the authors test the possibility of such a situation and find that airports do not behave strategically.

2.3 Differences-In-Differences

DID was invented by Ashenfelter (1978). The method removes the endogenous variation from a model instead of including exogenous variables like IV and RD (Angrist and Pischke, 2008). However, the exogenous variation in DID models is captured by the treatment variable which represents an exogenous random shock such as a new legislation, an opening of a store in a region, or a promotional campaign funded by a third party. In DID designs the assignment mechanism is interpreted as a function of time, but the researcher should control for fixed effects and time-varying effects in order to prevent self-selection bias. It should also be noted that DID requires a panel data structure in order to be implemented.

DID combines two single difference estimators to identify the treatment effect. The first estimator applies cross-sectional comparisons to wipe out any omitted trends (time-varying effects) and the second estimator applies time series comparisons to wipe out any unobserved-but fixed effects (Roberts & Whited, 2012). Fixed effects represent any time invariant characteristic of an observation in the sample. Alternatively, one might say that fixed-effects control for unobserved heterogeneity which is constant over time. Sometimes the treatment variable varies only at a more aggregate level such as state, market, segment, website or distribution channel. The unobserved parameters, when examining such fixed-effects, must therefore be identified at the same level of aggregation.

The key mechanism of DID lies within the additive structure for the outcome values of the control group. Specifically, the method implies that in the absence of the treatment, the outcome values are determined by the additive manner of a time-invariant level-specific effect and a time-varying effect that are common across level-specific observations (Angrist and Pischke, 2008). What remains after differencing the differenced mean outcomes of the treatment and control groups just after the treatment with the differenced mean outcomes of the same groups just before treatment is the effect of the treatment.

There is one condition that should be met in DID estimations. This condition is called 'common trends' and can be formally checked using multi-period data. In particular, the outcomes of the treatment and control groups should exhibit common trends in the pre-treatment period. Common trends implies that in the absence of treatment, the average change in the outcomes of the two groups would have the same. Therefore, any deviation from the common trends can be attributed to the treatment (Angrist and Pischke, 2015). If this condition is met, then DID consistently estimates the treatment effect.

The papers included in this Section discuss the quasi-experimental setting which is created after exploiting information from time-specific events and check when the data allows for the condition of common trends. Table 3 provides an overview of the 6 papers that are reviewed in this Section. Table 3 describes the name of the paper, academic journal, panel data, outcomes, external shock (treatment) and fixed effects.

Table 3: DID Applications in Marketing Research

Author(s)	Name of the Paper	Journal	Panel Data	Dependent Variable(s)	External Shock (Treatment Assigned due to an Event)	Proposed Fixed Effects (FE)
Chevalier and Mayzlin (2006)	The effect of word of mouth on sales: Online book reviews	Journal of marketing research	Consumer Reviews and Book Characteristics Data	Sales	Consumer Reviews	Book-site FE and Book FE
Danaher et al. (2010)	Converting pirates without cannibalizing purchasers: the impact of digital distribution on physical sales and internet piracy	Marketing science	Consumption of Pirated TV Content, Sales of DVD Season Box Sets	Piracy Levels, DVD sales	Removal of NBC content from iTunes	Date and Episode Level FE, Date and DVD Season Box FE
Dhar and Baylis (2011)	Fast-food consumption and the ban on advertising targeting children: the Quebec experience	Journal of marketing research	Household-level Annual Survey Data	Expenditures	Ban of Advertisement	Language, Province and Children FE
Anderson et al. (2010)	How sales taxes affect customer and firm behavior: The role of search on the Internet	Journal of Marketing Research	Customer-level Historical Transactions and Customer Characteristics Data	Catalog and Internet Sales	Opening of the First Store Bricks-and-Mortals Store in a State	Customer Characteristics and State FE
Aker (2010)	Information from markets near and far: Mobile phones and agricultural markets in Niger	American Economic Journal: Applied Economics	Market and Trader-level Data	Price Dispersion Across Market Pairs	Introduction of Mobile Phone Services	Market-pair and Time FE
Busse et al. (2006)	$1,000 cash back: The pass-through of auto manufacturer promotions	The American Economic Review	Car-Level Data from Automobile Transactions and Promotions Listings Data	Price	Dealer Cash and Customer Promotion funded by Manufacturers	Car and Week-Vehicle Segment FE

Chevalier and Mayzlin (2006) use sales, consumer reviews and book characteristics data to measure the effect of consumer reviews on the relative book sales at Amazon.com and Barnesandnoble.com (bn.com). The econometric analysis of Chevalier and Mayzlin (2006) aims to answer the question: *"If a cranky consumer posts a negative review of a book on bn.com but not on Amazon.com, would the sales of that book at bn.com fall relative to the sales of that book at Amazon.com? "* Hence, the change in reviews on one site is modeled as an external shock which affects the sales of the specific site relative to the sales of the other. The effect of online reviews though, may be confounded because consumers buy books for many reasons other than the reviews about those books. Chevalier and Mayzlin (2006), in their DID estimation, control for book-site fixed effects and book fixed effects to mitigate endogeneity concerns. Book fixed effects are constituted by offline promotion, quality of the book and popularity of the author. Book-site fixed effects represent readership preferences of website users.

Checking for common trends in the study of Chevalier and Mayzlin (2006) might be very complicated, if not unrealistic, due to differences in the ranking systems and user behavior of both sites. It is a fact that Amazon.com has many more reviews than its competitors which are rigorous and positive. One can therefore expect that the sales at Amazon.com to be higher than they would be without the provision of reviews. Overall, the results indicate that a change in reviews at bn.com cannot predict a change in sales at Amazon.com.

Danaher et al. (2010) measure the effect of digital distribution on physical sales and internet piracy. The authors examine what would have happened to the privacy level of a TV content had it taken this content out from iTunes and what would have happened to DVD sales of TV seasons had they taken these DVDs out from iTunes store (the iTunes store provides digital distribution of DVDs through Amazon)? Specifically, the removal of NBC content from iTunes due to conflicted interests between the two companies, is treated as an exogenous shock to the legal digital supply. NBC content was available for sale on iTunes, on DVD webstores like Amazon.com and on piracy channels (for free). The effects of this shock on piracy and DVD sales levels are estimated in two separate DID specifications.

The authors use datasets from Mininova.com and Amazon.com which contain information about piracy and DVD sales levels for NBC and other TV networks. Consumption data of pirated TV content on major Torrent sites, and sales data of DVD season box sets at Amazon.com are employed. For the first DID specification, Danaher et al. (2010) include date and episode-level fixed effects to examine the impact of the aforementioned event on piracy levels. For this analysis, the authors provide an identification check regarding the common trends between NBC and non-NBC piracy levels. For the second DID specification, the authors include date and DVD season box fixed effects to examine the effect of the event on physical sales. For the second study the authors do not provide any discussion about common trends.

Dhar and Baylis (2011) measure the effect of banning an advertisement on the expenses of fast-food products. Such banning is used as an exogenous demand shock. Household level survey data from a Canadian a food-expenses questionnaire is employed. The authors apply a DID model to examine whether fast food expenses are smaller for groups affected by the ban of the advertisement as compared to those who were not. The effect is estimated after differencing the fixed effects of province, language, and families with or without children. The ban should affect expenditures of households familiar with the language of the advertisement (the study was conducted in a multilingual area-Canada) located in provinces where the ban took place and with at least one child. However, not all households consume fast foods in a given week which leads to measurement error and does not allow for checking common trends due to the sparseness of the data. Dhar and Baylis (2011) overcome this problem by using PSM which is explained in the next Section.

Anderson et al. (2010) use customer-level historical transactions and customer characteristics data to assess the effect of opening a bricks-and-mortals store on the internet and catalog demand of a multichannel retailer. If this is the first store of the retailer in a specific state, then this introduces a legal obligation to gather sales taxes on all Internet and catalog orders shipped to that state. The authors analyze the buying behavior across customers who live in the same state where the retail store has opened, but at a distance from the retail trade area, and must pay sales taxes for their Internet and catalog orders for products offered by this retailer. These customers represent the treatment group as compared to those customers who live in a neighboring state and do not have to pay taxes for their Internet and catalog orders. The authors control for individual customer characteristics fixed effects but also for time trends and other group-level parameters such as state fixed effects.

Anderson et al. (2010) discuss three important caveats before the analysis. The first two are related to selection issues. Initially, the opening of the store may attract customers to buy more frequently from the store or to interchange more channels of distribution. To isolate the tax effect the authors concentrate on customers who are located more than a hundred miles away from the retail trade area assuming that these customers will keep buying through orders. Second, store location may be endogenous because managers are expected to consider consumer demographics in the specific area. That is another reason why the treatment group of customers is chosen to be located at a distance from that area. Third, the interpretation of the quasi-experimental setting demands that any factors which could affect the sales of one group should also affect the sales of the other. Otherwise the effect of the treatment is confounded. The major test of the third caveat is that of common trends. The authors report that after inspecting historical sales trends of both groups they find no difference.

Aker (2010) via the use of market and trader-level data investigates the effect of the introduction of mobile phones on price dispersion across grain markets. Particularly the introduction of a mobile phone service in Niger is used as the external shock. The idea of this study is that grain traders travel around markets to gain information with respect to price. Therefore, the introduction of mobile phone services should have decreased their search costs and hence increased the dispersion of prices. The dependent variable is defined as the absolute value of the price difference between two markets. The treatment represents the introduction of mobile phone service in the two markets which define the differential price. The first set of controlled parameters used in this model is related to the market-pair fixed effects such as geographic location, urban status, and market size. The second is time fixed effects and the third represents pair-specific time trends. The identification challenge of this study lies in the common trends of dispersed prices between treated and untreated market pairs. Through the use of many formal tests the author reports that, overall there are no statistically significant differences between the treated and untreated group in the pretreatment period.

Recalling the study of Busse et al. (2006) which is described in the RD Section, the authors also apply a DID specification in order to measure the effect of dealer cash and customer cash promotions funded by manufacturers. Hence, this paper represents an exemplified example of what one can do with observational data in a quasi-experimental setting. Busse et al. (2006) use the customer cash and dealer cash promotion events as external shocks. For the individual level parameters, the authors include car fixed effects such as model, model year, body type, doors, cylinders, transmission and trimmed level. For the aggregate level parameters, the authors use week-vehicle segment fixed effects representing the cross-sectional effects of segments such as SUV and compact cars. The identification assumption of common trends is related to the indifference of price trends of non-promoted cars with that of the promoted ones which belong to the same segment just before the period of promotion. This assumption is checked and ensures that the model consistently estimates the effects of dealer and cash promotions on transaction prices.

