Industrial Marketing Management 127 (2025) 14–28

Contents lists available at ScienceDirect

Industrial Marketing Management

journal homepage: www.elsevier.com/locate/indmarman

Propensity score modeling for business marketing research

Peter Guenther[a,*], Miriam Guenther[a], Shekhar Misra[b], Mariia Koval[c], Ghasem Zaefarian[d]

[a] University of Liverpool Management School, Liverpool, UK
[b] J.E. Cairnes School of Business & Economics, University of Galway, Galway, Ireland
[c] IESEG School of Management, Lille, France
[d] Leeds University Business School, Leeds, UK

ARTICLE INFO

Keywords:
Propensity score modeling
Review
Results assessment
Guidelines

ABSTRACT

Propensity score modeling (PSM) is a powerful statistical technique that, in the appropriate data contexts, addresses biases from confounding and selection, which can otherwise distort results and lead to erroneous inferences. However, while the number of PSM applications in business marketing research is growing, many studies mistakenly assume that PSM is a universal solution for all endogeneity issues. Often, studies lack sufficient detail about the specific endogeneity problem they aim to address, which is a critical issue, as PSM is appropriate only for certain types of endogeneity. Additionally, essential tests to confirm the validity and robustness of PSM results are frequently overlooked or insufficiently reported, raising concerns about the reliability of findings. This article aims to enhance the rigor of PSM applications in business marketing research by offering updated practical guidance on its appropriate use, key aspects to report, and common misconceptions and errors to avoid. A practical example of PSM implementation in Stata is included, along with a comprehensive checklist of justifications and best practices to guide business marketing researchers in their future PSM-based studies.

1. Introduction

Addressing endogeneity is crucial in business marketing research to ensure robust findings and meaningful insights. Endogeneity arises when explanatory variables and the error term in a predictive model are correlated, introducing bias and undermining the validity of research findings (Hill et al., 2020; Li, 2012; Wooldridge, 2010). This bias can lead to effect estimates that substantially under- or overestimate the true effects, and in some cases, even reverse the inferred direction of effects. Recognizing this issue, business marketing researchers increasingly employ advanced methods to safeguard against endogeneity, thereby enhancing the credibility of their results.

Zaefarian et al. (2017) provide a comprehensive overview of contemporary methods for addressing endogeneity problems. Importantly, Ullah et al. (2018) outline a detailed step-by-step guide for the generalized method of moments and the instrumental variable approach to address endogeneity in business marketing research. Among the statistical techniques discussed by Zaefarian et al. (2017), propensity score modeling (PSM) stands out as a powerful tool for researchers seeking to tackle endogeneity stemming from selection bias that is often present in observational studies. Specifically, the bias stems from confounded selection of treatment, where units (e.g., firms) with certain characteristics chose treatment while others do not. PSM effectively neutralizes the impact of confounding variables by creating comparable treatment and control groups, thereby enabling more accurate estimation of causal relationships (Dehejia & Wahba, 2002; Li, 2012; Rosenbaum & Rubin, 1983; Rubin, 2005; Shipman et al., 2017).

Despite its potential, PSM remains relatively underutilized in business marketing research, although notable applications exist. For instance, PSM has been used to examine the profitability of suppliers following a major distributor's market entry (Huang et al., 2012), shifts in firms' R&D investments in emerging economies facing competition from unregulated/unregistered entities (Heredia Pérez et al., 2018), returns on solution offerings (Restuccia & Legoux, 2019), advantages of direct customer value propositions (Mishra et al., 2020), and relationships with financial service firms that lead to investor relations awards (Cheng et al., 2021).

However, while the number of PSM applications in business marketing has been growing, especially in recent years, many of these applications fall short of best practices outlined in the PSM literature.

* Corresponding author.
E-mail addresses: Peter.Guenther@liverpool.ac.uk (P. Guenther), Miriam.Guenther@liverpool.ac.uk (M. Guenther), Shekhar.Misra@universityofgalway.ie (S. Misra), M.Koval@ieseg.fr (M. Koval), G.Zaefarian@leeds.ac.uk (G. Zaefarian).

https://doi.org/10.1016/j.indmarman.2025.03.006
Received 21 June 2024; Received in revised form 4 February 2025; Accepted 19 March 2025
Available online 27 March 2025

Essential methodological details—such as the specific endogeneity issue being addressed, conditioning approach, covariate balance, and common support—are frequently omitted or insufficiently reported, raising concerns about result reliability. With critical PSM literature dispersed across numerous journals, researchers may find it challenging to stay informed about methodological developments. Consequently, with few exceptions (e.g., Golovko et al., 2022), many studies lack sufficient justification for using PSM to address specific endogeneity problems, often treating it as a universal remedy for endogeneity. Yet, different endogeneity issues require distinct methodologies, and it is important to recognize that PSM is not a panacea for all types of endogeneity problems (Certo et al., 2016; Hill et al., 2020; Shipman et al., 2017; Zaefarian et al., 2017). Authors need to consider the different possible endogeneity causes and provide conceptual justification that the cause at hand is one that PSM handles well. Moreover, even fewer studies explain how their analysis adheres to the main principles of PSM regarding the chosen conditioning approach, covariate balance, and common support, raising questions about the robustness of their findings. This backdrop highlights the urgent need for comprehensive guidance—including a step-by-step guide—on how to apply PSM convincingly.

This article seeks to demystify PSM for business marketing researchers by providing updated, tailored guidance on its appropriate use, what aspects to report, and common misconceptions or errors to avoid. An application example, focusing on the payoff to acquisitions in business markets, offers hands-on guidance on implementing PSM in Stata. Additionally, the article includes a concise checklist to support future PSM applications by business marketing researchers.

From a methodological perspective, this article contributes to ongoing efforts to enhance rigor in business marketing research, responding to recent critiques by business marketing researchers that "the methodological approaches that have been used to study B2B [phenomena] have fallen short of the sophistication and rigor necessary to address … many unanswered questions" (Swani et al., 2020, p. 589). By providing clear and current guidance on the essential assumptions and steps for effective PSM application, this study contributes to enhancing methodological rigor in business marketing research. It complements the stream of business marketing studies that develop step-by-step guides for addressing endogeneity in observational studies (e.g., Ullah et al., 2018), filling the gap in the literature that has not covered the PSM approach. Moreover, we clarify when PSM is appropriate given the different types of possible endogeneity causes in business marketing research. Furthermore, we discuss how PSM can be augmented to address endogeneity more broadly than in the standard approach. By following our comprehensive guide, researchers can navigate the intricacies of endogeneity, ensuring their analyses yield more reliable results and, hence, actionable insights.

This guide aims to enhance understanding of PSM's fundamental concepts and demonstrate its application in the B2B context, equipping researchers with the necessary knowledge and tools to elevate the rigor of their studies. By adhering to PSM's main principles and following the checklist provided, researchers can effectively implement PSM to enhance the validity of causal inferences in their observational studies.

In the following sections, we provide the methodological background for PSM, encompassing the counterfactual framework, endogeneity, and appropriate data scenarios for PSM use. We then offer tailored guidance on justifying and employing PSM effectively, comparing our recommendations with prior PSM applications in business marketing to highlight areas requiring special attention. We present a checklist of key considerations for researchers embarking on future PSM-based projects and illustrate essential steps using an example business marketing dataset. We conclude with a general discussion and recommendations for further PSM-related research.

2. Methodological background

Business marketing researchers often rely on observational data to infer causal relationships of interest. However, making valid causal inferences from such data can be challenging due to various data-related issues. Addressing these issues requires the use of appropriate methods, such as PSM. In this section, we outline the methodological backdrop for PSM, beginning with the counterfactual framework, which represents the ideal condition for unbiased causal inferences. Next, we examine different causes of endogeneity, a topic that received significant attention in recent business marketing literature (e.g., Zaefarian et al., 2017). It is essential for researchers to recognize that endogeneity can have multiple causes and no single method, including PSM, can address all of them. Finally, we introduce the PSM approach and clarify the specific endogeneity issues it is best suited to address.

2.1. The counterfactual framework

The motivation of using PSM is to increase a researcher's confidence that a variable's identified effect on an outcome is not an artefact of unaccounted data characteristics and truly exists in the population (i.e., is causal) (Li, 2012; Rosenbaum & Rubin, 1983; Shipman et al., 2017). For example, business marketing researchers may want to determine whether the use of artificial intelligence (AI) helps B2B firms to become more successful, as reflected in their financial performance. As a thought experiment, the causal effect would be identified—in the sense of a perfect counterfactual—if the very same firm could be observed introducing AI with outcome Y_1 and not introducing AI with outcome Y_0 at the same point in time. In this scenario, the effect of the variable of interest (i.e., the AI introduction) could simply be calculated as $Y_1 - Y_0$. However, the perfect counterfactual does not exist, which means that we can always only observe one of the two possible outcomes for a given unit (Rubin, 2005), which is the well-known "fundamental problem of causal inference" (Holland, 1986).

The basic idea of the PSM approach can be derived from considering the gold standard of causal inference, which is the controlled laboratory experiment (Li, 2012). Experiments typically compare the outcomes of a treatment group (e.g., exposed to a stimuli) and control group (not exposed). The control group provides an estimate of the counterfactual outcome, which is not the perfect counterfactual but is considered close to perfect when participants are allocated randomly to the two groups. Randomization is to ensure that participants in the two groups are similar with regard to potentially consequential characteristics (e.g., personality or mood), so that these characteristics do not asymmetrically affect the (mean) outcomes in the two groups and hence bias the effect estimate of treatment (Li, 2012; Shipman et al., 2017). This idea of estimating an unbiased treatment effect through balancing of characteristics in the treatment and control groups is the crux of the PSM approach (Dehejia & Wahba, 2002).

2.2. Endogeneity causes

Against the backdrop of recent methodological articles in business marketing research (Zaefarian et al., 2017), the identification challenge described by the counterfactual framework can also be understood as an endogeneity issue (Li, 2012). However, business marketing researchers must be mindful that different causes of endogeneity exist (Certo et al., 2016), and that "there is no generic way to address every possible cause of endogeneity" (Hill et al., 2020, p. 118). A concern is that many PSM studies in business marketing do not explicitly explain the type of endogeneity they aim to address using PSM. PSM is just one tool in the methods toolbox for addressing endogeneity, and each method is suited to certain endogeneity issues while being unsuitable for others (Certo et al., 2016; Hill et al., 2020; Zaefarian et al., 2017). To support the clarity of business marketing studies using PSM, we briefly review the main causes of endogeneity and clarify which cause PSM can address.

Formally, endogeneity occurs when an independent variable is correlated with the error term of a model. This correlation violates the standard statistical assumption that the error term is uncorrelated with the model's independent variables. Violations of this assumption lead to bias of unknown directionality (up or down) and magnitude in the effect estimates, creating the risk of fundamentally wrong inferences (Wooldridge, 2010).

Four different causes of endogeneity can be distinguished, namely, omitted variables, simultaneity, measurement error, and selection (into sample or of treatment) (Hill et al., 2020; Wooldridge, 2010). Omitted variables create endogeneity to the extent that they drive both the independent variable of interest (X) and the dependent variable (Y). As they are omitted, they are part of Y's unexplained residual (i.e., the error term). Their correlation with X means that X is correlated with the error term, creating an endogeneity issue. Simultaneity occurs when X affects Y, but Y also affects X at the same time. As the residual is a part of Y, the effect of Y on X creates a correlation between X and the residual (i.e., the error term), leading to endogeneity. Measurement error in X creates endogeneity when it is also related to Y and hence becomes part of Y's residual (i.e., the error term), causing a correlation between X and the error term. Finally, selection leads to endogeneity when the selection process restricts the range of values observed for Y (selection into sample) or when an unmeasured factor determines both Y and the selection of X (selection of treatment). A selection into sample scenario occurs, for instance, when certain eligibility criteria based on Y (e.g., a minimum size/revenue requirement to become a publicly listed company) prevent the observation of Y (and its residuals) for certain units (e.g., revenue data cannot be observed for private startup firms). When X drives Y, the truncation of Y and its residuals (e.g., small or large residuals are systematically truncated) also depends on X, effectively creating a correlation between X and the remaining residuals (i.e., the error term). A selection of treatment scenario occurs when an unaccounted factor drives a unit's selection of the level of X (i.e., the treatment) and the outcome Y (i.e., its unexplained residual), causing a correlation between X and the error term (Certo et al., 2016). The PSM approach is mainly used to address endogeneity caused by selection of treatment.

2.3. Endogeneity caused by observable or unobservable factors

Regarding the appropriate use of PSM, an additional important distinction is whether the factors, which are related to the selection of treatment and cause endogeneity, are observable or unobservable to the researcher (Li & Prabhala, 2008). Observable factors are variables that can be, and have been measured, by the researcher. In contrast, unobservable factors are variables that have not been measured and are usually very difficult or even impossible to observe. The distinction between observable and unobservable underlying factors is relevant for the first and last endogeneity causes discussed above—omitted variables and selection—while measurement error and simultaneity are usually considered unobservable (Li & Prabhala, 2008; Wooldridge, 2010). Considering this distinction results in eight types of endogeneity causes: observable omitted variables, unobservable omitted variables, measurement error, simultaneity, observable selection into sample, unobservable selection into sample, observable selection of treatment, and unobservable selection of treatment.

PSM is traditionally used to address observable selection of treatment (Rosenbaum & Rubin, 1983; Rubin, 2005).[1] Nevertheless, it can be augmented to also capture effects from unobservable underlying factors causing the (level of) treatment, as we discuss later in this paper.

[1] Different terms are used in the literature to describe the assumed selection of treatment based on observable factors, including 'selection on observables,' 'conditional independence,' 'ignorable treatment assignment,' and the 'unconfoundedness assumption.'

2.4. How does PSM work and when is it appropriate?

In essence, PSM redresses imbalance in a set of characteristics between the treatment and control groups, as this imbalance—caused by selection of treatment on observable factors—can lead to endogeneity. Specifically, PSM matches to the treatment group observations selected control group observations that are similar on multiple observed characteristics (Hill et al., 2020; Li, 2012; Shipman et al., 2017; Wooldridge, 2010). The PSM approach reduces this multidimensional matching problem to a single dimension with the help of the propensity score using a two-step procedure (Li, 2012; Rosenbaum & Rubin, 1983; Shipman et al., 2017). In the first step, the probability of treatment is modelled, using the characteristics as explanatory variables for treatment (e.g., in a logit or probit model). Predicted values of the probability of treatment (i.e., conditional on the considered characteristics) are then used to derive propensity scores for each of the treatment and control group observations in the dataset. Despite being an estimated score, research demonstrates that using it for matching can improve statistical efficiency in common PSM contexts (Abadie & Imbens, 2016). In the second step, observations are matched so that the treatment group and matched control group observations have similar propensity scores, effectively addressing any imbalances in the considered characteristics between these groups. The estimated effects are thus based on comparing outcomes only between treatment and control firms with very similar characteristics, rather than those with differing characteristics. If no sufficiently similar comparison can be identified in the data, corresponding observations are discarded, which can reduce the generalizability of findings into these specific areas of the data (Shipman et al., 2017). At the same time, effects would be biased if observations without suitable counterfactuals were to be used in the estimation.

PSM can be a powerful tool for business marketing researchers to derive robust inferences in certain data scenarios. Like for any other method, researchers nevertheless need to verify that the approach is the most suitable, given available alternatives and considerations about the data at hand and phenomenon that researchers intend to draw inferences about. In the following, we provide guidance for researchers to determine the general appropriateness of PSM, along two key questions that draw on the important points regarding endogeneity causes discussed above. In addition, we created a decision tree (Fig. 1) to help researchers verify their choice of PSM against alternative methods. The discussion of the alternative methods is beyond the scope of this paper, as our focus is on PSM, including guidance on key decision parameters, which remain unaddressed in general articles on endogeneity and methods. Our study therefore constitutes an important complement to prior work in business marketing research that discussed endogeneity and remedies more broadly (e.g., Zaefarian et al., 2017).

What is the nature of endogeneity? Business marketing researchers using PSM must ensure that all consequential characteristics are considered and, therefore, are observable factors; this is commonly referred to as the strongly ignorable assumption (Li, 2012; Rosenbaum & Rubin, 1983; Shipman et al., 2017). This requirement may seem challenging in the typical data context of PSM (i.e., secondary data downloaded from data syndicators), where researchers effectively have no control over the variables that have been collected. However, not all missing characteristics automatically create an endogeneity issue and are a potential source of bias. Recall that for endogeneity to occur, the missing characteristics must simultaneously affect the selection decision (i.e., the treatment) and the outcome (e.g., financial performance). Against this backdrop, a PSM approach is appropriate when characteristics that are both observable and consequential *fully* account for the endogeneity, which is then sufficiently redressed by consideration of these characteristics (Li, 2012).

To assure readers that all important consequential characteristics have been considered (i.e., are observable), business marketing researchers using PSM, therefore, should explicitly discuss theory or conceptual considerations about the phenomenon at hand. Any

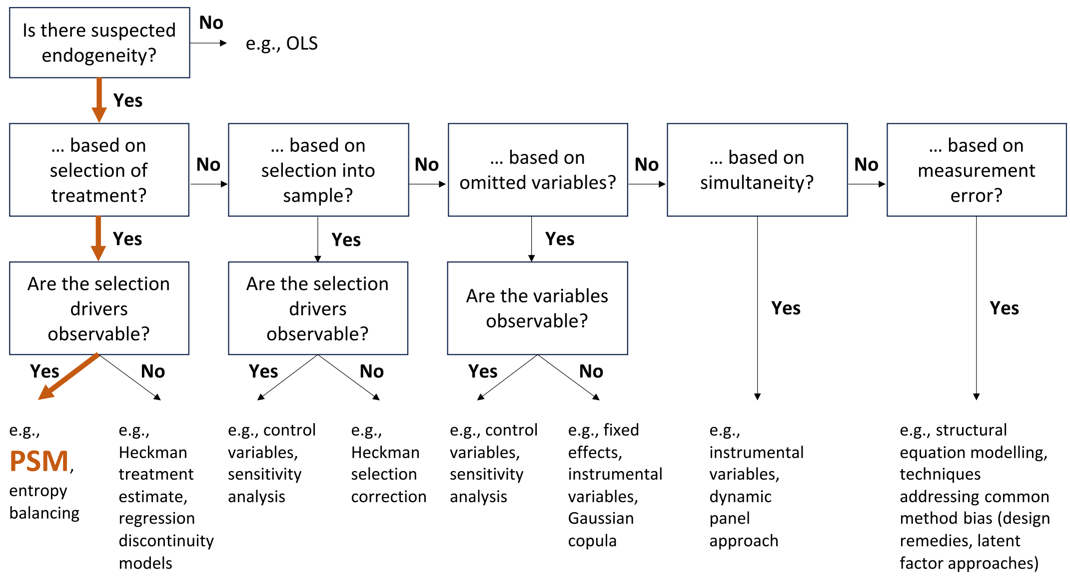

Fig. 1. Decision tree to verify the appropriateness of PSM use.
Note: The decision path to PSM is highlighted for illustrative purposes only.

unobservable characteristics, which the researcher did/could not measure, are an issue if they can be expected to simultaneously affect the selection of treatment and outcome based on the conceptual considerations. If a minority of consequential characteristics remains unobserved, researchers should demonstrate robustness of results to the use of approaches that address endogeneity based on unobservable characteristics, such as endogenous treatment effects models (e.g., Heckman treatment estimate), regression discontinuity models, or synthetic control groups (Certo et al., 2016; Hill et al., 2020). If a majority of consequential characteristics remains unobserved, researchers should exclusively use these alternative approaches instead of PSM or, as we discuss later in this paper, combine PSM with methods that are robust against endogeneity caused by selection of treatment based on unobservable factors.

What is the comparative advantage to a regression model? It is important for business marketing researchers to consider methodological alternatives, including approaches' possible interchangeability or data and context specifics that can make one approach more advantageous than its alternatives. In this regard, researchers should especially consider that a standard regression model aims to achieve a similar objective to PSM, which is balancing on the considered characteristics (Benedetto et al., 2018; Li, 2012). Specifically, a regression model estimates effects ceteris paribus (all else equal), which means that the effect of treatment (versus non-treatment) is estimated by implicitly comparing treatment and control firms with the same levels on the covariates (e.g., important firm characteristics) included in the regression model. In that sense, PSM and regression are "are not really different animals, at least not until we specify a model for the propensity score" (Angrist et al., 2009, p. 83). However, certain research scenarios that we discuss next make PSM more advantageous and business marketing researchers can directly refer to these in their future PSM applications to justify use of PSM.