2.4 Propensity Score Matching

Considering the papers that have been retrieved so far, most of the empirical studies in marketing are confronted with the evaluation problem of unobservability in potential outcomes and the possible appearance of self-selection bias in the ATT. PSM has received significant attention to causal research since the works of Rubin (1973a and 1973b). The technique is based on matching treatment and control observations according to their observable characteristics. The underlying procedure is divided into two steps. First is the estimation of propensity score, which is the probability of an observation to be treated or non-treated, and estimated via limited depended variables models such as probit, logit and tobit among others. The independent variables in these models are the ordinary control variables which are also used in quasi-experimental econometric models and relate to observed characteristics of observations included in the sample. After estimating the propensity score of each observation, a matching algorithm, such as Nearest Neighborhood or Kernel functions, is applied to select the closest non-treated observations to the treated ones given their propensities where the algorithm drops the remaining observations out of the sample (Caliendo and Kopeinig, 2008). In essence, PSM methods are used to correct for selection issues by controlling for observable pre-treatment differences between the treatment and control groups.

PSM is a very promising approach for estimating counterfactuals because it links the statistical properties of estimating treatment effects with the experimental setting which allows for causal inference. The basic rationale of the method is that one has to find, within a group of non-treated individuals, those observations whose characteristics before the treatment are similar to those of the treated individuals. PSM is not based on identifying the correct model that fits the data hand, but minimizes the dimensionality of the matching problem by the propensity score. The method is also well known for its flexibility on the functional form, because the difference in the outcome means between the treated and non-treated group does not have to be modeled as linear in the difference in covariates (Caliendo and Kopeinig, 2008). A major advantage of the method is that it can be applied to any study where there is a treatment and a control group. Even if there are no suitable instruments, discontinuities or external shocks, PSM can always be used to assign observations to treatment .

As any other quasi-experimental technique, PSM is based on a certain assumption which is called conditional independence (or unconfoundedness). This condition implies that the potential outcomes are independent of the treatment assignment, given a set of baseline covariates (Heinrich et al., 2010). In practice though, conditional independence cannot be checked. Therefore, the estimates are assumed to be consistent when the selection mechanism is well known by the researcher or PSM is combined with other quasi-experimental techniques which control for the selection mechanism in the first place. Regarding the knowledge for the selection mechanism, the researcher may ask the marketing manager of the company under examination about the criteria by which customers are selected. Last but certainly not least, PSM provides good robustness checks (Roberts and Whited, 2013).

The 6 papers included in Table 4 were selected because they explicitly discuss the selection problem that may arise with respect to their studies. Specifically, Table 4 describes the name of the paper, academic journal, data, outcomes, treatment and the observed covariates.

Table 4: PSM Applications in Marketing Research

Author(s)	**Name of the Paper**	**Journal**	**Data**	**Dependent Variable(s)**	**Propensity of the Treatment Assignment**	**Proposed Covariates**
Dhar and Baylis (2011)	Fast-food consumption and the ban on advertising targeting children: the Quebec experience	Journal of Marketing Research	Household-level Annual Survey Data	Expenditures	Purchase in a given week	Demographics, Year Specific and Seasonal Effects
Mithas et al. (2005)	Why do customer relationship management applications affect customer satisfaction?	Journal of Marketing	Archival data for a Cross-section of Firms	Customer Knowledge and Customer Satisfaction	Ability to Handle Sophisticated CRM systems	IT intensity, Industry Sector and Supply Chain Integration
Tsai et al. (2015)	What's In A Name? Assessing the impact of rebranding in the hospitality industry	Journal of Marketing Research	Hotel-level Annual Survey Data	Occupancy Rate among others	Rebranding	Marketing Expenses, Number of Rooms, Management Fee and Renovation among others
Huang et al. (2012)	Wal-Mart's impact on supplier profits	Journal of Marketing Research	Serial data for a Packaged Goods Category offered by a Single Supplier	Profits	Wal-Mart Entry	Per Capita Income, Population, Population Growth Rate, Number of Other Supercenters within 20 Miles Radius, Herfindahl Index, and Median Household Size
Datta et al. (2015)	The Challenge of Retaining Customers Acquired with Free Trials	Journal of Marketing Research	Customer-Level Marketing, Transaction, and Usage-related Data	Customer Retention and Customer Usage	Subscription for a Free Trial Version of iTV	Age, household size, Income, Time to Adoption, Direct Marketing, Advertising Intensity, Total Fees for a Regular Subscription
Wangenheim and Bayon (2007)	Behavioral consequences of overbooking service capacity	Journal of Marketing	Customer-Level Transaction and Revenue Data	Customer Transaction Behavior and Revenues	Downgrade Customer Status, Upgrade Customer Status, Deny Boarding	Number of flights/High Value Bookings, Accrued Miles and Socio-demographic Characteristics

Dhar and Baylis (2011) face the challenge of sparse data on households' expenditures because not all households consume fast-food in a given week. However, according to economic theory purchase decisions directly affect expenditures' measurements. The authors estimate the propensity of a household to buy fast-food in a given week as a function of demographics, year specific and seasonal effects. Given this information the expected amount to be spent per week on fast-food can be estimated for each household. Then a DID specification is applied to estimate the causal effect of banning an advertisement. Common trends cannot be tested though, so the matching exercise is employed as a robustness check. Dhar and Baylis (2011) report similar results between PSM and DID analyses.

Mithas et al. (2005) use archival data of a cross-section of firms to evaluate the effect of Customer Relationship Management (CRM) on customer satisfaction, where CRM describes the sophistication of a company in handling customer-related data. Selection issues arise because CRM sophistication is endogenous due to unobserved differences among firms, hereafter Mithas et al. (2005) use a PSM procedure to match the sample on IT intensity, industry sector, firm size, supply chain integration and customer knowledge of a firm. These observables are assumed to determine the propensity of a firm to be sophisticated in CRM practices.

Tsai et al. (2015) use data from an annual survey of the hotel industry to quantify the impact of rebranding on performance. Emphasis is given in occupancy rate as an important performance indicator. The effect of rebranding is estimated for a franchised chain of hotels which operate under a major umbrella brand. The treatment effect is estimated for hotels which switched their brand affiliations. The treatment and control groups were selected according to their zip code (i.e 10 miles distance) in order to prevent the effect of common demand shocks. This step recalls the paper of Anderson et al. (2010) which is retrieved in the DID Section. However, this precautionary step is not enough to prevent endogeneity. The treated and non-treated hotels still differ in a variety of ways. The treated hotels have more rooms, lower reservation rates per room, larger frequency in renovations, and lower overall performance. The differences between the two groups raise self-selection bias concerns where an underperforming hotel is more likely to be rebranded. Rebranding might also occur because of marketing activities, renovation and management structure.

Tsai et al. (2015) combine fixed-effects and IV. The hotel-fixed effects, such as the total number of rooms, are included in order to account for time-invariant unobserved factors. But, this is not enough since the effect of rebranding can be attributed to hotel-and time-varying unobserved factors which are correlated with performance. For instance, if a new recreation park is going to be developed in the area and the demand of a hotel is expected to increase then the owner may rebrand the hotel in order to attract customers who like leisure activities. Additionally, rebranding could boost the reactions from rivals. The authors also control for aggregate time-varying factors, market fixed effects and time fixed effects in order to correct for the aforementioned endogeineity concerns.

Tsai et al. (2015) use PSM as a robustness check in order to address the problem of non-comparability of the treated and non-treated hotels. The PSM is based on the propensity of a hotel to be rebranded given certain covariates such as marketing expenses, number of rooms, management fee and renovation among others. The reason behind the extended discussion about the study of Tsai et al. (2015) is to stimulate the importance of PSM, especially when there are multiple selection issues and even the most advanced specifications cannot guarantee that observations are not self-selected into treatment.

Huang et al. (2012) use serial data of a packaged goods category offered by a big supplier to examine the effect of Wal-Mart's market entry on supplier profits. Wal-Mart because of its bargaining power has been criticized of squeezing the performance of its suppliers. Selection issues arise because the entry of Wal-Mart in certain markets is not random. Huang et al. (2012) explain the difficulty in finding appropriate instruments and then apply a PSM procedure. The propensity Wal-Mart entering a particular market is based on per capita income, population, number of other supercenters within 20 miles, median household size, Herfindahl index and population growth rate of a market. In turn, the authors apply a DID model since the entry of Wal-Mart in a market can be interpreted as a time-related treatment that is exogenous in the suppliers' profit equation. The paper of Huang et al. (2012) exhibits informative graphs concerning the distributions of the propensity scores and matching estimates for treated and non-treated markets. The graphs are impressive as they illustrate the process of matching a sample.

Datta et al. (2015) test the effects of free-trial promotion service on customer retention and customer usage via the use of a large sample of customers who adopted iTV when both free trial and the ordinary subscription were available. The dataset is composed of customer-specific, marketing, transaction and usage-related information provided by the company (CRM data). Datta et al. (2015) apply PSM in order to control for the selection of consumers into the free-trial and the ordinary subscription customer groups. The propensity of an iTV user of trying the free-trial version or not is estimated as a function of age, household size, income, time to adoption, direct marketing, advertising intensity, and the total fees for a regular advertising subscription at the time of a customer sign-up. Dissimilarities in consumer behavior can then be attributed to free-trial acquisition rather than the consumers' characteristics.

Wangenheim and Bayon (2007) evaluate the effects of downgrading and upgrading customer status, and denying boarding on customer transaction behavior and revenues. The authors use transaction and revenue data (CRM data) from a customer database of a global airline company. Airlines though, do not randomly upgrade (from economy class to business seats) or downgrade (from business seats to economy) a customer, or deny his boarding. In fact, airlines apply certain rules for selecting customers into such treatments. These rules relate to availability of seats on the aircraft, customer status (gold, silver, or bronze), amount of points collected in the loyalty program and the price paid for the current flight . The authors also report that the usage patterns were considerably different in all treatment groups as compared to the usage patterns of typical customers. In this study the control group represents typical customers who are not members of the loyalty program.

Wangenheim and Bayon (2007) argue that when the selection mechanism is well known, PSM gives the best results as compared to other quasi-experimental techniques. The covariates used for the estimation of propensity scores are close approximations, but not identical, to the rules of the company. As the authors explain, the purpose of the model is to obtain parameter estimates for each individual and not to replicate the company's classification procedure. The propensities of each treatment are based on the number of flights or high value bookings (for downgrading), the accrued miles and socio-demographic characteristics. Wangenheim and Bayon (2007) apply a DID specification to model the treatments as external shocks in consumer responses.

3. DISCUSSION

3.1 Problems that have been solved with the use of quasi-experimental econometric methods

Table 5: Comprehensive representation of papers reviewed with a short description of the endogeneity problem which was tackled via the use of a specific quasi-experimental econometric method.