First, the general data setup can make the use of PSM a more appealing choice. Specifically, the more the data setting mimics an experimental setup, the more convincing is the application of PSM (Goldfarb et al., 2022). The setup especially refers to the categorization of observations in treatment and control observations. This categorization should be naturally occurring (i.e., not require calculation on the part of the researcher) and be clear-cut. For example, the introduction of AI by B2B firms ensures a clear-cut, binary (yes or no) categorization without the need for calculation. In contrast, the case for PSM may be less convincing if a continuous variable (e.g., B2B advertising spending) first needs to be dichotomized in order to arrive at a binary treatment categorization (e.g., high and low spenders) required for PSM. In fact, dichotomization tends to result in less precise effect estimates, reducing the chance of identifying effects that truly exist in the population (i.e., increased type II error) (Shipman et al., 2017). The reason for the reduced precision is that a standard PSM procedure would tend to overrepresent treatment and control observations close to the dichotomization cut-off value, resulting in matched observations with almost equivalent levels of treatment (i.e., on the original continuous treatment variable) for which, therefore, the treatment effect (based on the dichotomized categorization) should be small by design.

Second, business marketing researchers can use the underrepresentation of certain covariate levels in the treatment or control group (i.e., lack of common support) as a justification for their PSM use (Li, 2012). A regression model would extrapolate into areas of the multidimensional covariate space that are not actually observed in the treatment or control groups (Angrist et al., 2009). PSM uses a more cautious approach and only considers treatment and control observations for which satisfactory counterfactuals exist, while other observations are not used in the estimation of the treatment effect (Langworthy et al., 2022; Li, 2012). As we discuss in the review and guidance section below, we recommend that business marketing researchers routinely check for characteristics' balance and common support between the treatment and control groups. Doing so is relatively straightforward in the PSM context, which is one of the method's advantages over a regression analysis.

Third, suspected nonlinearity in the effect of consequential characteristics on the outcome (e.g., financial performance) can support business marketing researchers' argument to use PSM. A standard regression only considers the linear effects of covariates (Benedetto et al., 2018). Although nonlinear covariate terms can be added, researchers would need to choose up to which power this is being done and the resulting model can become complex quickly. In contrast, PSM makes no assumption about the effects' functional form as covariate effects on the outcome are not explicitly modelled (Hirano et al., 2003; Li, 2012). Instead, covariates are used for matching and then the outcomes—which the covariates may affect at an arbitrary level of linearity or nonlinearity—are directly compared between the matched treated and control units.

Finally, certain data scenarios can provide a rational for business marketing researchers to prefer PSM. Specifically, when the number of identified covariates is large relative to the number of observations of the outcome, PSM offers the advantage that it effectively collapses the covariates into the propensity scores in the first step before it models the outcome (Benedetto et al., 2018).

3. Guidance based on the review of prior applications

In this section, we discuss common design choices in PSM, review related recommendations of current literature on the method, and examine their implementation in business marketing applications of the method. To identify relevant business marketing applications, we searched six electronic databases—*ABI/INFORM Global*, *JSTOR*, *Business Source Premier*, *Google Scholar*, *SSRN*, and *ECONLIT*—using the search term 'propensity score' alongside 'B2B', 'business-to-business', and 'business market'. Additionally, to ensure completeness for relevant leading journals, we conducted journal-specific searches on the websites of *Journal of Marketing*, *Journal of Marketing Research*, *Production and Operations Management*, and *Industrial Marketing Management*. We focused on studies in which the authors used PSM for the empirical analysis in a business marketing research context. Based on this search strategy, we identified 26 articles, which are listed in the Appendix.

In the following sections, we present our review and discussion by chronologically viewing a PSM project, considering (1) the justification for using the method, (2) the selection of considered covariates/characteristics in the propensity score model, (3) the specification of the conditioning approach on the estimated propensity score, (4) the evaluation of the achieved covariate balance, and (5) the evaluation of the common support. Table 1 presents the results of our review of prior PSM applications along these considerations. A split of the total sample of reviewed papers into before and after 2022, with approximately equal-sized groups, demonstrates that the issues we discuss in the following are relatively persistent over time.

3.1. Justification of PSM use

As discussed above, PSM does not redress endogeneity in general. Specifically, the approach addresses endogeneity from observed characteristics but not unobserved characteristics. We therefore recommend that business marketing researchers using PSM draw on theory and conceptual considerations to explicitly justify that the simultaneous impact of unobserved characteristics on treatment/selection and the outcome is likely to be inconsequential.

Furthermore, endogeneity from observed characteristics can alternatively be addressed by means of a regression model. Thus, comparative advantages of PSM should be presented. As discussed above, this can be achieved by means of (1) the general data setup involving a naturally occurring, clear-cut categorization of observations in treatment and control observations (e.g., a B2B firm's *yes* or *no* decision to take a certain action), (2) lack of common support on important characteristic levels across the treatment and control groups, and/or (3) suspected nonlinearity in the characteristics' effects on the outcome (e.g., financial performance). Our review of prior PSM applications in business marketing shows that fewer than one-third of the studies provided the appropriate justification for PSM's use.

Table 1
Review of prior PSM applications in business marketing research.

Consideration	Total	Before 2022	From 2022
Appropriate justification of PSM use	27 %	36 %	17 %
Justification of selected covariates	15 %	14 %	17 %
Justification of the conditioning approach	8 %	7 %	8 %
if applicable, justification of the parameters	27 %	21 %	33 %
Evaluation of the achieved covariate balance	38 %	36 %	42 %
Evaluation of the common support	42 %	57 %	25 %

3.2. Selection of covariates/characteristics

The covariates considered in the logit or probit model, which is used to determine the propensity score, need to be carefully selected. Theoretical and conceptual considerations should guide researchers' selection. Specifically, a convincing case needs to be made that the covariates affect *both* the treatment (e.g., B2B firm's decision) *and* the outcome. A common misconception is that all covariates that are likely to affect the treatment should be included. This is not the case. In fact, to safeguard the precision of estimates, business marketing researchers should ensure that covariates that only affect the treatment but not the outcome (i.e., instrumental variables) are excluded (Austin et al., 2007; Myers et al., 2011). Similarly, covariates that conceptually are mediators between treatment and outcome or downstream consequences of the outcome should be excluded (Andrew et al., 2023).

Another related common misconception assumes that the objective is to identify covariates that maximize the explanation of treatment. However, when covariates almost perfectly predict treatment (i.e., they are deterministic), it implies that only firms with certain characteristics take the action of interest, while the other firms do not. In such a scenario, identification of suitable counterfactuals is problematic, which means that the treatment effect cannot be estimated or only be estimated with large error due to a small number of successful matches (Thoemmes & Kim, 2011). Therefore, business marketing researchers should strictly only include covariates supported by theory and avoid empirical searches (e.g., based on goodness-of-fit) for covariates with explanatory power (Ali et al., 2015). Notably, our review of prior PSM applications in business marketing shows that only about one in seven studies provided the appropriate justification of selected covariates.

Excluded covariates can be useful for sensitivity analyses. For instance, certain covariates may have been excluded from the main model due to high conceptual overlap with other covariates or because they are only partially backed by theory. In robustness tests, business marketing researchers can compare their original results with results when these covariates are included instead of excluded (Dehejia, 2005).

3.3. Specification of the conditioning approach

After the propensity score is estimated, business marketing researchers need to determine how it is used to ensure that outcomes are compared between similar (i.e., regarding the propensity score) treatment and control firms. Five common approaches exist and each of these requires its own specific considerations. We discuss the approaches and considerations next and explain why one of the approaches should be avoided all together. To evaluate the approaches, we considered recent simulation study results that we summarize in Table 2 (Guo et al., 2020). Our analysis of prior business marketing research shows that fewer than one in ten studies provided justification for the conditioning approach.

Table 2
Rank-order of PSM approaches based on simulation study results by Guo et al. (2020).

	Rank (1 = best, 5 = worst)	
Approach	Selection on observables	Selection on unobservables
Nearest neighbor matching	2	2
Subclassification/stratification	1	5
Optimal matching	5	1
Weighting	3	3
Kernel-based matching	*not considered*	*not considered*
Direct inclusion in outcome model	*not considered*	*not considered*
Benchmark: Ordinary least squares	4	4

3.3.1. Nearest neighbor matching

The idea of nearest neighbor matching is to identify the treatment effect by comparing the outcomes of treatment and control firms that are matched based on their similar propensity scores (Rubin & Thomas, 1996). The approach is popular and valid. However, it necessitates researchers to make decisions about multiple design parameters, increasing the demand to demonstrate the sensitivity of results to these design choices.

First, researchers need to decide whether matching is performed *without replacement* or *with replacement*. In other words, researchers must determine if a given control observation can only be used once for matching purposes or multiple times. Both approaches have advantages and disadvantages and suitability depends on the specific research context, which means that no approach is generally superior to the other. However, we recommend that business marketing researchers discuss their design choice against the backdrop of the following considerations (Shipman et al., 2017). On the one hand, replacement can enhance matching quality, as it ensures that the control observation with the closest propensity score (i.e., the most similar counterfactual) is always matched, regardless of whether the observation has been matched before. Sample size is also preserved more effectively, considering that matching to treatment observations is less likely to fail as the pool of potential control candidates is never exhausted. On the other hand, replacement increases the risk that the estimated counterfactual outcome is only based on certain, and potentially few, control observations to the degree that selected control observations are matched many times. This is especially problematic if the matched control observations are outliers (i.e., have atypical outcome values), which then have an undue influence on the estimated treatment effect owing to their repeat matching. We hence recommend that business marketing researchers using matching with replacement assess the outcome values of control observations that were matched many times. This should be done against the backdrop of the average outcome value in the control group to identify possible influential outliers. Moreover, from a statistical perspective, repeat observations should be weighted downwards according to the frequency of their use, and standard errors need to be adjusted upwards (Armstrong et al., 2010; Stuart, 2010). As repeated matching of the same observation can mean that data is no longer independent, researchers should consider using robust error estimators, such as sandwich estimators (Benedetto et al., 2018).

Second, researchers need to decide whether one (i.e., *one-to-one matching*) or multiple (i.e., *one-to-many matching*) control observations are matched to each of the treatment observations. The relative sample sizes of the treatment and control groups are useful to determine the approach. One-to-one matching is a convincing choice when the groups are approximately equally sized or the number of treatment observations exceeds to number of control observations. When the control group is substantially larger, one-to-many matching becomes a compelling approach, using a ratio that approximately corresponds to the factor by which the number of control group observations exceeds the number of treatment observations (e.g., 1:2 if the control group is twice as large). In this scenario, one-to-many matching can increase statistical power (i.e., precision), which determines the ability to identify a treatment effect that truly exists in the population (Shipman et al., 2017). However, we do not recommend more than five matched control observations. Precision gains tend to be small for more than five matched control observations and matching quality can suffer if a large number of observations is matched (Li, 2012; Rosenbaum, 2020). Just over a quarter of the prior PSM applications in business marketing research that we reviewed provided justification for the matching approach.

Third, to ensure matching quality, business marketing researchers should specify and report the *caliper distance*. This distance can be understood as the maximum allowable dissimilarity between treatment and control observations for a successful match to occur. Use of the caliper distance has been shown to improve estimation results (Austin, 2014). To support interpretation and replicability, researchers should report the value of the employed caliper distance (Benedetto et al., 2018). It is crucial to note that this distance can be expressed in three distinct terms that are not interchangeable. Specifically, the caliper distance can be defined as a fraction of the raw propensity score, its standard deviation, or the standard deviation of the propensity score's logit (Shipman et al., 2017). We recommend using the standard deviation of the propensity score's logit because compelling evidence regarding optimal cut-off values exists for this measure. Specifically, a simulation study shows that a caliper distance of 0.2 based on this measure is optimal (Wang et al., 2013). However, business marketing researchers can opt for more conservative or liberal caliper distances in certain scenarios. For instance, a more liberal caliper distance is justified when many treated firms could otherwise not be matched. Unsuccessful matches reduce statistical power and threaten the validity of findings to the extent that the estimated treatment effect may not be representative for all treated firms. In contrast, researchers can justify a more conservative caliper distance when achievement of successful matches is not an issue, as doing so ensures matches of highly similar treatment and control observations, improving the quality of results.

Overall, recent simulation study results show that nearest neighbor matching is ranked second-best among the considered PSM approaches when observable characteristics determine the treatment and outcome (Table 2). The approach ranks equally well when unobservable characteristics determine the treatment and outcome, showing a certain desirable robustness when PSM assumptions about the observability of consequential characteristics are violated.

3.3.2. Subclassification/stratification

Compared with nearest neighbor matching, the stratification approach has the advantage that researchers need to decide about fewer parameters. In a nutshell, the approach sorts observations based on their estimated propensity score and divides the sample into n strata, which are usually equal-sized (Guo et al., 2020; Thoemmes & Kim, 2011). Thus, the number of strata is the key parameter that researchers select. In this regard, the literature recommends the use of five strata, which approximately mitigates 90 % of the bias that the included observed characteristics would otherwise have created (Benedetto et al., 2018; Cochran, 1968). Additional precision gains are possible by flexibly selecting the strata size, through an optimization algorithm, so that the variance of the treatment effect estimate is minimized (Hullsiek & Louis, 2002). A reason for this potential gain is that flexible sizing reduces the risk that certain strata cannot be used as they contain only treatment or only control observations. However, researchers have to weigh the advantage of using all strata (i.e., sample size and generalizability) with the potential disadvantage of poor matching quality within strata. The treatment effect based on the stratification approach is estimated through calculation within each stratum and aggregation across strata. Specifically, the mean and variance of the outcome difference between treated and control observations are calculated within each stratum and are then aggregated.

Simulation study results show that the stratification approach is superior compared with alternative propensity score-based approaches discussed here in terms of balancing the typical trade-off between bias reduction (e.g., ensuring quality matches) and estimation precision (e.g., retaining sample size) when observable characteristics drive the treatment and outcome (Table 2). However, the approach becomes the worst performing approach when unobservable characteristics determine treatment and outcome. Therefore, business marketing researchers would only opt for the stratification approach when they are highly confident that all consequential underlying factors driving treatment and outcome have been captured with the set of matching variables.

3.3.3. Optimal matching

Like the stratification approach, optimal matching uses strata but the approach is designed to always achieve full matching, using the full sample for estimation without loss of observations (Rosenbaum, 2002).

Strata composition and sizes are determined flexibly based on the objective that the summed distance in the propensity scores across strata are minimized. The approach's main advantage is that sample size is preserved, which can make optimal matching especially beneficial for small samples. Small samples are not uncommon in business marketing research, because researchers often find it difficult to recruit respondents as the population of interest is small (e.g., only a few firms adopt a new technology or approach) and the typical target informants (i.e., employees or managers) have limited time resources. However, business marketing researchers should consider the possible loss in matching quality due to matching of potentially dissimilar treatment and control observations, which can bias inferences.

Simulation study results indicate that optimal matching is inferior to nearest neighbor matching and stratification in terms of recovering an unbiased treatment effect when observable characteristics determine treatment and outcome (Table 2). The approach also underperforms a standard ordinary least squares regression model. Therefore, we currently cannot recommend the approach for PSM's typical application context, although further simulation study research is needed that specifically considers small samples. Interestingly, the approach works surprisingly well when unobservable factors drive treatment and outcome. However, for this contexts, different methods (e.g., treatment effect models) provide more reliable results.

3.3.4. *Weighting*

The propensity score-based weighting approach estimates a regression model based on weighted observations instead of the original observations (Angrist et al., 2009). The approach borrows its logic from the use of sampling weights in regression models to redress nonrandomness of sampling. In the context of PSM, observations are weighted by the reciprocal of their estimated treatment probability, which is given by the propensity score.[2] Similar to optimal weighting, the approach's advantage is that the original sample size is retained. However, we strongly recommend that business market researchers using this approach carefully assess the estimated propensity scores. Treatment (control) observations with scores close to zero (one) are assigned extreme weights, warranting robustness tests that consider truncation of extreme weights. Inattention to this issue can result in substantially biased estimates (Kang & Schafer, 2007). Moreover, robust standard errors using the Huber-White sandwich estimator should be used for weighted regressions as, otherwise, standard errors tend to be understated (Lohr, 2022). Alternatively, the standard errors can be determined through bootstrapping.

Simulation study results show that the propensity score-based weighting approach ranks in the mid-range of the considered PSM approaches and is relatively stable when assumptions about the observability of consequential characteristics are violated (Table 2). However, it is outperformed by nearest neighbor matching in all data scenarios.

3.3.5. *Kernel-based matching*

The kernel-based matching estimator constructs counterfactuals per each of the treatment observations in the sample (Heckman et al., 1997). The counterfactuals are weighted averages of all of the control observations within a bandwidth that the researcher specifies (Heckman et al., 1997 use a bandwidth of 0.06). The weights are determined so that control observations with propensity scores closer to the treatment observation's propensity score are weighted more highly, while a lower weight is placed on more distant control observations (Heckman, Ichimura, & Todd, 1998). A nonparametric local regression is used to determine the weights (Heckman et al., 1997).

Simulation studies have not yet compared kernel-based matching with alternative PSM approaches (Table 2). We therefore recommend that business marketing researchers use one of the established PSM approaches but remain open to future simulation studies that consider kernel-based matching and its comparative performance.

3.3.6. *Direct inclusion in outcome model*

We do not recommend to directly include the estimated propensity score in the outcome equation, which is sometimes done in the literature. Doing so defeats a key advantage of PSM over a regression approach, which is to control for suspected nonlinearity in the effect of consequential time-invariant firm characteristics on the outcome. Covariates in regression models only control for linear effects (Thoemmes & Kim, 2011). Business marketing researchers should therefore avoid this approach and use one of the established approaches discussed before.

3.4. *Evaluation of the achieved covariate balance*

As discussed, the propensity score effectively is a means to reduce the dimensionality of the matching covariate space to a single dimension. However, this raises the question of how similar the treatment and control groups are in terms of the original covariates (Hansen, 2008). In general, larger samples are preferred for PSM, as they ensure higher-quality matches and, consequently, more robust inferences (Peikes et al., 2008). Moreover, we recommend that business marketing researchers use *t*-tests to statistically assess the mean differences between the treatment and control groups across the covariates used to calculate the propensity score. Our analysis of literature reveals that fewer than two out of five reviewed studies in business marketing assessed covariate balance. For the stratification approach this assessment should be done per strata. Typically, non-significant differences are found for the majority of covariates. Researchers should include covariates with significant differences directly in the outcome equation, adding at least the covariates' linear and squared terms to control for linear and nonlinear effects.

However, researchers should verify that the t-tests have sufficient statistical power, which can be an issue in small samples. For small samples, researchers should use Cohen's d as a standardized difference that relates the mean difference between the treatment and control groups to the pooled standard deviation (Cohen, 1988). The typical guideline values are 0.2, 0.5, and 0.8 for small, medium, and large effects. Researchers assess Cohen's d to ensure that any differences between the covariates are small.

3.5. *Evaluation of the common support*

PSM can result in biased treatment effect estimates when the final matched sample is not a good representation of the original sample (Shipman et al., 2017). For instance, the matched sample can be unrepresentative due to unsuccessful matches (i.e., when no suitable control observation can be identified). Unsuccessful matches typically occur for treatment observations with high propensity scores (i.e., a high likelihood of treatment). However, the treatment effect may be most likely to fully materialize for such observations and, hence, their exclusion can impede the effect's accurate estimation. Moreover, the matched sample can be unrepresentative if only certain control observations are used. For example, control observations with a very low likelihood of treatment may never be matched to a treatment observation. This issue can be exacerbated in nearest neighbor matching with replacement as then only control observations with high or moderate propensity scores may be (repeatedly) used for matching while other control observations are excluded.

We recommend business marketing researchers to carefully assess the extent of common support by plotting the estimated propensity scores for the treatment and control groups. For each group, observations with a similar propensity score should exist from the other group across the propensity score continuum. Observations without

[2] Treatment observations are weighted by $1/PS^*$ and control observations are weighted by $1/(1 - PS^*)$, where PS^* is the estimated propensity score.

equivalents are likely to be excluded from the PSM analysis or be matched with very dissimilar counterfactuals if researchers select a liberal tolerance level (e.g., caliper distance). Moreover, we recommend that researchers compare the list of observations that were included in the final analysis with the list of observations in the original sample to assess whether certain observations were excluded systematically. If a systematic exclusion of observations is found, we recommend that researchers discuss its consequences as part of the research limitations. Notably, our analysis of the existing business marketing research reveals that fewer than half of the studies assessed the common support.

4. Augmented PSM approaches to address treatment selection based on unobservable factors

As explained above, the PSM approach as a stand-alone method is suitable to address endogeneity caused by selection of treatment based on observable factors. While initial simulation study results indicate that PSM based on nearest neighbor matching is relatively robust against selection bias from unobservable factors (Guo et al., 2020), approaches that directly account for this bias are nevertheless preferable. Moreover, examples exist in the literature where inferences based on PSM can be incorrect, most likely due to consequential unmeasured unobservable factors, even when a large number of observed factors is considered (Peikes et al., 2008). In the following, we discuss combinations of PSM with alternative approaches to more exhaustively address effects from selection of treatment, accounting for unobservable factors in addition to observable factors. We distinguish between two types of combinations. The first type builds on unique data settings where events or interventions create treatment and control groups. The second type combines PSM with methods discussed earlier that are robust to treatment selection based on unobservable factors.