Author/Year	Endogeneity Issue	Method
Villas—Boas and Winner (1999)	The effect of price on demand may be confounded by unobserved input prices and demand shocks.	IV
Kuksov and Villas-Boas (2008)	The effects of price, display and feature on demand may be confounded by unobserved cost-related product characteristics.	IV
Chintagunta and Dube (2005)	The effect of price on demand may be confounded by unobserved brand characteristics.	IV
Barroso and LLobet (2012)	Price, demand and advertising are simultaneously determined through unobserved cost-related product characteristics.	IV
Petrin and Train (2010)	The effect of price on demand may be confounded by unobserved product quality.	IV
Danaher and Dagger (2013)	The effects of advertising on sales, profits and store visits may be confounded by unobserved parameters with respect to unobserved targeting rules and competitive information.	IV
Dinner et al. (2014)	The effects of online and offline advertising on sales, search impressions, click trough rates may be confounded by unobserved b demand shocks	IV
Rutz et al. (2012)	Conversion rates might be dominated by zeros which leads to substantial measurement error. The effect of keyword position on conversion and click through rates may be confounded due to lack of competitive information.	IV
Busse et al. (2006)	The effect of cash promotions on transaction price may be confounded by time-varying demand shocks, product fixed effects and time-segment fixed effects, meaning that manufacturers do not apply promotions randomly	RD & DID
Hartmann et al. (2011)	The effect of casino e-mail promotion on future profitability and the effect of direct-mail promotion on probability of response, both of them may be confounded by unobserved parameters with respect to targeting rules	RD
Yuan (2008)	The effect of using GSP on bidding values may be confounded by bidders' self-selection in to treatment in an effort to use a more efficient auction mechanism.	RD
Luca (2011)	The effect of consumer reviews on revenues may be confounded by unobserved product quality which leads to selection of of treated restaurants	RD
Narayanan and Kalyanam (2015)	The effects of higher keyword position on sales and click through rate may be confounded by unobserved parameters with respect to targeting, strategic behavior of advertisers in their bid setting and lack of complete competitive information.	RD
Chen et al. (2009)	Hotels customers may self-select themselves into the guarantee service program treatment because of this program is promotes via lobbies and cards placed in the guest rooms	RD
Snider and Williams (2014)	The effect of applying legislation on increasing competition in airports may be confounded by unobserved airport characteristics which lead to selection of treated airports	RD
Chevalier and Mayzlin (2006)	The effect of consumer reviews on the book sales of a web-site may be confounded by book-related and web-site related fixed effects.	DID
Danaher et al. (2010)	The effect of removing the NBC content from iTunes on piracy levels may be confounded from date episode-level fixed effects. The effect of removing NBC content from iTunes on NBC DVD sales may be confounded from date and DVD Box season fixed effects.	DID
Dhar and Baylis (2011)	The effect of banning an advertisement on consumers' fast food expenditures in a given week may be confounded by province, language and family status fixed effects. Not all households consume fast food in a given week which leads to measurement error and selection issues.	DID & PSM
Anderson et al. (2010)	The effect of opening a bricks-and-mortals store in a state on catalog and internet sales may be confounded from customer characteristics fixed effects, time trends and state fixed effects.	DID
Aker (2010)	The effect of introducing mobile phone services in a country on price dispersion across markets may be confounded by market-pair fixed effects and time fixed effects.	DID
Mithas et al. (2005)	The effect of CRM sophistication on customer knowledge may be confounded by differences in firm-related characteristics.	PSM
Tsai et al. (2015)	The effect of rebranding on a hotel's performance such as occupancy rate may be confounded by unobserved hotel- and time-varying factors, and market, time and hotel fixed effects.	PSM & IV
Huang et al. (2012)	The effect of Wal-Mart's entry on supplier profits may be confounded by differences in market specific characteristics.	PSM & DID
Datta et al. (2015)	The effect of promoting a free-trial version of iTV on customer retention and customer usage may be confounded by differences in consumer characteristics and consumer-specific marketing actions	PSM
Wangenheim and Bayon (2007)	The effects of upgrading and downgrading customer status and denying the boarding of a customer on transaction behavior and revenues may be confounded by the rules of the company in selecting that customer into one of those treatments.	PSM & DID

Table 5 provides an overview of the endogeneity problems (OVB/SSB, SB and ME) which have been solved via the use of quasi-experimental econometric methods (IV, RD, DID and PSM) in the 26 revised papers. It is evident that these techniques can be used in a wide variety of marketing concepts such as Demand, Pricing, Profitability, Competition, Sales, Consumer Reviews (which can be related to Word of Mouth), Advertising, Promotion, CRM, Branding and Segmentation. The revised studies in this thesis describe marketing problems to be solved, endogeneity concerns, identification strategies and the causal mechanisms which relate marketing variables to business outcomes. Marketing researchers should elaborate on the these issues in their observational studies because only with concise and convincing theoretical arguments and analysis can one cope with endogeneity (Roberts and Whited, 2013). The papers revised in this thesis combine quasi-experimental econometric modelling, marketing theory and data structure in order to identify causality in various marketing areas.

In the absence of a completely randomized design, there is no way to ensure that endogeneity problems are removed or effectively alleviated. Endogeneity in marketing models will distort the estimates if one does not control for factors that may affect the outcome even in the absence of the marketing variable. Such factors are depicted in Table 5. Quasi experimental econometric models provide researchers with opportunities to address, at least, the most obvious sources of endogeneity by controlling for the variation of the marketing variables (IV) or by controlling for the selection of observations into marketing treatments (RD, DID and PSM). The cost of a product may be used to instrument its price (Villas—Boas and Winner, 1999). Time may be used to model the assignment of a treatment for example, the ban of an advertisement as an external shock in expenditures of customers (Dhar and Baylis, 2011). The gambling activity of a casino player may be used as a forcing variable in order to explain the selection of that player into a promotion treatment (Hartmann et al., 2011). The CRM rules of a company may be used to determine the propensity of assigning a customer into a different status such as upgrading or downgrading (Wangenheim and Bayon, 2007).

This thesis provides sound examples of how difficult it is to causally assess marketing effectiveness. Specifically, the examination of marketing effects entails strenuous research efforts due to a variety of reasons. The necessary conditions of the method may not be satisfied because of differences in observations other than the treatment (Yuan, 2008). The exogenous variables may not vary in a similar way to that of the endogenous variable (Chintagunta and Dube, 2005). There may be a scarcity of exogenous variables in the marketing area where the research takes place (Rutz et al., 2011). Moreover, the data may not allow for testing assumptions (Chevalier et al., 2006; Dhar and Baylis, 2011) and endogeneity may be caused by multiple sources (Tsai et al., 2015). Therefore, the researcher should be aware of all the quasi-experimental methods in order to examine marketing effects in a meaningful way.

3.2 The marketing researcher should be aware about all these methods because of three reasons

According to Rossi (2014) and Larker and Rusticus (2010), although IV is the standard textbook solution for endogeneity both papers indicate the strain of finding appropriate instruments. Concerning the endogenous variable of price in demand models, some of the most common instruments such as lagged prices or wholesale prices are inappropriate (Rossi, 2014). On the other hand, cost-related instruments are difficult to find due to lack of information in variable costs. Even if cost-related instruments are available they may not contain the required amount of variation, with respect to time or market, in order to shift the marketing variable. Rutz et al. (2012) explain the difficulty in finding available instruments for endogenous variables such as keyword position in click-through or conversion rate equations due to missing information in paid search auctions (such as GSP). Furthermore, Hartmann et al. (2011) indicate the limited availability of customer-related instruments for endogenous variables like that of promotion or advertising due to the fact that most of these instruments are inappropriate. An instrument is considered inappropriate when there is a serious theoretical concern which may not satisfy the exclusion restriction. Lagged variables might be correlated with demand error due to autocorrelation. Customer related variables such as demographics or customer usage are assumed to directly affect demand so cannot be used to instrument endogenous variables in demand models. It is also difficult to measure the effect of promotion on sales without including customers' income or past-spending patterns in the sales equation as controls. Therefore, these variables will be correlated with the error term of the sales equation.

Rossi (2014) states that the emphasis on price endogeneity in IV applications within marketing has been misplaced. Judging from the retrieved studies, the marketing field provides a prosperous ground for quasi-experimental applications via targeting rules (RD), external shocks (DID) and treatment propensities (PSM). As such, the emphasis on IV, as the main remedy for endogeneity in marketing models, perhaps has also been misplaced.

Therefore, the first reason that should hint the marketing researcher to be aware of all quasi-experimental methods is that IV may not always be useful. There are many natural experiments in marketing which can legitimate other techniques for causal inference. Observational marketing data is composed of various kinds of structures and a plethora of treatment selection cases. Such cases refer to cash promotions according to demand conditions (Busse et al., 2006), customized targeting according to expected profitability (Hartmann et al. 2011), the adoption of sophisticated CRM systems according to certain firm capabilities (Mithas et al., 2005) or the advancement of a keyword at a higher position according to its quality score (Narayanan and Kalyanam, 2015). Such cases should trigger the researcher to realize that RD, DID and PSM are also crucial means for exploiting the potential of a quasi-experimental design given the structure of the data at hand.

The second reason for being aware of all these methods is that there are cases when data does not allow for checking the necessary assumptions. For example, the data in the study of Dhar and Baylis (2011) is not appropriate for testing common trends due to sparseness. In the study of Chevalier and Mayzlin (2006), the assumption of common trends cannot be checked because there are large dissimilarities between the two book-sites (Amazon.com and bn.com). Although the authors apply many configurations to restrict their sample in comparable characteristics they cannot guarantee the absence of self-selection bias.

A third reason refers to the use of more than one method in order to strengthen the validity of the results. In the study of Busse et al. (2006), the authors use both DID and RD in order to strengthen their inference regarding the causal effect of cash promotions on sales. All of the papers included in the PSM Section are examples of combing different methods on order to mitigate limitations in the structure of the data, address endogeneity arising from multiple sources and test for robustness.

3.3 A proposed theoretical framework regarding the process of how these methods help to identify the problem, and how the problem in turn helps to identify the correct method

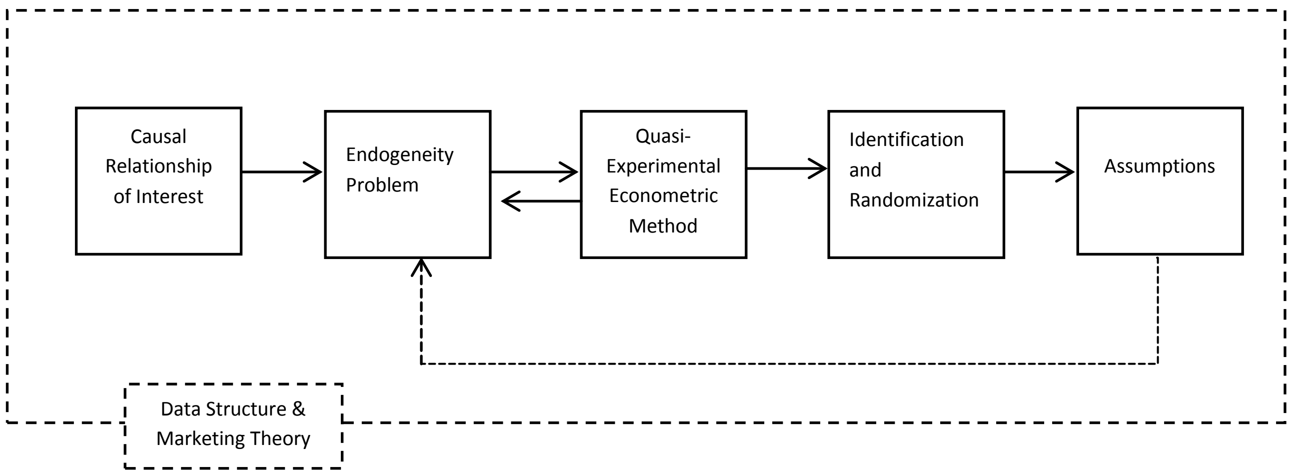

Figure 1: Theoretical framework which describes the dynamic process of how the causal relationship of interest determines the steps from the method to be used until the assumptions to be checked and the "feedback loop" of identifying an alternative method if assumptions are not met or for further robustness. Additionally, the steps of the above depicted process should be implemented with respect to marketing theory and the structure of the available data.