The first combination type requires a certain cleverness on part of the researcher to identify rare data settings, which involve an exogenous event or intervention that affects the endogenous variable of interest (Card & Krueger, 1994; Wooldridge, 2010). Exogenous means that the event or intervention must be outside of a unit's (e.g., a B2B firm) control *and* the unit must not anticipate it.[3] To give an example in the recent marketing literature, albeit outside of the business marketing domain, a recent study used a relatively sudden change in accounting rules (i.e., an unanticipated policy intervention) in combination with PSM to identify the benefits and costs of marketing accountability, which is likely to be driven by many underlying (unobservable) factors that would be impossible to measure (Guenther et al., 2024). Another recent example study in strategic management used PSM in combination with the Fukushima nuclear accident in Japan as an exogenous shock that drove staff turnover in US firms near nuclear power plants in order to address the simultaneity issue that complicates proper identification of the relationship between turnover and performance (Stern et al., 2021). Both studies used a difference-in-differences (DID) approach to identify effects of interest by comparing outcome differences between treatment and control groups before and after the event or intervention (for details on the DID approach, see Ryan et al., 2015). PSM is used to match control firms that are highly similar to treatment firms in terms of a set of important observable characteristics that are likely to drive both the treatment selection and outcome.

A key challenge of the approach based on natural experiments is that appropriate data settings are so rare that it can make the investigation of certain phenomena impossible, thereby limiting academic insight in potentially important areas. Moreover, even when a suitable intervention such as a policy change can be identified, it may not be truly exogenous, as the upcoming change is often known years in advance, allowing firms with certain unobservable characteristics to benefits from it, including in terms of the outcomes of interest (e.g., Guenther et al., 2024). Effect estimates would then be biased by these unobservable characteristics, which means that the natural experiment would not solve the underlying endogeneity issue. Nevertheless, if a suitable, truly exogenous event or intervention can be identified, the estimated effects would be causal effects, providing an otherwise unmatched robustness and reliability in insights.

When the first combination approach is infeasible for a lack of suitable data, business marketing researchers can resort to a fully methods-based approach combining PSM with methods specifically designed to address treatment selection based on unobservable factors (Heckman, Ichimura, Smith, & Todd, 1998). Specifically, Heckman treatment estimates and regression discontinuity models can be combined with PSM (Linden & Adams, 2012; Makepeace & Peel, 2013). In the first case, following the standard Heckman treatment effects approach, treatment selection is modelled using a probit model, and the inverse Mills ratio is calculated to represent the unobservable part of the selection decision (Heckman, 1979; Wooldridge, 2010). The inverse Mills ratio is then included in an outcome equation, estimating treatment effects by comparing treatment and control observations matched through PSM (Makepeace & Peel, 2013). Alternatively, PSM can be combined with the regression discontinuity approach if the data requirements for the latter are met (Linden & Adams, 2012). Specifically, regression discontinuity models require that treatment versus non-treatment is determined by a clear cut-off value on a continuous underlying variable; for example, only B2B firms with revenue below a certain threshold may qualify for funding or support. Units just above and below the threshold are used as counterfactuals, assuming similar unobservable characteristics, to estimate the treatment effect (Thistlethwaite & Campbell, 1960). In this context, PSM can further refine the comparison by matching treatment and control observations not only close to the cut-off but also similar in important observable characteristics (Linden & Adams, 2012).

5. Step-by-step guide of PSM

Our step-by-step guide follows the guidelines established in the PSM literature. As discussed above, typically PSM is used to study the treatment effect on an outcome variable of interest. However, since secondary data often do not come from randomized trials but from (nonrandomized) observational studies, the treatment group can suffer from selection bias, which can undermine the validity of the analysis. PSM is one of the key ways to reduce selection bias. In their seminal work, Rosenbaum and Rubin (1983) proposed PSM as a method to reduce the bias in the estimation of treatment effects with observational datasets. Therefore, PSM enables us to confidently evaluate the causal effect of treatment on selected outcomes, assuming that the method's assumptions are met.

To guide business marketing researchers effectively, we present PSM implementation steps chronologically based on their occurrence in a project. Specifically, we discuss:

- Identifying a context
- Initial data selection
- Identifying the subsample
- Creating the final sample
- Implementing the propensity score matching process, including software implementation and advanced matching consideration
- Presenting the results

5.1. Identifying a context

PSM is highly useful to identify treatment effects in business marketing contexts. In this step-by-step guide, we consider the context of

[3] Researchers should also conceptually verify that the event or policy is not driven by unobservable factors that could also affect the outcome of interest, as in that case the event or policy itself would be endogenous (Besley & Case, 2000).

acquisitions in business markets. The financial value of acquisitions (i.e., the treatment effect) manifests itself in B2B firms' stock returns following the acquisition (see Fig. 2). Formally, we are interested in evaluating the causal effect of the treatment (acquisitions) on outcome Y (stock returns) of the firms in the population.

Mathematically, this problem can be denoted as:

$$Y_i = Y_{0,i} + D_i\left(Y_{1,i} – Y_{0,i}\right) \quad (1)$$

which is the actually observed outcome for firm i, where $D_i \in \{0,1\}$ indicates whether firm i actually received the treatment. $Y_{1,i}$ is the outcome if firm i were exposed to the treatment, $Y_{0,i}$ is the outcome if firm i were not exposed to the treatment, and X is a set of important pretreatment characteristics.

The causal effect for firm i is then:

$$Y_{1,i} – Y_{0,i} \quad (2)$$

As noted above, the 'fundamental problem of causal inference' is that it is impossible to observe the individual treatment effect, making it impossible to derive causal inference without making assumptions – that are usually untestable.

The average treatment effect on the treated, across firms, is given as:

$$E(Y_1 – Y_0|D = 1) = E(Y_1|D = 1) – E(Y_0|D = 1) \quad (3)$$

where E(.) denotes the expected outcome. $E(Y_1|D = 1)$ is the outcome of firms with treatment (i.e., which performed an acquisition), while $E(Y_0|D = 1)$ is the hypothetical counterfactual of the same firms had they not been treated. Usually, $E(Y_1|D = 1)$ can be directly observed in the data, while $E(Y_0|D = 1)$ cannot. To be able to estimate the treatment effect, researchers therefore need to construct the counterfactual $E(Y_0|D = 1)$.

5.2. Initial data selection

We use the SDC platinum database to identify the B2B firms that have made one or more acquisitions in the year 2017. For the purposes of this step-by-step guide, we chose to restrict the sample to a single year but note that PSM can be extended to multiple years, in which case the matching should be done on a yearly basis. After downloading the B2B acquisitions data and cleaning it, we merged this data with the financial data from Compustat database. The financial data is not restricted to the firms in the B2B acquisitions data but instead contains financial details of all the firms publicly listed on US stock exchanges. This data is used to identify counterfactuals of the treatment firms (i.e., firms without an acquisition) in order to create a matched control group.

5.3. Identifying the subsample

As discussed above, we investigate the effect of acquisitions by B2B firms on their stock returns. To identify the subsample, we first considered all the acquisitions made by public US firms in 2017. We restricted the analysis to public US firms due to the availability of stock returns and financial data that we used to measure the dependent and control variables (Fig. 2). We were able to identify 2117 acquisitions. Given our context, we then selected only the acquisitions by B2B firms, defined as firms that sell their products or services predominantly to other firms or the government instead of consumers (Delgado & Mills, 2020). Applying this criterion resulted in a subsample of 1334 acquisitions by B2B firms.

5.4. Creating the final sample

We merged the data on B2B acquisitions obtained from the SDC platinum database with the Compustat data for the year 2017. We used Compustat data to create measures of our dependent, matching, and control variables. We used stock returns as the dependent variable as stock returns are an effective forward-looking measure of firm performance and indicate investors' expectations of firms' future performance (Baillie & DeGennaro, 1990; Zhang, 2006). To measure stock returns, we used the change in share prices in the current year (Fama, 1990). Furthermore, we used the financial data to calculate the control variables that can potentially impact firms' stock prices based on the past research on acquisitions (e.g., Lambkin & Muzellec, 2010; Palmatier et al., 2007; Richey et al., 2008).

5.5. Implementing the propensity score matching process

In order to study the effect of acquisitions on stock returns, a key identification challenge is to determine close counterfactuals of the B2B firms with acquisitions. Identification of suitable counterfactuals enables us to directly compare the returns of the firms with and without acquisitions during the sampling year. For illustration purposes and to simplify the following discussion, we used 1:1 matching and matched each treatment firm (i.e., a firm that engaged in acquisitions) with one control firm that did not acquire another firm. The results are robust to matching with more control firms (e.g., five).

The matching process involves pairing each treated firm with a control firm that has the closest propensity score during the sampling year. The propensity score is estimated using relevant variables that are likely to affect both firms' decision to engage in acquisitions and the outcome (i.e., stock return) of this activity. In our context, the goal of matching is to derive effects based on the comparison between firms with highly similar probabilities of engaging in acquisitions and realizing benefits from it. For our application example, we match firms on their size (total assets), resource slack (ratio of cash by assets), financial leverage (ratio of long-term debt to assets), and liquidity (ratio of current assets to current liabilities). The matching variables (criterion) can be changed as per the requirements of the research. However, the greater the number of criterion firms the more difficult it will be to find a match for every firm. Therefore, business marketing researchers should keep in mind that, as the number of matching parameters increases, the final sample size tends to decrease. Researchers thus need to find a balance in choosing matching criteria. Past research on the topic can generally be a useful guide for this purpose.

5.5.1. Software support

We focus here on the implementation of PSM in Stata, as it provides business marketing researchers with a convenient platform, offering packages, fine-tuning options, and support through Statalist. For researchers who prefer open-source software, R and Python are viable alternatives. R offers specialized packages, such as *Matching* for optimal matching (Sekhon, 2011). Moreover, *MatchIt* is available, which is a generalized package that supports various PSM conditioning approaches, including nearest neighbor matching, subclassification/stratification, and optimal matching (Ho et al., 2011). Python also features a dedicated package, *PsmPy*, for PSM using nearest neighbor matching (Kline & Luo, 2022), with additional packages like *pysmatch* under development for alternative conditioning approaches.[4]

Stata provides access to several user-written commands for propensity score matching such as *gmatch*, *psmatch2, pscore, and nnmatch*. While there are minor differences in the specifics of the commands' matching approaches, the results tend to be very similar across commands. For instance, the matching criteria used tend to be more consequential than the used algorithm. For this step-by-step guide, we used *gmatch* and describe its implementation next. In our example, we used 1:1 matching (i.e., one treatment firm to its counterfactual control firm). As discussed, alternatively, one treatment firm can be matched to more than one control firm, although this number should not exceed five. The results in our example are robust to these choices. Moreover, *gmatch* allows the researcher to specify the caliper distance, although

[4] https://github.com/miaohancheng/pysmatch

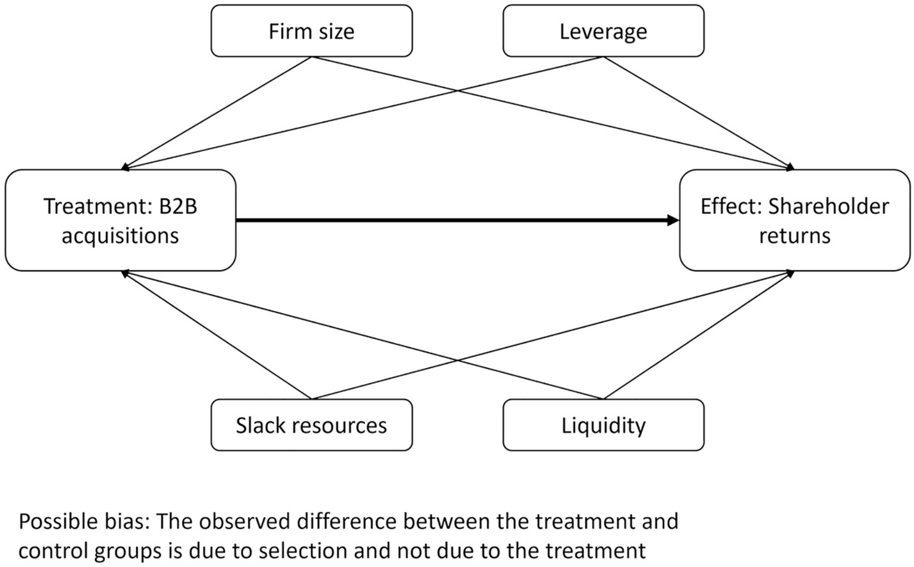

Fig. 2. Causal relationship example.

this needs to be done in terms of the maximum absolute propensity score difference instead of the standard deviation of the propensity score's logit. The code for implementing *gmatch* is given in Fig. 3.

First, we ran a pooled logit regression that models the probability of a firm being in the treatment or control group (i.e., the probability that a B2B firm engaged in acquisitions) based on the set of matching variables discussed above. Second, we used the estimation results to predict the probabilities per observation, saving it as a new variable (*pscore*). Third, *gmatch* matches treatment and control firms with the most similar estimated *pscore*. The command generates a new variable, *set,* which provides a unique value for each matched pair, facilitating identification of the matched pairs in the dataset. From our sample of 1334 acquisitions by B2B firms (i.e., treatment firms), the algorithm was able to identify matches for 1039 firms and did not find matches for 295 treatment firms due to insufficiently similar propensity scores in the control group. The final sample, therefore, comprises 2078 firms (1039 treatment and 1039 control firms).

5.5.2. Matching on industry codes

In large samples comprising firms from different industries, and a large number of firms per industry, matching within industries is likely to improve the matching quality with regard to the achieved similarity between treatment and control firms. For example, matching a pharmaceutical treated firm to another pharmaceutical control firm, enhances the level of control for industry-level characteristics that can affect both the assignment to the treatment (i.e., engaging in acquisitions) and the outcome (i.e., stock returns). Here, a challenge is that popular industry classification codes, such as SIC (Standard Industrial Classification) or NACIS (North American Industry Classification System), are categorical variables and therefore cannot be directly used in the first-stage logit model. However, this challenge can be addressed with a minor tweak by adding one additional line to the code. Specifically, instead of matching the estimated probability score (*pscore*), researchers can create a new variable (*pscore2)* that combines the *pscore* with the sic code (sic) for each observation. This new variable (*pscore2*) is then used for matching. The modification to the code is also given in Fig. 3. When we performed the matching based on SIC, the number of matched firms dropped to 926, resulting in a final sample of 1852 firms (926 treatment plus 926 control).

5.5.3. PSM with selection on unobservables

Although we match the treatment firms and their counterfactuals on key firm characteristics, underlying unobservable differences in the two groups could remain and affect both the treatment selection and the dependent variable (i.e., stock returns). For example, cultural differences between treatment and control firms may affect both treatment and stock returns. Not accounting for such unobservable characteristics in the model can undermine the model's ability to accurately identify the causal effect of the treatment on the outcome. To account for such unobservable factors, PSM can be combined with a Heckman-style selection model (Heckman, 1979). In our example analysis, we therefore ran a probit regression to predict a firm's probability of engaging in acquisitions based on the antecedents of corporate acquisition activity, as per the existing literature in this area. We then derived the inverse Mills ratio from the first-stage probit model and included it as one of the predictors in the final regression equation. Adding this variable to the model effectively accounts for omitted unobservable factors.

5.6. Presenting the results

Before running the final analysis, as discussed, covariate balance needs to be established and demonstrated between the treatment and control firms. To this effect, we performed *t*-tests on the differences in the mean values of the matching variables between the treatment and control firms. Covariate balance is established when the differences are not statistically significant. For our example, the results of the mean differences are reported in Table 3 and show that the differences are not significant. This result confirms that similar treatment and control firms have been matched, indicating the absence of systematic differences between the groups on the matching criteria that could otherwise influence the dependent variable.

After covariate balance is confirmed, the final sample can be analyzed using the appropriate statistical technique. As we were interested in the impact of B2B firms' acquisitions on stock returns, we ran a regression model that estimates the effect of *Treat*, which is 1 for the firms that engaged in acquisitions and 0 for control group firms. The results of the model are reported in Table 4. We considered three models, using PSM only, additionally matching on SIC, and including the inverse Mills ratio. The sample sizes for the three models differ and

Code for Section 4.5.1.

Line 1: *logit Treat Size Resource_Slack Leverage Liquidity*

*** Logit regression that predicts the probability of an observation belonging in the treatment group based on the predictors ***

Line 2: *predict double pscore if e(sample), pr*

*** Generates the probability of a positive outcome ***

Line 3: *drop if missing(pscore)*

*** Removing the observations for those no pscore was generated; this can be possible if any of the variables for an observation is missing ***

Line 4: *gmatch Treat pscore, cal(0.1)*

*** Matching using treatment firms with control firms based on pscore, "cal(0.1) allows to control for quality; if for a treatment firm, the difference between its pscore and control firms is more than 0.1 gmatch will not match is with any control firms; the value 0.1 can be changed; check the gmatch help file for details***

Line 5: keep if set!=.

*** Keeping all the treatment and control firms that were matched ***

Code for Section 4.5.2.

*** To match the firms based on their standard industry classification (SIC) codes following line should be added before line 3 of the above code ***

gen double pscore2 = sic+pscore if pscore!=.

*** in this case now the matching should be done using the variable pscore2 ***

Fig. 3. Implementing *gmatch*.

Table 3
Difference in mean values of the matching variables between treatment and control firms.

Variable	Mean Treat (T)	Control (C)	T-C
Size	7.392	7.388	0.004 N.S.
Slack resources	0.127	0.120	0.007 N.S.
Leverage	0.245	0.249	−0.004 N.S.
Liquidity	2.948	2.591	0.357 N.S.

are marginally smaller than the total number of identified treatment firms due to missing values for the various variables used in the models. The estimated treatment effect varies across the three models. The results show that the effect of acquisitions by B2B firms on stock returns is negative and significant when PSM considers the SIC codes, and when the effect of unobservable characteristics is controlled by means of the inverse Mills ratio. Without these additional controls, the estimated effect is not significant.

6. Conclusion and directions for further research

In business marketing research, there is a constantly growing interest to utilize advanced statistical techniques such as PSM that effectively addresses certain endogeneity issues in observational studies. Nonetheless, our review of prior PSM applications in business marketing has revealed several issues with the understanding of when PSM is appropriate and the consideration of necessary application routines required to justify the approach and demonstrate its robustness. These issues are

Table 4
Effect of acquisitions by B2B firms of stock returns.

Stock returns	(1) Coeff.	(1) (S.E.)	(2) Coeff.	(2) (S.E.)	(3) Coeff.	(3) (S.E.)
Treat	***−0.169***	***(0.119)***	***−1.145****	***(0.639)***	***−1.078****	***(0.640)***
Size	−0.004	(0.030)	−0.251	(0.170)	3.458	(2.188)
Return on assets	0.016	(0.024)	−1.095***	(0.108)	−1.224***	(0.122)
Slack resources	−0.047	(0.426)	−2.188	(2.431)	−6.500*	(3.569)
Liquidity	−0.011	(0.009)	0.016	(0.053)	−0.038	(0.062)
Matched on SIC	No		Yes		Yes	
Inverse Mills ratio	No		No		Yes	
F-statistics	0.37		3.09***		3.50***	
Observations	1775		1603		1590	

Table 5
Checklist: PSM motivation and recommended application routines.