Given all the information extracted from this thesis, the theoretical framework in Figure 1 should help empirical marketing researchers in their scientific practice of performing quasi-experimental econometric studies with non-experimental data. Such study should start with the causal relationship of interest. This primary step will define the outcome and treatment variables. In turn, according to marketing theory the potential correlation between the treatment variable and the error term of the outcome equation should be explained. For example, in demand models the price is considered to be endogenous because of unobsereved demand shocks or unobserved product quality. Then, with the use of marketing theory and the available data the researcher should proceed to the selection of the most appropriate method. In demand models IV is the common solution but there should be available instruments such as cost-related variables. The exclusion restriction should also be justified in terms of marketing theory, i.e. costs of similar products in other markets are not expected to be correlated with the error term of the demand equation (Villa-Boas and Winner, 1999).

Each of these methods interprets the problem of endogeneity in a different way. IV gives an OVB or SB interpretation whereas DID, RD and PSM give an SSB interpretation. According to the definition of the major endogeneity problem to be solved, each method identifies certain parameters in order to approximate randomization. With respect to the selected method, endogeneity is tackled either by instruments or natural selection processes defined by external shocks, forcing variables or covariates which are expected to increase the propensity of the treatment. For example, in the RD analysis the researcher should identify a continuous forcing variable which forces observations into treatment with scores above a specific threshold. If there is a discontinuous jump in the outcome values this can be attributed to the casual effect of the treatment parameter. In RD the randomization of the treatment parameter is approximated through the forcing variable as a natural way of selecting treated observations. However, marketing theory should be used to explain the rules which are used by a company to select consumers into treatment, i.e according to the expected profitability to be gained by them after being treated with specific promotions (Hartmann et al. 2011). In addition, the researcher should verify that the company will provide the necessary data in order to mimic the selection process.

Last but not least, each of these methods has certain assumptions to be satisfied. If these assumptions are not satisfied then the researcher should opt for an alternative method. Even if these assumptions are satisfied the researcher can re-run the analysis with an alternative method to check for the robustness of the results. Figure 1 describes the way that these methods help to identify the endogeneity problem, and how the problem in turn helps to identify the correct method. Table 6 exhibits supplemental information in reference to Figure 1.

Table 6: Analytic explanation of the steps in Figure 1 for each quasi-experimental econometric method

Method	**Interpretation of the Major Source of Endogeneity**	**Identification of Major Control Parameters**	**Assignment Mechanism and Approximation of Randomization**	**Assumptions**
IV	OVB,SB	Instruments	The randomization of the treatment assignment is approximated via the exogenous variation of instruments	Relevance and Exclusion Restriction
RD	SSB	Forcing Variables and a predetermined threshold	Forcing variables approximate randomization since the treatment is enforced by a natural way	Observations just below and just above the threshold must have the same characteristics, observations cannot affect their selection into treatment, only treatment is discontinuous
DID	SSB	External Shocks and Fixed Effects	External shocks approximate randomization since the treatment is naturally assigned by time after controlling for fixed effects and time trends	Outcome values of the treated and non-treated group have common trends before the treatment period
PSM	SSB	Control Variables which are expected to determine the propensity to be treated	Selection issues are mitigated after matching, hence the assignment mechanism can be considered random	Conditional Independence

3.4 Contributions

This thesis is the first attempt in reviewing more than one quasi-experimental econometric application within the marketing context. Roberts and Whited (2013) have made a similar attempt but this is related to the field of corporate finance. While Rossi (2014) provides an extensive review on IV Hartmann et al. (2011) provides a small review on RD, yet so far, there is no paper to simultaneously review IV, RD, DID and PSM applications within the field of marketing. Thus, 26 exemplified applications, published in top-tier journals, including a wide variety of causal relationships, endogenous variables and controls are revised. The paper also provides a detailed theoretical discussion on the various sources of endogeneity with respect to marketing theory regarding the realization of the problem and the way to rule it out. Moreover the paper describes many endogeneity problems that have been solved so far via the use of quasi-experimental econometrics but also signifies the importance of being aware of all of these methods in order to cope with the ambiguity of examining marketing effects. Finally, a theoretical framework is provided to help researchers identify the problem with the appropriate method and vice versa with respect to marketing theory and available data.

3.5 Limitations

Tables 1-4 are not exhaustive. This thesis does not review all the published quasi-experimental studies in marketing literature. The aim of tables 1-4 is to present the reader with examples of quasi-experimental econometric applications in various marketing areas, but not to fully cover the available marketing literature on these techniques. Therefore, the amount of papers is not sufficient to claim that the review is systematic. Systematic reviews represent the optimal evidence-based procedure of providing an exhaustive summary of the extant literature relevant to the topic in question. The literature review conducted in this thesis has an explorative and qualitative nature in an effort to provide a first introduction of the use of quasi-experimental econometric applications within the marketing context. The answers provided for the research questions should therefore be verified through a systematic review. Moreover, the revised papers could have been selected with criteria other than those explained in the IV Section. Alternative criteria could have been the number of publications and the appearance of the name of the method and the term of endogeneity in the abstract, title or keywords.

Furthermore, in this thesis the results of the papers in terms of main effects are not discussed because emphasis is given to the way the methods are used. Another significant aspect which is not discussed is related to the quantitative manner that the assumptions of the methods can be checked. This would have provided valuable advice for the reader with respect to alternative ways of testing these assumptions, but such effort is beyond the scope of this thesis. However, the reader is advised to carefully read the retrieved papers which are related to his own research. Finally, the paper does not introduce the mathematical equations which are specified in the papers in order to examine the effects of interest. However, the appendix in the end of the paper provides some illustrative examples.

4. CONLCUSION

The most intrinsic characteristic of applying quasi-experimental econometric techniques, when well justified and applied, is that they bring exogenous variation in the model. The major challenge is that the researcher should provide convincing arguments in favor of the methods applied with respect to marketing theory, endogeneity concern and the data at hand. The reviewed papers in this thesis are outcomes of such scientific practice. These papers make strong cases for a causal interpretation of the relationship between marketing variables and business outcomes.

This thesis reviews 26 quasi-experimental econometric papers within the marketing context and discusses possible solutions of endogeneity in the empirical marketing research. This extended literature review was conducted in an effort to inform the reader on how to take advantage of observational data in order to estimate the causal effects of marketing variables on business outcomes within a quasi-experimental setting. Although the majority of marketing data is non-experimental hence suffers from endogeneity, a wide range of marketing situations allows for the estimation of causal effects in a quasi-experimental setting.

The means for tackling endogeneity in various marketing contexts are described in terms of four different quasi-experimental methods. These methods, under certain assumptions, improve over OLS and achieve consistent estimates of marketing effects. The structure of data in combination with marketing theory should provide guidance in the selection of the method according to the causal relationship of interest. Each method interprets endogeneity differently with respect to the source of bias, the identification of the causal and control parameters, and of course the mimicking of a randomization. It is of the utmost importance that the reader is aware that no matter how well the research design is created and the research is implemented, the assumptions' check may always invalidate the results. If the assumptions are not satisfied then researcher should opt for alternative methods or, at least provide caveats and apologize.

The emphasis on IV as the most common textbook solution for endogeneity perhaps has been misplaced regarding the marketing context. On the one hand, valid instruments are rare and on the other hand there is an abundance of external shocks and selection rules in order to conduct an RD, DID or PSM analyses. The papers presented in the Tables 1-4, although not exhaustive, should provide good guidance for the reader to perform a quasi-experimental study in order to estimate the causal effect of a marketing variable. Tables 1-4 present a large amount of instruments, forcing variables and fixed effects among other covariates to control for the confounding effects described in Table 5. This information should provide guidance on possible ways to control for endogeneity in marketing studies.

This thesis is a first exploratory step to give some indication of the types of endogeneity problems which have been solved in marketing using IV, RD, DID and PSM. The performance of a systematic review is the next necessary step to provide a more concrete and comprehensive evidence-based analysis on the use of these methods. Furthermore, domain and content analyses can be applied to shed light on questions such as: Which method should be used to correct for endogeneity in a specific marketing variable? Figure 1 may help researchers apply such analyses. After setting a strong theoretical ground for the application of quasi-experimental econometric methods in the marketing context then a meta-analytic review can be conducted to corroborate the proposed causal relationships between marketing variables and business outcomes. Therefore, the resources extracted for marketing investments will be optimally allocated to boost business performance.

REFERENCES

Aker, J. C. (2010). Information from markets near and far: Mobile phones and agricultural markets in Niger. American Economic Journal: Applied Economics, 2(3), 46-59.

Anderson, E. T., Fong, N. M., Simester, D. I., & Tucker, C. E. (2010). How sales taxes affect customer and firm behavior: The role of search on the Internet. Journal of Marketing Research, 47(2), 229-239.

Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Angrist, J. D., & Pischke, J. S. (2014). Mastering'metrics: The Path from Cause to Effect. Princeton University Press.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings.The Review of Economics and Statistics, 47-57.

Barroso, A., & Llobet, G. (2012). Advertising and consumer awareness of new, differentiated products. Journal of Marketing Research, 49(6), 773-792.

Bass, F. M. (1969). A simultaneous equation regression study of advertising and sales of cigarettes. Journal of Marketing Research, 291-300.

Bolton, R. N., & Drew, J. H. (1991). A multistage model of customers' assessments of service quality and value. Journal of consumer research, 375-384.

Busse, M., Silva-Risso, J., & Zettelmeyer, F. (2006). $1,000 cash back: The pass-through of auto manufacturer promotions. The American Economic Review, 1253-1270.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of economic surveys, 22(1), 31-72.

Chintagunta, P. K., & Dube, J. P. (2005). Estimating a stockkeeping-unit-level brand choice model that combines household panel data and store data. Journal of Marketing Research, 42(3), 368-379.

Chen, X., John, G., Hays, J. M., Hill, A. V., & Geurs, S. E. (2009). Learning from a service guarantee quasi experiment. Journal of marketing research, 46(5), 584-596.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. Journal of marketing research, 43(3), 345-354.

Danaher, B., Dhanasobhon, S., Smith, M. D., & Telang, R. (2010). Converting pirates without cannibalizing purchasers: the impact of digital distribution on physical sales and internet piracy. Marketing science, 29(6), 1138-1151.

Danaher, P. J., & Dagger, T. S. (2013). Comparing the relative effectiveness of advertising channels: A case study of a multimedia blitz campaign. Journal of Marketing Research, 50(4), 517-534.

Datta, H., Foubert, B., & Van Heerde, H. J. (2015, April). The Challenge of Retaining Customers Acquired with Free Trials. Journal of Marketing Research., 52(2), 217-234

Dhar, T., & Baylis, K. (2011). Fast-food consumption and the ban on advertising targeting children: the Quebec experience. Journal of Marketing Research, 48(5), 799-813.