Panel A: Motivating PSM use

- ✓ Conceptual justification provided that a selection of treatment based on observable characteristics creates the endogeneity issue
- ✓ Conceptual justification provided that all important consequential characteristics are captured with the available data
- ✓ Regression approach conceptually ruled out based on the data setup (e.g., clear-cut and naturally occurring treatment and control groups), imperfect common support (i.e., underrepresented covariate levels in any of the groups), and/or possible nonlinearity in covariate effects

Panel B: Recommended application routines

- ✓ Justification of selected covariates
  - o based on theory or conceptual considerations
  - o the majority of important consequential characteristics have been captured
  - o consequential = affect *both* treatment assignment/selection *and* the outcome
  - o strikes a balance between similarity of the treatment/control groups and preservation of sample size (i.e., statistical power)
- ✓ Justification of the conditioning approach
  - o based on current knowledge, use the nearest neighbor matching (NNM) approach
  - o NNM is relatively more robust than other PSM approaches even when unobserved characteristics create part of the endogeneity issue
  - o use the subclassification/stratification approach instead, if the effect of unobserved characteristics can be ruled out
  - o avoid the following approaches: optimal matching, weighting, kernel-based matching (more comparative research needed), direct inclusion in the outcome model
  - o reporting and justification of matching parameters
    - ▪ NNM: with or without replacement: if with replacement, check for influential observations used multiple times; 1:1 or 1:m matching based on relative treatment/control groups sizes with $m \leq 5$; caliper distance of 0.2 based on the standard deviation of the propensity score's logit, unless the loss in statistical power justifies a more liberal value
    - ▪ Subclassification/stratification: use five strata; unequal strata sizes can be justified if the statistical power gains (increased sample size) outweigh the costs of potentially more imbalanced treatment/control groups
- ✓ Evaluation of the achieved covariance balance
  - o perform and report t-tests to compare the treatment/control groups regarding the mean levels of the covariates used to calculate the propensity score
  - o for covariates with statistically significant differences, include their linear and squared terms directly in the outcome equation
  - o for small samples, use Cohen's $d < 0.2$ as a cut-off value instead of *t*-tests to identify covariances with substantial imbalances
- ✓ Evaluation of the common support
  - o report the distributions of propensity scores in the treatment/control groups
  - o check for systematic patterns in unsuccessful matches contributing to underrepresentation of certain types of treatment/control firms
  - o if underrepresentation is found, discuss its consequences as part of the research limitations

equally present in older and more recent studies, highlighting the need for a comprehensive guide for business marketing researchers on PSM to improve the rigor of future empirical studies.

This article provided guidance along three main dimensions. First, we clarified PSM's methodological backdrop, including the data context in which the approach is appropriate. This backdrop is often somewhat buried, even in the methodological literature on the approach. However, business marketing researchers require it to convincingly justify choosing PSM over alternatives. Second, we outlined the necessary application routines for PSM use. These routines comprise the justification of the method, the considered covariates for the propensity score, and the chosen conditioning approach. The routines also entail evaluations of the achieved covariance balance and common support. Third, we provided a step-by-step guide, comprising code to implement PSM in Stata.

We developed the concise checklist in Table 5, which provides business marketing researchers with a shortcut for their next PSM project. The checklist contains the application routines discussed in this article in one place.

Through this comprehensive guidance, we contribute by bridging the gap between the conceptual methodology literature on PSM and its practical application in the business marketing context. As noted, the PSM methodology literature spans across far more journals than business marketing researchers may read regularly, making it challenging to follow the method's recent developments. This challenge can promote inappropriate use and/or incomplete reporting of important tests pertaining to PSM—shortcoming that our review of prior PSM applications in business marketing revealed. Equipped with this guide, researchers will effectively navigate the method and be able to conduct reliable and impactful empirical research.

We anticipate that the trend to utilize advanced statistical techniques to account for endogeneity, such as through the use of PSM, will only grow due to the increasing awareness about the issue and its profound consequences in business marketing research. Our empirical step-by-step guide uses one example research context, B2B firms' acquisitions and performance effects, to demonstrate how scholars can effectively apply PSM. However, research opportunities to apply the method are abundant. In fact, consideration of the appropriate data contexts for the method (i.e., naturally occurring 'treatments') can inspire business marketing researchers to think of a new category of phenomena worth exploring. Specifically, PSM can be used to investigate the outcomes of discrete, significant decisions made by B2B firms, such as the adoption of innovative marketing strategies, supply chain practices, technologies, or internationalization strategies.

With regard to future methodological PSM research, our review revealed specific areas that require further research. First, PSM offers several conditioning approaches, each with its distinct trade-offs between preserving sample size and ensuring covariate balance in the treatment and control groups. However, with the exception of Guo et al. (2020), there is a lack of simulation studies that systematically assess the comparative strengths and weaknesses of PSM approaches across different data scenarios. Second and relatedly, approaches such as optimal matching and weighting enhance statistical power by ensuring full matching and retaining the original sample size. However, simulation study results using large samples show that these approaches underperform compared with other PSM alternatives due to lower quality matches. This raises the question of comparative performance in small samples, when preservation of sample size might be more consequential for inferences than matching quality. Third, the comparative performance of kernel-based matching against other PSM approaches is unexplored, creating a blind spot regarding its potential effectiveness. Fourth, some PSM assessment routines (e.g., check of whether certain types of firms are underrepresented in the matched sample) are relatively subjective and may yield different conclusions among researchers. Subjectivity and discretion in model choices can introduce bias (King & Nielsen, 2019). Hence, there is an opportunity for future research to develop and validate systematic workflows for these routines. Finally, matching approaches that are not based on propensity scores are used in the recent empirical literature, such as entropy balancing[5] (e.g., Kyaw et al., 2022; Le et al., 2023) or the synthetic control approach[6] (e.g., Guenther et al., 2024). Simulation studies are needed that compare these methods with the various PSM approaches. For other approaches, such as coarsened exact matching (CEM),[7] initial simulation results suggest that PSM methods perform equally well or better in various data scenarios (Guo et al., 2020). However, this depends on the specific PSM approach and the design choices made in CEM, such as the number of matching categories considered. Some studies show that CEM and other approaches such as Mahalanobis distance matching (MDM) that directly use covariates for matching, instead of aggregating them in a propensity score first, can reduce covariate imbalance and improve inferences (King & Nielsen, 2019). Further simulation studies are needed to compare the relative performance of these different matching approaches simultaneously, while accounting for the key design choices specific to each method.

CRediT authorship contribution statement

Peter Guenther: Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Investigation, Conceptualization. **Miriam Guenther:** Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Investigation, Conceptualization. **Shekhar Misra:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Mariia Koval:** Writing – original draft, Methodology, Investigation, Conceptualization. **Ghasem Zaefarian:** Writing – review & editing, Writing – original draft, Project administration, Conceptualization.

Appendix A. Appendix

Authors	Journal
Akrout and Diallo (2017)	*Industrial Marketing Management*
Ambulkar et al. (2023)	*Production and Operations Management*
Ba et al. (2022)	*Production and Operations Management*
Bai and Astvansh (2025)	*Production and Operations Management*
Belhadi et al. (2023)	*Journal of Business Research*
Bimpikis et al. (2020)	*Management Science*
Cheng et al. (2021)	*Journal of Business Research*
Claro et al. (2023)	*Industrial Marketing Management*
Craig et al. (2016)	*Manufacturing & Service Operations Management*
Friess and Kassemeier (2023)	*Journal of International Marketing*
Golovko et al. (2022)	*Journal of Business Research*
Habel et al. (2024)	*Organizational Behavior and Human Decision Processes*
Harmeling et al. (2015)	*Journal of Marketing*
Huang et al. (2012)	*Journal of Marketing Research*
Janani et al. (2022)	*Journal of the Academy of Marketing Science*
Leung and Sharma (2021)	*Journal of Business Research*
Mishra et al. (2020)	*Industrial Marketing Management*
Heredia Pérez et al. (2018)	*Technological Forecasting and Social Change*
Restuccia and Legoux (2019)	*Industrial Marketing Management*
Schmitz et al. (2019)	*Journal of Marketing*
Shi et al. (2017)	*Journal of Marketing*
Tinits and Fey (2022)	*Management International Review*
Vendrell-Herrero et al. (2021)	*Industrial Marketing Management*
Wang et al. (2023)	*Industrial Marketing Management*
Zhou and Wan (2022)	*Production and Operations Management*
Zou et al. (2024)	*Journal of Business & Industrial Marketing*

Data availability

Data will be made available on request.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association, 105*(490), 493–505.
Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica, 84*(2), 781–807.

[5] Entropy balancing reweights control group observations to ensure that observable characteristics are distributed as equally as possible between the treatment and control groups, based on distributional moment conditions (Hainmueller, 2012).

[6] The synthetic control approach creates an optimally weighted combination of control group observations (i.e., a synthetic control), instead of using actual control group observations, to approximate the treatment observations as closely as possible (Abadie et al., 2010). This approach is suitable for natural experiments, where some units are affected by an exogeneous intervention (e.g., a policy change) while others are not. The optimal weighting is determined so that the synthetic control best resembles the treatment units during the pre-intervention period on key observable outcome drivers.

[7] CEM coarsens observable characteristics to identify suitable matches of treatment and control observations. For instance, categorical characteristics are aggregated and continuous characteristics (e.g., firm size) are coarsened into broader categories based on natural breakpoints (e.g., small, medium and large firms based on governmental classification schemes). For an application example combining CEM with difference-in-differences, see Zhang and Tong (2020).

Akrout, H., & Diallo, M. F. (2017). Fundamental transformations of trust and its drivers: A multi-stage approach of business-to-business relationships. *Industrial Marketing Management, 66*, 159–171.

Ali, M. S., Groenwold, R. H., Belitser, S. V., Pestman, W. R., Hoes, A. W., Roes, K. C., … Klungel, O. H. (2015). Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: A systematic review. *Journal of Clinical Epidemiology, 68*(2), 112–121.

Ambulkar, S., Arunachalam, S., Bommaraju, R., & Ramaswami, S. (2023). Should a firm bring a supplier into the boardroom? *Production and Operations Management, 32*(1), 28–44.

Andrew, B. Y., Alan Brookhart, M., Pearse, R., Raghunathan, K., & Krishnamoorthy, V. (2023). Propensity score methods in observational research: Brief review and guide for authors. *British Journal of Anaesthesia, 131*(5), 805–809.

Angrist, J. D., Pischke, J., & r.-S.. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press.

Armstrong, C. S., Jagolinzer, A. D., & Larcker, D. F. (2010). Chief executive officer equity incentives and accounting irregularities. *Journal of Accounting Research, 48*(2), 225–271.

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine, 33*(6), 1057–1069.

Austin, P. C., Grootendorst, P., & Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: A Monte Carlo study. *Statistics in Medicine, 26*(4), 734–753.

Ba, S., He, S., & Lee, S.-Y. (2022). Mobile app adoption and its differential impact on consumer shopping behavior. *Production and Operations Management, 31*(2), 764–780.

Bai, M., & Astvansh, V. (2025). How and why does a business-to-business firm's corporate social responsibility disclosure impact its dependence on its major customers and major suppliers? *Production and Operations Management, 34*(1), 60–78.

Baillie, R. T., & DeGennaro, R. P. (1990). Stock returns and volatility. *The Journal of Financial and Quantitative Analysis, 25*(2), 203–214.

Belhadi, A., Kamble, S., Benkhati, I., Gupta, S., & Mangla, S. K. (2023). Does strategic management of digital technologies influence electronic word-of-mouth (ewom) and customer loyalty? Empirical insights from b2b platform economy. *Journal of Business Research, 156*, Article 113548.