Dinner, I. M., Van Heerde, H. J., & Neslin, S. A. (2014). Driving online and offline sales: The cross-channel effects of traditional, online display, and paid search advertising. Journal of Marketing Research, 51(5), 527-545.

Ghose, A., & Yang, S. (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. Management Science, 55(10), 1605-1622.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research.Educational researcher, 3-8.

Hartmann, W. R., Nair, H. S., & Narayanan, S. (2011). Nonparametric Estimation of Marketing-Mix Effects Using a Regression Discontinuity Design. Marketing Science, 30(6), 107-109.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In The economics of new goods. University of Chicago Press. 207-248)

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In Annals of Economic and Social Measurement, NBER. 5(4), 475-492.

Heckman, J. (1997). Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. Journal of Human Resources, 441-462.

Heckman, J. (2000), “Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective” Quarterly Journal of Economics, 115, 45-97.

Heinrich, C., Maffioli, A., & Vazquez, G. (2010). *A primer for applying propensity-score matching*. Inter-American Development Bank.

Huang, Q., Nijs, V. R., Hansen, K., & Anderson, E. T. (2012). Wal-Mart's impact on supplier profits. Journal of Marketing Research, 49(2), 131-143.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. Journal of econometrics, 142(2), 615-635.

Jacob, R. T., Zhu, P., Somers, M. A., & Bloom, H. S. (2012). *A practical guide to regression discontinuity* (pp. 1-100). MDRC.

Kuksov, D., & Villas-Boas, J. M. (2008). Endogeneity and individual consumer choice. Journal of Marketing Research, 45(6), 702-714.

Larcker, D. F., & Rusticus, T. O. (2010). On the use of instrumental variables in accounting research. Journal of Accounting and Economics, 49(3), 186-205.

Luca, M. (2011). Reviews, reputation, and revenue: The case of Yelp.com. (September 16, 2011). Harvard Business School NOM Unit Working Paper, (12-016).

Mithas, S., Krishnan, M. S., & Fornell, C. (2005). Why do customer relationship management applications affect customer satisfaction?. Journal of Marketing, 69(4), 201-209.

Meyer, B. (1995), “Natural and Quasi-Experiments in Economics,” Journal of Business and Economic Statistics, 12, 151-162.

Nair, H., Dubé, J. P., & Chintagunta, P. (2005). Accounting for primary and secondary demand effects with aggregate data. Marketing Science, 24(3), 444-460.

Narayanan, S., & Kalyanam, K. (2015). Position Effects in Search Advertising and their Moderators: A Regression Discontinuity Approach. Marketing Science, 34(3), 388-407.

Petrin, A., & Train, K. (2010). A control function approach to endogeneity in consumer choice models. Journal of Marketing Research, 47(1), 3-13.

Reiss, P. C., & Wolak, F. A. (2007). Structural econometric modeling: Rationales and examples from industrial organization. *Handbook of econometrics*, *6*, 4277-4415.

Roberts, M. R., & Whited, T. M. (2012). Endogeneity in empirical corporate finance.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

Rossi, P. E., R. E. McCulloch, G. M. Allenby. 1996. The value of purchase history data in target marketing. Marketing Science, 15(4), 321–340.

Rossi, P. E. (2014). Invited Paper—Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications. Marketing Science, 33(5), 655-672.

Rubin, D., (1973a), “Matching to Remove Bias in Observational Studies”, *Biometrics*, 29, 159-183.

Rubin, D., (1973b), “The Use of Matched Sampling and Regression Adjustments to Remove Bias in Observational Studies”, *Biometrics*, 29, 185-203.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of educational Psychology, 66(5), 688.

Rubin, D. B. (2006). Matched sampling for causal effects. Cambridge University Press.

Rutz, O. J., Bucklin, R. E., & Sonnier, G. P. (2012). A latent instrumental variables approach to modeling keyword conversion in paid search advertising. Journal of Marketing Research, 49(3), 306-319.

Snider, C., & Williams, J. W. (2014). Barriers to entry in the airline industry: A multi-dimensional regression-discontinuity analysis of AIR-21. Review of Economics and Statistics, (0).

Srinivasan, S., & Hanssens, D. M. (2009). Marketing and firm value: Metrics, methods, findings, and future directions. Journal of Marketing research, 46(3), 293-312.

Thistlethwaite, D. L., & D. T. Campbell (1960). Regression Discontinuity Analysis: An Alternative to the Ex-post Facto Experiment. Journal of Educational Psychology 51(6), 309–317

Tsai, Y. L., Dev, C. S., & Chintagunta, P. (2015). What's In A Name? Assessing the impact of rebranding in the hospitality industry. Journal of Marketing Research, ahead of print.

Verbeek, M. (2008). A guide to modern econometrics. John Wiley & Sons.

Villas-Boas, J. M., & Winer, R. S. (1999). Endogeneity in brand choice models. Management Science, 45(10), 1324-1338.

Wangenheim, F. V., & Bayón, T. (2007). Behavioral consequences of overbooking service capacity. Journal of Marketing, 71(4), 36-47.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.

Wright, P. G. (1928). Tariff on animal and vegetable oils.

Yuan, J. (2008). Estimating the Efficiency Improvement of the Resource Allocation in the Yahoo! Keyword Auction. International Journal of Humanities and Social Science, 1(18)

APPENDIX

HOW TO TEST A CAUSAL RELATIONSHIP?

Introduction to the Rubin Causal Model

Since the early seventies, Rubin (1973a, b, 1974, 1977, and 1978) via a sequence of influential papers developed the current leading approach for analyzing causal effects in observational studies. Rubin argued that causal effects should be interpreted in terms of their potential outcomes. Specifically, according to the different levels of exposure of a unit to the treatment, a pair of outcomes should be defined for that unit. Therefore, the development of the models which test causal effects is based upon a pair of potential outcomes rather than just the observed outcome. The examination of a causal effect according to its potential outcomes has two distinct advantages (Wooldridge, 2010).

Initially, this formulation allows for unobserved heterogeneity in the treatment effects which helps researchers to theoretically and practically cope with endogeneity. Unobserved heterogeneity is the variation among the individual units that participate in a given sample and the primal root of endogeneity. Although there is no way to observe all the factors that cause variation in the sample, by assuming heterogeneous treatment effects one can identify a model which rules out at least the obvious sources of endogeneity. Additionally, within the frame of potential outcomes, the relevant parameters and assumptions can be determined without any reference to specific parametric models (Wooldridge, 2010). This gives great flexibility to the researcher to identify the pertinent model according to the problem in question and the data at hand.

Indexing Units and Indicating Treatment Status

Indexing units and indicating status is a crucial procedure for testing causal relationships since provides the researcher with substantial information regarding the structure of the data at hand and helps him to identify the relevant model according to the causal effect of interest. Through the use of the following imaginary example the reader should get an idea how to index variables by indicating treatment status. Subsequently, the story unfolds in conjunction with the description of the quasi-experimental methods in order to bridge the use of quasi-experimental techniques with marketing research problems according to theory of causal inference. The notation is not always consistent between the theory about how each method works and the practical example of how each method is applied. The differences in notation between theory and practical applications provide the reader with the idea about the formal theory and then help him to translate it into actionable research. In addition these differences save space and can be used as an exercise for becoming familiar with modeling since modeling skills are crucial for a successful econometric analysis. This appendix is focused on correcting for self-selection bias in binary treatment variables.

Research Scenario

Suppose that Wagenignen University would like to examine the effect of a promotional campaign on the sales of Better Leven, which are animal friendly products packed with a specific label, in the ten biggest Dutch cities using N observational units, indexed by $i = 1,\ldots,N$. In general, observational units can be plants, animals, consumers, firms, states or any other individual entity which is examined on potential outcomes according to its exposure on different levels of a treatment. For the example of the promotional campaign assume that individual stores of the two big supermarket chains, Lidl and Alber Heijn, represent the units of observation with N being the total number of these stores. The rationale is that stores of these supermarket chains can be found in every big city in the Netherlands. After indexing the observations, the indicator W_i should be used to indicate whether the store i implemented the promotional campaign, with $W_i = 1$ if yes and $W_i = 0$ if not.

The campaign was implemented in the stores of these supermarket chains in six out of the ten cities. The stores in the remaining four cities were running other campaigns for similar products and they were afraid of possible cannibalization. For the sake of completeness, suppose that Albert Heijn was running a campaign to make the Albert Heijn Excellence animal products more salient in Rotterdam because their market research about local consumer trends indicated that the citizens of this area prefer high quality meat. On the other hand, Lidl was running a discount promotion in Utrecht because there was excessive stock of animal products in their local stores, other than the Better Leven ones, which were supposed to be sold before their expiration date. So, Utrecht and Rotterdam should be excluded from the promotional campaign.

Formulating the Potential Outcomes

For store i, there are two potential outcomes, labelled as Y_{1i} and Y_{0i}. Regarding the specific example, the sales of Better Leven products of a particular store represent the observed outcome Y_i, of that store after implementing the promotion. The first (Y_{1i}) describes the resulted outcome of the store i, if the promotion was implemented to that store. On the other hand, the second (Y_{0i}) describes the resulted outcome for the same store if the promotion was not implemented. Store i is able to either implement or not the promotion, but it is not possible to do both concurrently, and thus only one of the two outcomes can be actually observed. The outcome which cannot be observed is defined as a counterfactual outcome. However, the major aim of causal inference is to provide insights about this outcome. The scope of the next sections then, is to introduce the most prominent methods for estimating the counterfactual as a function of the given data.

Following the above discussion, the potential outcomes regarding the sales of Better Leven products in store i, considering also the promotion station status (Yes/No) of that store, can be formulated as:

$$Y_i = Y_i(W_i) = \underbrace{Y_{0i}\cdot(1-W_i)+Y_{1i}\cdot W_i}_{A} = \underbrace{\begin{cases} Y_{0i} & if \quad W_i = 0 \\ Y_{1i} & if \quad W_i = 1 \end{cases}}_{B} \qquad \text{Equation (1)}$$

The potential outcomes are depended to the particular handling that would have made one of them to be observed. According to Eq. (1), if the store i decides to implement the campaign, the outcome Y_i becomes Y_{1i} because the indicator W_i equals 1. If the store i decides to not implement the campaign then the outcome Y_i becomes Y_{0i} because W_i equals 0. Specifically, the substitution of the value of W_i from part B into part A provides the observed outcome. This demarcation between the pair of potential outcomes (Y_{1i},Y_{0i}) and the observed outcome Y_i is the cornerstone of econometric analysis of treatment effects (Wooldridge, 2010).

Determine the Causal Effect and the Selection Bias

Performing the simple mathematical operations in part A of Eq. (1), the observed outcome Y_i, can also written as:

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i}) \cdot W_i \quad \text{Equation (1.1)}$$

The causal effect of promoting Better Leven products on the sales of individual stores of the two supermarket chains is the difference Y_{1i} - Y_{0i} (Wooldridge, 2010). Generally, there should be a unique distribution for Y_{1i} and Y_{0i} in population which means that the effect of promotion can be different for different stores. According to Eq. (1), there is no way to observe both potential outcomes. Hence, the researcher should examine the effect of promotion by differencing the average sales of the stores that run and did not run the promotion, where the latter is used as the counterfactual. This is the easiest way to remedy unobservability to some extent, given the control of other factors.

The difference in average sales regarding the population distributions of the two potential outcomes can be expressed by Eq.(2):

$$\mathrm{E}[Y_{1i} - Y_{0i}], \ \text{Equation (2)}$$

where E refers to the mathematical expectation of the difference in the population means of the two outcome variables. For a specific sample consider the formula $\frac{1}{n}\sum_{i=1}^{n}[Y_{1i} - Y_{0i}]$.

The potential outcomes are also related to treatment status. Particularly, the formal link of the compared average sales conditional on promotion status and the average causal effect is given by the equation (2.1):

$$\underbrace{\mathrm{E}[Y_i|W_i = 1] - \mathrm{E}[Y_i|W_i = 0]}_{\text{Observed differences in average sales}} = \underbrace{\mathrm{E}[Y_{1i}|W_i = 1] - \mathrm{E}[Y_{0i}|W_i = 1]}_{\text{Average treatment effect on the treated}} \quad \text{Equation (2.1)}$$

$$+ \underbrace{\mathrm{E}[Y_{0i}|W_i = 1] - \mathrm{E}[Y_{0i}|W_i = 0]}_{\text{Selection Bias}},$$

where $A\,|B$ indicates the conditional dependence of A to B.

The term $\mathrm{E}[Y_{1i}|W_i = 1] - \mathrm{E}[Y_{0i}|W_i = 1] = \mathrm{E}[Y_{1i} - Y_{0i}|W_i = 1]$ represents the average causal effect of promotion on the stores that implemented the promotion (ATT). This term examines the average difference between the sales of the stores which run the promotion, $\mathrm{E}[Y_{1i}|W_i = 1]$, and the counterfactual outcome of that stores had they not run promotion $\mathrm{E}[Y_{0i}|W_i = 1]$.

Eq. (2.1) expresses the fundamental identity of causal inference and for the further understanding of the reader can be also described with simple words:

Outcome for treated – outcome for untreated = [outcome for treated - outcome for treated if not treated]
+ [outcome for treated if not treated - outcome for untreated]

These words can be also translated in our example:

Outcome for the supermarket that run the promotion – outcome for the supermarket that did not run the promotion = [Outcome for the supermarket that run the promotion - outcome for the supermarket that run the promotion if did not run the promotion] + [outcome for the supermarket that run the promotion if did not run the promotion - outcome for the supermarket that did not run the promotion]

The observed difference in promotion status though, adds to this causal effect another term which is called selection bias, and it is directly related with the unobserved heterogeneity and therefore endogeneity. This term reflects the difference in average Y_{0i} between the stores that run and did not run the promotion. Selection bias refers to cases when differences in the results are due to differences in the groups and not so much because of the treatment.

For instance, there might be stores with lower sales for Better Leven products that need this promotion in any case. Better Leven products are positioned as high quality products which imply higher prices. Nevertheless, Lidl is a supermarket chain which is positioned as a low cost provider. So, the Lidl stores are more likely to have lower sales for Better Leven products than Albert Heijn stores which are known for their high quality products but also for their higher prices as compared with the other chains. In other words, Lidl stores are more likely than Albert Heijn stores to seek promotion, hence even if they would implement a promotion their Y_{0i} 's are expected to be worse, making the selection bias negative in this case.

Sometimes, the selection bias can be quite large, in absolute terms, that it completely hides a positive treatment effect. The major aim of econometric research is to remedy selection bias, and therefore provide insight about the causal effect of the treatment variable (W_i).

Random Assignment

Random assignment of W_i corrects the selection bias since it creates independence between the treatment and potential outcomes. This independence is expressed by Eq. (2.2):

$$\mathrm{E}[Y_i|W_i = 1] - \mathrm{E}[Y_i|W_i = 0] = \mathrm{E}[Y_{1i}|W_i = 1] - \mathrm{E}[Y_{0i}|W_i = 0]$$

$$= \mathrm{E}[Y_{1i}|W_i = 1] - \mathrm{E}[Y_{0i}|W_i = 1], \qquad \text{Equation (2.2)}$$

where the independence of Y_{0i} and W_i allows the researcher to exchange $\mathrm{E}[Y_{0i}|W_i = 1]$ with $\mathrm{E}[Y_{0i}|W_i = 0]$ in the second line of the equation. Given a randomized process of assigning the treatment, the result of Eq. (2.2) simplifies further to $\mathrm{E}[Y_{1i} - Y_{0i}]$.

The most intrinsic aspect of Rubin's method is the relationship between the assignment of the treatment and the potential outcomes. The ideal case for analyzing this relationship is when the treatment assignment is randomized. Randomization ensures that the treatment assignment and the outcomes are not affected by other variables that may confound the causal effect of the treatment on the outcome. In other words, randomization ensures that ceteris is paribus. Through the use of randomization the researcher can gain very nice estimators of the average treatment effects (ATT and ATE), i.e. the mean differences between the treatment and the control group (Wooldridge 2010). The major advantage of randomization is that the outcomes of the control group can be straightforwardly used as counterfactual outcomes because they are drawn from a random distribution.

Unconfoundedness, Assignment Mechanism and Estimation Procedures

The notion of Unconfoundedness was formally expressed by Rosenbaum and Rubin (1983) as :

$$W_i \perp\!\!\!\perp (Y_{0i}, Y_{1i}) \mid X_i, \text{ Assumption 1}$$

where $A \perp\!\!\!\perp B \mid C$ indicates conditional independence of A and B, given C that represents the vector of chosen covariates which are assumed to be associated with the outcomes . In other words, the treatment status is not depended to the potential outcomes given certain covariates. The major premise of unconfoundedness is based on the researcher's confidence that he has a sufficient amount of predictors for the outcome variable, such that after controlling for their effects he can get valid estimates of causal effects (Wooldridge, 2010). However, the inclusion of covariates is limited only by the available data, the resources needed to collect that data, the software that will be used for the analysis and last but not least the ingenuity of the researcher.

Nevertheless, just adding covariates is not the best strategy. There should be extensive research, specific to the causal effect of interest, before one decides which variables co-vary with the treatment and the potential outcomes. Since this assumption cannot be actually checked the researcher should make some statements about the probability function of the assignment mechanism in order to test how well this assumption is approximated.

Imbens and Rubin (2009) define the assignment mechanism as the conditional probability of receiving the treatment as a function of potential outcomes and observed covariates. Specifically, the probability of the Better Leven products in store i to be promoted can be stated as:

$$Pr_i(W_i|X_{ji}, Y_{0i}, Y_{1i}), \text{ Statement A}$$

where X_{ji} is the j^{th} covariate, indexed with j=1,...,K, of the store i . Covariate can be any variable that may help the researcher to explain heterogeneity in the sample. The city that store i is located, the supermarket chain that store i is part of, the time when store i implemented the promotion are some of the most obvious covariates that the researcher can think of.

Experimental Trials

Wooldridge (2010) distinguishes three alternative ways of defining assignment mechanisms. The first class refers to randomized experiments, where the probability of assignment does not vary with potential outcomes given a known function of certain covariates. So, the probability of the Better Leven products in store i to be promoted under a completely randomized experimental design can be stated as:

$$Pr_i \{ W_i \mid (Y_{0i}, Y_{1i}) \mid X_{ji} \} = \begin{cases} 1/\binom{N}{N_1} & if \ \sum_{i=1}^{N} W_i \\ 0 & otherwise \end{cases}, \text{ Statement B}$$

In Statement B each unit has an assignment probability Pr_i= N_1/ N .Otherwise the treatment is not assigned to that unit since assignment probability is 0. St. B is the most complete expression of the relationship between unconfoundedness and complete randomization, given by Imbens and Rubin (2009).

This ideal case reflects a completely randomized experiment where, in a population of N units, $N_1 < N$ units are randomly selected to be treated and the remaining $N_0 = N - N_1$ units are defined as the control group. The use of complete randomization has become more salient in social sciences the last decades, and is regarded as a justifiable procedure to allocate scarce resources in order to make causal claims. The estimation of treatment effects in randomized experiments is quite simple. The researcher can use the mean differences as explained before or linear regression after controlling for certain covariates. The latter is given by Eq.(3):

$$Y_i = a + \rho\, W_i + \sum_{j=1}^{K} \gamma_j X_{ji} + \varepsilon_i, \quad \text{Equation (3)}$$

where ρ is the parameter of the treatment variable W_i which is a dummy variable that takes the values of 0 and 1 according to status and represents the causal effect of the treatment variable on the outcome variable Y_i ; ρ equals the difference between the potential outcomes, $\rho_{reg} = (Y_{1i}\text{-}Y_{0i})$, for the observational unit i.

In addition, γ_j is the value of the j^{th} parameter among K covariates, $j = 1, \dots, K$, where $X_{ji}'s$ can be dummies or continuous variables; a is the typical constant of the regression but in the specific context represents also the mathematical expectation of Y_{0i}, and ε_i is the typical error term but in this context represents also the difference between Y_{0i} and its mathematical expectation given a certain number of defined covariates, $Y_{0i} - E\big[Y_{0i}|X_{ji}\big]$. If the treatment variable and the error term are correlated, then the causal effect suffers from selection bias. However, this is prevented by complete randomization. Considering the example of Better Leven products Eq. (3) can rewritten as:

$$Sales_i = a + \rho\, Promotion_i + \sum_{c=1}^{10} \gamma_{1c} City_{ci} + \sum_{s=1}^{2} \gamma_{2s} SupermarketChain_{si} + \varepsilon_i \quad \text{Equation (3.1)}$$

This regression does not include all the potential covariates. But, it is intended to give a good taste to the reader about the process of structuring the data and indexing observation to test a causal effect in a more elegant way rather than just differencing the means. Also, all the other methods that are being discussed in this paper are expanding the ordinary multivariate regression models, therefore one must become familiar with Eq. (3.1) before is been introduced to more advanced estimation techniques.

Moreover, multiple regression analysis is more liable to ceteris paribus condition because provides the researcher with the capability to control for many factors other than the treatment that may affect the dependent variable. So, if the researcher does not expect a large selection bias, which seriously compromises the effect of the treatment on the potential outcomes, and if there is substantial number of controls, then the simple multivariate regression model is suggested. The problem is that no researcher knows all the relevant controls in the model are included and in the absence of random assignment this uncertainty may seriously harm the inference of causality.

Although random assignment is a vital tool for evaluating causal effects, it is not always feasible to implement it. Not only is expensive and time consuming to administer it, a randomized experimental design must be created and implemented before the treatment (Heinrich et al., 2010). Additionally, completely randomize designs mandate programming skills to create the algorithm which will randomly assign the treatment and advanced machine learning skills to estimate the counterfactual.

HOW TO APPROXIMATE COMPLETE RANDOMIZATION

Propensity Score Matching

The second class of assignment mechanisms maintains the independence between assignment probabilities and potential outcomes. But, contrary to randomized experiments, the assignment probabilities are not assumed to be a known function of the covariates (Wooldridge; 2010). The most prominent methods used to estimate treatment effects under this class of assignment mechanisms are known as Propensity Score and Matching Algorithm.

The propensity score was defined by Rosenbaum and Rubin (1983) to be the probability of treatment assignment conditional on observed baseline covariates. Rosenbaum and Rubin proved that unconfoundedness still holds even after replacing covariates with the propensity score. For the Propensity Score method the probability of the Better Leven products in store i to be promoted can be sated as:

$$Pr_i \{ W_i \perp\!\!\!\perp (Y_{0i}, Y_{1i}) \mid e(X_{ji})\}, \text{Statement C}$$

which equals St. B.

Propensity score, is defined as $\nu_w(e)$ and equals the mathematical expectation of the outcome variable given the covariates. Actually, Rosenbaum and Rubin showed that $\nu_w(e)= \mathrm{E}[Y_i|W_i = w, e(X_{ji})= e]$. Wooldridge (2010) indicates that the combination of propensity score and unconfoundedness leads to the following relation, $\nu_w(e)= \mathrm{E}[Y_i(w)| e(X_{ji})= e]$.

$\nu_w(e)$ can be estimated via limited dependent variable models given that the treatment is a binary variable (Heinrich et al., 2010). The crucial component of determining propensity score is the specification of the selection model after defining the variables that determine the participation in the treatment. So, the probability of the Better Leven products in store i to be promoted can be estimated by the following participation model:

$$Pr(Promotion_i) = \mathrm{F}(\mathrm{a}+\textstyle\sum_{c=1}^{10} \gamma_{1c} City_{ci} + \sum_{s=1}^{2} \gamma_{2s} SupermarketChain_{si}), \text{ Equation (3.2)}$$

Eq. (3.2) describes the general form of a limited dependent variable model, which simply estimates the probability of Better Leven products in store i to be promoted as a function of the two aforementioned covariates. The dependent variable $Pr(Promotion_i)$ can take the values of zero and one. Reasons, such as the expected profitability in a specific city or the strategic positioning of the supermarket chain that store i operates within, are plausible determinants of the decision to conduct or not a promotion for Better Leven. $F(\cdot)$ indicates the flexibility of the functional form.

Since $Pr(Promotion_i) = promotion_i(e)$, according to Rosenbaum and Rubin (1983), the causal effect of promoting Better Leven products which is based on the propensity score can be defined as:

$$\rho_{prop} = \frac{1}{n}\sum_{i=1}^{n}\{[promotion_1(e\ (\textstyle\sum_{c=1}^{10} \gamma_{1c} City_{ci} + \sum_{s=1}^{2} \gamma_{2s} SupermarketChain_{si})] - [promotion_0(e\ (\sum_{c=1}^{10} \gamma_{1c} City_{ci} + \sum_{s=1}^{2} \gamma_{2s} SupermarketChain_{si})]\}, \text{ Equation (3.2.1)}$$

On the other hand, Matching Algorithm exploits information from the control group to identify what would have happened to the treated group in the absence of the treatment. Simply, these algorithms use the control group to estimate the counterfactual and thus remedy unobservability. By comparing how outcomes differ between treated relative to observationally similar non-treated units, it is possible to estimate the causal effect of the treatment. Matching procedures directly match treated with non-treated units who have similar characteristics, referring to similar values in covariates.

Abadie and Imbens (2006) explicitly explain how to match treated and non-treated units with replacing non-treated units that do not fit well with treated ones. Specifically, given a sample$\{(Y_i, X_{ji}, W_i)\}_{i=1}^{N}$, $\| X_k - X_i \|$ (where k and i represent two observations from the treatment and control group accordingly, and W may contain also the propensity score estimated from Eq. 3.2.1) is defined to be the nearest neighbourhood to i, which implies that l_1

(i) equals a non-negative integer k, for $k \in \{1,...,N\}$, if $W_k \neq W_i$ and $\| X_k - X_i \| = \min_{k: W_z \neq W_i} \| X_z - X_i \|$. Regarding a more general situation, let $l_m(i)$ be the index that satisfies $W_{l_m(i)} \neq W_i$ and that is the m^{th} nearest control unit to treatment unit i, so that $\sum_{l: W_l \neq W_i} 1\{ \| X_l - X_i \| \leq \| X_{l_m(i)} - X_i \| \} = m$, where $1\{\cdot\}$ is the indicator function which, according to the matching algorithm, equals one if the expression in the brackets is true and zero if not. If a control unit m cannot receive the value of one as compared with every n unit is the sample, then is dropped out as ineligible.

Alternatively, $l_m(i)$ is the index of the unit in the control group that is the m^{th} nearest to treatment unit i in terms of the distance measure according to the norm $\| \cdot \|$. Further, let $K_m(i) \subset \{1,...,N\}$ represent the total set of indices for the initial M matches for i: $K_m(i) = \{l_1(i),..., l_M(i)\}$.

The matching algorithm for the sales of store, which is explicitly described by Wooldridge (2010), it imputes the missing potential outcomes as averages of the outcomes for the matches, by defining the potential outcomes Y_{0i} and Y_{1i} as:

$$Sales_{0i} \begin{cases} Sales_i & if \quad Promotion_i = 0 \\ \frac{1}{M} \sum_{k \in K_m(i)} Sales_k & if \quad Promotion_i = 1 \end{cases}$$

Equation (3.3)

$$Sales_{1i} \begin{cases} \frac{1}{M} \sum_{k \in K_m(i)} Sales_k & if \quad Promotion_i = 0 \\ Sales_i & if \quad Promotion_i = 1 \end{cases},$$

Considering Eq. (3.3) the causal effect of promoting Better Leven products based on the matching algorithm can be defined as:

$$\rho_{match} = \frac{1}{n} \sum_{i=1}^{N} (Sales_{1i} - Sales_{0i}).$$

Matching algorithms and Propensity score are powerful estimators but in the context of causal inference they are mostly used in order to estimate the counterfactual outcome and to balance the sample rather than estimating the causal effect itself (Wooldridge, 2010). The major aim of using these techniques is to restructure the data in order to support regression analysis or any of the techniques which represent the third class of assignment mechanisms and will be discussed in the next subsections. Thanks to the tremendous rate of advancement in econometric techniques, given the skills of the researcher and the data at hand, the possibilities of mixing different methods are almost countless. The third class of assignment mechanisms refers to the adjustment of the functional form of the model according to the data at hand and the causal effect in question. Simply, these techniques encompass the data generating process in order to correct for selection bias. However, in quasi experimental research the outcomes are not totally independent from the assignment mechanism. The most salient quasi-experimental methods are IV, RD, and DID.

Instrumental Variables

IV methods are in the vanguard of econometric tools which aim to solve the problem of endogeneity. The challenge in IV is to find an instrument, which is an exogenous variable correlated with the treatment variable, but uncorrelated with the error term and it indirectly affects the potential outcomes. This particular technique uses the assumption of unconfoundedness, but the conditional independence lies within the instrument and the potential outcomes. So, Eq. (1) is alternated to:

$$W_i = W_{0i} \cdot (1 - Z_i) + W_{1i} \cdot Z_i = \begin{cases} W_{0i} & if \;\; Z_i = 0 \\ W_{1i} & if \;\; Z_i = 1 \end{cases}, \quad \text{Equation (4)}$$

where Z_i is the instrument which directly affects the value of the treatment.

According to Wooldridge (2010), the exogeneity of the instrument can be expressed by modulating Ass.(1) as:

$$Z_i \perp (Y_{0i}, Y_{1i}, W_{0i}, W_{1i}), \quad \text{Assumption 2}$$

where $A \perp B$ means that all the elements which belong to the vector B are independent from A. Ass.(2) contains two properties of the instrument. Initially, $\perp$ indicates the conditional independence of the instrument and the potential outcomes and ensures its random assignment. Second, the absence of z in the definition of Y_w notifies the indirect effect of the instrument on the potential outcomes (Wooldridge, 2010).

In addition to these two properties, Imbens and Angrist (1994) have introduced the concept of compliance as a vital prerequisite for the appropriate estimation of the causal effect in IV settings. Compliers are the units who are complied with their assignment to the treatment. In other words, these units always accept the treatment status that is assigned to them without further reaction. Imbens and Angrist (1994) present four categories of observational units according to their reaction about the treatment they are assigned to. Explicitly:

$$Q_i = \begin{cases} never - taker & if \;\; W_{0i} = W_{1i} = 0\,, \\ complier & if \;\; W_{0i} = 0,\, W_{1i} = 1, \\ defier & if \;\; W_{0i} = 1,\, W_{1i} = 0\,, \\ always - taker & if \;\; W_{0i} = W_{1i} = 1\;. \end{cases} \quad \text{Equation (4.1)}$$

where Q_i indicates the category that the observational unit belongs according to response on the treatment status. Particularly, never-takers and always takers are the units whose outcomes are unaffected the assignment mechanism since the former will never take the treatment even if they are assigned to it and the latter will always take the treatment even if they are assigned to control group. Defiers always will try to get in the opposite group than that they will be assigned. Therefore, except complying, any other reaction is considered inappropriate for the IV estimator since the assignment is not randomized. Hence, in the IV case Eq.(2) becomes:

$$\mathrm{E}[Y_{1i} - Y_{0i} | Q_i = Complier] \quad \text{Equation (4.2)}$$

The estimation of the causal effect in IV is implemented by the Two-Stage Least Squares (2SLS) model. This is a two stage regression. The first stage checks the effect of the instrument on the treatment variable. The second stage uses the fitted values of the treatment variable, which are derived in the first stage, as the independent variable with the outcome variable to be the dependent one. This implies that the effect of the instrument on the potential outcomes is captured in the fitted values of the treatment variable. It is helpful for the reader to imagine IV as a chain reaction which starts from the effect of Z_i on W_i and continues with the effect of the fitted values of W_i ($\widehat{W_i}$) on Y_i. Nevertheless, before proceeding in the second stage the link between the instrumental variable and the outcome variable should be examined by another regression which is called the reduced form equation.

Concerning the example of Better Leven, imagine that Wageningen University was funded from the Dutch Government in order examine the causal effect of promotion on store sales of these products. Wageningen conducted a lottery for distributing these funds in certain stores according to that budget. Hence, Wageningen Lottery will be the instrument which affects the randomization of promotion of Better Leven in specific stores.

The 2SLS first stage regression is described by Eq.(4.3.1):

$$Promotion_i = a_1 + \varphi WageningenLottery_i + \varepsilon_{1i}, \text{ Equation (4.3.1)}$$

where φ is the effect of $WageningenLottery$ status (receive or not funds for promotion) on $Promotion$ status (implement not implement). This coefficient should be significant and relatively large. Otherwise the instrument is weak or inappropriate.

Then, the first stage fitted values of promotion can be defined as:

$$\widehat{Promotion}_i = a_1 + \varphi WageningenLottery_i, \text{ Equation (4.3.2)}$$

The reduced form equation for checking the link between $WageningenLottery$ and $Sales$ will be of the reduced form:

$$Sales_i = a_0 + \rho WageningenLottery_i + \varepsilon_{0i}, \text{ Equation (4.3.3)}$$

where ρ is the effect of $WageningenLottery$ status on the sales of store i.

Further, the 2SLS second stage equation will be of the form:

$$Sales_i = a_2 + \lambda_{2SLS}\widehat{Promotion}_i + \varepsilon_{2i}, \text{ Equation (4.3.4)}$$

where λ_{2SLS} is the causal effect of the fitted values of promotion status, which captures the effect of the lottery status, on the sales of store i. Moreover, it can be shown that $\lambda = \frac{\rho}{\varphi}$, see Angrist and Pischke (2015). It has to be stated that the second stage equation can be a logit model.

Regression Discontinuity

The rationale behind RD is that the treatment is assigned, either partly or completely, by the value of an exogenous continuous variable D_i, which is called running or forcing variable, being on either side of a common threshold (Wooldridge, 2010; Angrist and Pischke, 2015). Therefore, a discontinuity is created in the conditional probability of receiving a treatment as a function of D_i. This variable is smoothly related with the potential outcomes, which implies linearity. Any observed discontinuity, or jump according to graphical representation of the data, of the conditional distribution of the outcome variable as a function of this covariate at the threshold signals a causal effect of the treatment variable.

In RD design, the assignment of the treatment status, W_i, is assumed to be a deterministic function of D_i:

$$W_i = 1[D_i \geq r], \text{ Assumption 4}$$

where c is the threshold and 1[·] notifies the indicator function which equals on if the even in brackets is true and zero otherwise. Observational units with D_i value equal or bigger than r are classified within the treatment group and the rest are classified within the control group. Deterministic means that once the researcher knows the value of r, he also knows the status D_i. The critical property of RD is that the outcome values which correspond to the values of the running variable below the threshold can be used as a valid counterfactual for the outcome values equal and above this point. The premise is that by including the running variable in the model the researcher takes account for the data generating process of the treatment. Since the values of the running variable are not controlled by the researcher, but directly affect the status of the treatment, then the assignment of the treatment can be conceived as randomized.

Now, recall that a serious reason for supermarket stores to conduct a promotion for Better Leven is the estimated willingness to pay for those products according to the region that the supermarket stores operate within. For example, if the willingness to pay of the citizens of a certain city exceeds a certain threshold, which can be easily defined as the break-even point after defining the unit costs, then the stores of this city implement the promotion since they expect an increase in their profit margins. For simplicity, assume a common price threshold for both Lidl and Albert Heijn and that these supermarkets use the same logistics system which results in the same cost structure. Special price deals between each supermarket chain and suppliers are possible with respect to ordered quantities, but they are not considered in this example.

So, willingness to pay is the running variable to be included in the RD model to test the causal effect of promoting Better Leven on Sales. In addition, the sales of the stores which are located in the cities with estimated willingness to pay lower than the threshold can be used as a counterfactual for the sales of the supermarket stores located in the cities with estimated willingness to pay equal or above the threshold. Then, the local linear regression RD model can be identified as:

$$\overline{Sales}_{WTP_{ci}} = a + \rho\ Promotion_{WTP_{ci}} + \ d_1 WTP_c + d_2 {WTP_c}^2 + \sum_{c=1}^{10} \gamma_c City_{ciWTP} + \theta_c (\sum_{c=1}^{10} \gamma_c City_{ciWTP} \times WTP) + \varepsilon_{WTP}\ , \quad \text{Equation (5)}$$

where $\overline{Sales}_{WTP_{ci}}$ represents the sales rate of Better Leven in store i indexed according to the estimated willingness to pay of the city c (WTP can be defined as an interval counting forward and backwards from the threshold r) that the store is located, $Promotion_{WTP_{ci}}$ is the treatment dummy for every store defined by willingness to pay of the city that the store is located ,the parameter d_1 captures the linear control for $\overline{Sales}_{WTP_c}$according to willingness to pay of each city c , d_2 checks for the rate of increase in WTP and if this parameter is significant might harm the assumption of the smooth association between the running variable and the outcome, θ_c captures the composite effect of the interaction term $\sum_{c=1}^{10} \gamma_c City_{cWTP} \times WTP$ on the sales of store i, and ρ as always represents the causal effect of promotion on sales.

Considering the deterministic nature of the running variable the causal effect of promoting Better Leven in the sales of store i can be expressed as:

$$\mathrm{E}[Sales_{1i} - Sales_{0i} | WTP = r], \text{ Equation (5.1)}$$

The results of RD can be violated if the stores that did not run the promotion and are located in the cities with just lower willingness to pay than that of the threshold differ in some way (other than the treatment) with the stores that run the promotion in the cities with just higher willingness to pay. For example the consumers who buy certain products from Albert Heijn stores they are expected to have higher willingness to pay than those who buy these products from Lidl regardless the city that these people live.

Differences-in-Differences

The intrinsic characteristic of DID is that it takes into account the differences between the treatment and control group given the absence of random assignment. The method assumes that treatment and control groups follow the same pattern, with respect to time, in their outcome values before and after the treatment. In other words, the distribution of outcomes of the treatment groups should be similar to that of the control group before and after the treatment. Then, the divergence of the post treatment path of the treatment group as compared with that of control group may indicate a causal effect (Angrist and Pischke, 2015).

DID is a quasi-experimental econometric technique that utilizes the time and cohort dimensions of data to control for unobserved-but-fixed covariates. This method is based on comparison in pre-treatment and post treatment outcome levels, and it is valid under the assumption that the counterfactual trend behavior of treatment and control groups is equal. The particular examination requires panel data which refers to repeated observations on the same units. Alternatively, one can say that panel data combines time series and observational data (Angrist and Pischke, 2015).

According to Wooldridge (2010), in the standard DID model the unit i belongs to group $G_i \in \{0,1\}$, where 1 indicates the treatment group and 0 the control group, and is observed in time period $T_t \in \{0,1\}$, where 1 indicates the post-treatment period and 0 the pre-treatment period. Moreover, for the random sample with N observations, $i = 1,\ldots,N$, the group status and the observational time of i can be operationalized as random variables.

Table 1

	TREATMENT	CONTROL
BEFORE	A	B
AFTER	C	D

Considering Table 1, the causal effect of the treatment can be estimated as (C-D)-(A-B). The formal expression of the DID causal effect can be written as follows:

$$\tau_{DID} = \mathrm{E}[Y_{1i} - Y_{0i}] = (\mathrm{E}[Y_i \mid G_i = 1, T_t = 1] - \mathrm{E}[Y_i \mid G_i = 1, T_t = 0]) - (\mathrm{E}[Y_i \mid G_i = 0, T_t = 1] - \mathrm{E}[Y_i \mid G_i = 0, T_t = 0]),$$

Equation (6)

Eq.(6) subtracts the difference in the population means over time in the control group ($G_i = 0$) from the difference in the population means over time in the treatment group ($G_i = 1$). The aim of this subtraction is to remove the biases related to time trends and fixed effects.

DID also uses regression analysis, but combines cross-sectional and time-series estimators, to estimate the causal effect of interest. The standard DID regression mole is written as:

$$Y_{it} = t_{DID}(G_i \times T_{it}) + \nu G_i + \mu T_t + \varepsilon_{it} \text{ , Equation (6.1)}$$

where Y_{it} denote the number of units according to their group status i at time t , t_{DID} is the causal parameter of the DID model as explained by Eq.(6) and $G_i \times T_{it}$ is the interaction term which is created by multiplying the two dummies which indicates observations with treatment status in the post-treatment period, G_i is a dummy for the treatment group which controls for the fixed differences between the observational units, moreover T_t which is the dummy of the post-treatment period and μ is the parameter which controls for time effects regarding the fact that things change over time no matter the presence or the absence of the treatment--the major aim of DID is to cancel out the time trends and fixed-effects of group characteristics via Eq.(6) and thus to eliminate selection bias at the cost of attaining information of these factors (see Wooldridge, 2010 and Angrist and Pischke 2008 for a detailed description about the elimination of the parameters ν and μ in eq. 6.1), and finally ε_{it} is the error term of the grouped units observed at period t.

Now remember that Wageningen University wants to particularly examine the causal effect of promoting Better Leven on the average weekly sales of these products. Moreover, remember that in some cities this promotion was not implemented due to danger of cannibalizing sales of other promoted products. Concerning the DID approach, serial weekly data of sales of the supermarket stores which run and did not run the promotion is needed. So, the variables should be indexed according to their implementation of the promotion and time. Therefore, the DID multistate regression model can be identified as follows:

$$Sales_{ct} = a + t_{DID} Promotion_{ct} + \sum_{t=1}^{3} \mu_c Week_{ct} + \sum_{c=1}^{10} \nu_c City_{sc} + \sum_{c=1}^{10} \theta_c (City_{cs} \times t) + \varepsilon_{ct}, \quad \text{Equation (6.2)}$$

where $Sales_{ct}$ denote the sales of the stores conducted a promotion in city c at week t. According to Angrist and Pischke (2014) the interaction term in Eq.(6.1) can be simplified by introducing a simple measure of exposure to the treatment, so $Promotion_{ct}$ measures the proportion of stores that promoted Better Leven in city c and week t. Moreover, the s^{th} city dummy equals one when the store is form city c, meaning $s = c$, and ν_c are the coefficients of the city dummies. The time effects μ_c are similarly coefficients of the week dummies, $Week_{ct}$.

In addition, samples that are constituted by many periods and groups allow for relaxation of the common trends assumption by introducing a degree of a nonparallel evolutionary time path in outcomes between the groups in absence of the treatment effect (Angrist and Pischke, 2015). So, $\sum_{c=1}^{10} \theta_c (City_{cs} \times t)$ controls for city-specific trends. Specifically, the model in Eq.(6.2) implies that in the absence of a promotion effect, sales in city c deviate from common week effects by following the linear trend captured by θ_c.

References

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1), 235-267.

Busse, M., Silva-Risso, J., & Zettelmeyer, F. (2006). $1,000 cash back: The pass-through of auto manufacturer promotions. *The American Economic Review*, 1253-1270.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 315-331.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, 65(2), 261-294.

Heinrich, C., Maffioli, A., & Vazquez, G. (2010). A primer for applying propensity-score matching. *Inter-American Development Bank*.

Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467-475.

Imbens, G. W., & Rubin, D. B. (2015). Causal Inference in Statistics, Social, and Biomedical Sciences. *Cambridge University Press*.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

Rubin, D., (1973a), “Matching to Remove Bias in Observational Studies”, *Biometrics*, 29, 159-183.

Rubin, D., (1973b), “The Use of Matched Sampling and Regression Adjustments to Remove Bias in Observational Studies”, *Biometrics*, 29, 185-203.

Rubin, D. (1974), “Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies,”*Journal of Educational Psychology*, 66, 688-701.

Rubin, D., (1977), “Assignment to Treatment Group on the Basis of a Covariate,”*Journal of Educational Statistics*, 2(1), 1-26.

Rubin, D. B., (1978), “Bayesian inference for causal effects: The Role of Randomization”, *Annals of Statistics*, 6:34–58.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.