Benedetto, U., Head, S. J., Angelini, G. D., & Blackstone, E. H. (2018). Statistical primer: Propensity score matching and its alternatives†. *European Journal of Cardio-Thoracic Surgery, 53*(6), 1112–1117.

Besley, T., & Case, A. (2000). Unnatural experiments? Estimating the incidence of endogenous policies. *The Economic Journal, 110*(467), F672–F694.

Bimpikis, K., Elmaghraby, W. J., Moon, K., & Zhang, W. (2020). Managing market thickness in online business-to-business markets. *Management Science, 66*(12), 5783–5822.

Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in new Jersey and Pennsylvania. *The American Economic Review, 84*(4), 772–793.

Certo, S. T., Busenbark, J. R., Woo, H.-S., & Semadeni, M. (2016). Sample selection bias and heckman models in strategic management research. *Strategic Management Journal, 37*(13), 2639–2657.

Cheng, L. T. W., Sharma, P., Shen, J., & Ng, A. C. C. (2021). Exploring the dark side of third-party certification effect in b2b relationships: A professional financial services perspective. *Journal of Business Research, 127*, 123–136.

Claro, D. P., Plouffe, C. R., & Vieira, V. A. (2023). Sales compensation plan type and sales opportunity coverage: "Double-edged" sword effects on sales performance. *Industrial Marketing Management, 113*, 153–167.

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics, 24*(2), 295–313.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J: L. Erlbaum Associates.

Craig, N., DeHoratius, N., & Raman, A. (2016). The impact of supplier inventory service level on retailer demand. *Manufacturing & Service Operations Management, 18*(4), 461–474.

Dehejia, R. (2005). Practical propensity score matching: A reply to smith and todd. *Journal of Econometrics, 125*(1), 355–364.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics, 84*(1), 151–161.

Delgado, M., & Mills, K. G. (2020). The supply chain economy: A new industry categorization for understanding innovation in services. *Research Policy, 49*(8), Article 104039.

Fama, E. F. (1990). Stock returns, expected returns, and real activity. *The Journal of Finance, 45*(4), 1089–1108.

Friess, M., & Kassemeier, R. (2023). Price increases and their financial consequences in international business-to-business selling. *Journal of International Marketing, 32*(1), 92–111.

Goldfarb, A., Tucker, C., & Wang, Y. (2022). Conducting research in marketing with quasi-experiments. *Journal of Marketing, 86*(3), 1–20.

Golovko, E., Lopes-Bento, C., & Sofka, W. (2022). Marketing learning by exporting – How export-induced marketing expenditures improve firm performance. *Journal of Business Research, 150*, 194–207.

Guenther, P., Guenther, M., Lukas, B. A., & Homburg, C. (2024). Consequences of marketing asset accountability—A natural experiment. *Journal of Marketing, 88*(5), 24–45.

Guo, S., Fraser, M., & Chen, Q. (2020). Propensity score analysis: Recent debate and discussion. *Journal of the Society for Social Work and Research, 11*(3), 463–482.

Habel, J., Kadić-Maglajlić, S., Hartmann, N. N., de Jong, A., Zacharias, N. A., & Kosse, F. (2024). Neuroticism and the sales profession. *Organizational Behavior and Human Decision Processes, 184*, Article 104353.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis, 20*(1), 25–46.

Hansen, B. B. (2008). The prognostic analogue of the propensity score. *Biometrika, 95*(2), 481–488.

Harmeling, C. M., Palmatier, R. W., Houston, M. B., Arnold, M. J., & Samaha, S. A. (2015). Transformational relationship events. *Journal of Marketing, 79*(5), 39–62.

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica, 66*(5), 1017–1098.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47*(1), 153–161.

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

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies, 64*(4), 605–654.

Heredia Pérez, J. A., Kunc, M. H., Durst, S., Flores, A., & Geldes, C. (2018). Impact of competition from unregistered firms on r&d investment by industrial sectors in emerging economies. *Technological Forecasting and Social Change, 133*, 179–189.

Hill, A. D., Johnson, S. G., Greco, L. M., O'Boyle, E. H., & Walter, S. L. (2020). Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research. *Journal of Management, 47*(1), 105–143.

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica, 71*(4), 1161–1189.

Ho, D., Imai, K., King, G., & Stuart, E. A. (2011). Matchit: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42*(8), 1–28.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945–960.

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.

Hullsiek, K. H., & Louis, T. A. (2002). Propensity score modeling strategies for the causal analysis of observational data. *Biostatistics, 3*(2), 179–193.

Janani, S., Christopher, R. M., Nikolov, A. N., & Wiles, M. A. (2022). Marketing experience of ceos and corporate social performance. *Journal of the Academy of Marketing Science, 50*(3), 460–481.

Kang, J. D. Y., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science, 22*(4), 523–539.

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis, 27*(4), 435–454.

Kline, A., & Luo, Y. (2022). Psmpy: A package for retrospective cohort matching in python. In *Paper presented at the 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*.

Kyaw, K., Treepongkaruna, S., & Jiraporn, P. (2022). Board gender diversity and environmental emissions. *Business Strategy and the Environment, 31*(7), 2871–2881.

Lambkin, M. C., & Muzellec, L. (2010). Leveraging brand equity in business-to-business mergers and acquisitions. *Industrial Marketing Management, 39*(8), 1234–1239.

Langworthy, B., Wu, Y., & Wang, M. (2022). An overview of propensity score matching methods for clustered data. *Statistical Methods in Medical Research, 32*(4), 641–655.

Le, H., Kibria, I., & Jiang, K. (2023). Is chief executive officer optimistic belief bad for workers? Evidence from corporate employment decisions. *Human Resource Management Journal, 33*(3), 748–762.

Leung, T. Y., & Sharma, P. (2021). Demystifying the dark side of board political capital. *Journal of Business Research, 126*, 307–318.

Li, K., & Prabhala, N. R. (2008). Self-selection models in corporate finance. In B. E. Eckbo (Ed.), *Handbook of empirical corporate finance* (pp. 37–86). Elsevier B.V.

Li, M. (2012). Using the propensity score method to estimate causal effects: A review and practical guide. *Organizational Research Methods, 16*(2), 188–226.

Linden, A., & Adams, J. L. (2012). Combining the regression discontinuity design and propensity score-based weighting to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice, 18*(2), 317–325.

Lohr, S. L. (2022). *Sampling design and analysis* (Vol. 3). Boca Raton: CRC Press.

Makepeace, G., & Peel, M. J. (2013). Combining information from heckman and matching estimators: Testing and controlling for hidden bias. *Economics Bulletin, 33*(3), 2422–2436.

Mishra, S., Ewing, M. T., & Pitt, L. F. (2020). The effects of an articulated customer value proposition (cvp) on promotional expense, brand investment and firm performance in b2b markets: A text based analysis. *Industrial Marketing Management, 87*, 264–275.

Myers, J. A., Rassen, J. A., Gagne, J. J., Huybrechts, K. F., Schneeweiss, S., Rothman, K. J., … Glynn, R. J. (2011). Effects of adjusting for instrumental variables on bias and precision of effect estimates. *American Journal of Epidemiology, 174*(11), 1213–1222.

Palmatier, R. W., Miao, C. F., & Fang, E. (2007). Sales channel integration after mergers and acquisitions: A methodological approach for avoiding common pitfalls. *Industrial Marketing Management, 36*(5), 589–603.

Peikes, D. N., Moreno, L., & Orzol, S. M. (2008). Propensity score matching: A note of caution for evaluators of social programs. *The American Statistician, 62*(3), 222–231.

Restuccia, M., & Legoux, R. (2019). B2b relationships on the fast track: An empirical investigation into the outcomes of solution provision. *Industrial Marketing Management, 76*, 203–213.

Richey, R. G., Kiessling, T. S., Tokman, M., & Dalela, V. (2008). Market growth through mergers and acquisitions: The role of the relationship marketing manager in sustaining performance. *Industrial Marketing Management, 37*(4), 394–406.

Rosenbaum, P. R. (2002). Constructing matched sets and strata. In P. R. Rosenbaum (Ed.), *Observational studies* (pp. 295–331). New York, NY: Springer New York.

Rosenbaum, P. R. (2020). Modern algorithms for matching in observational studies. *Annual Review of Statistics and Its Application, 7*(1), 143–176.

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. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association, 100*(469), 322–331.

Rubin, D. B., & Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics, 52*(1), 249–264.

Ryan, A. M., Burgess, J. F., & Dimick, J. B. (2015). Why we should not be indifferent to specification choices for difference-in-differences. *Health Services Research, 50*(4), 1211–1235.

Schmitz, C., Friess, M., Alavi, S., & Habel, J. (2019). Understanding the impact of relationship disruptions. *Journal of Marketing, 84*(1), 66–87.

Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: The matching package for r. *Journal of Statistical Software, 42*(7), 1–52.

Shi, H., Sridhar, S., Grewal, R., & Lilien, G. (2017). Sales representative departures and customer reassignment strategies in business-to-business markets. *Journal of Marketing, 81*(2), 25–44.

Shipman, J. E., Swanquist, Q. T., & Whited, R. L. (2017). Propensity score matching in accounting research. *Accounting Review, 92*(1), 213–244.

Stern, I., Deng, X., Chen, G., & Gao, H. (2021). The "butterfly effect" in strategic human capital: Mitigating the endogeneity concern about the relationship between turnover and performance. *Strategic Management Journal, 42*(13), 2493–2510.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1–21.

Swani, K., Brown, B. P., & Mudambi, S. M. (2020). The untapped potential of b2b advertising: A literature review and future agenda. *Industrial Marketing Management, 89*, 581–593.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology, 51*(6), 309–317.

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research, 46*(1), 90–118.

Tinits, P., & Fey, C. F. (2022). The effects of timing and order of government support mechanisms for sme exports. *Management International Review, 62*(2), 285–323.

Ullah, S., Akhtar, P., & Zaefarian, G. (2018). Dealing with endogeneity bias: The generalized method of moments (gmm) for panel data. *Industrial Marketing Management, 71*, 69–78.

Vendrell-Herrero, F., Bustinza, O. F., & Vaillant, Y. (2021). Adoption and optimal configuration of smart products: The role of firm internationalization and offer hybridization. *Industrial Marketing Management, 95*, 41–53.

Wang, X., Wei, R., Liu, Y., Xia, H., & Zhao, Y. (2023). The effects of relational knowledge emphasis on new product development strategy. *Industrial Marketing Management, 109*, 257–270.

Wang, Y., Cai, H., Li, C., Jiang, Z., Wang, L., Song, J., & Xia, J. (2013). Optimal caliper width for propensity score matching of three treatment groups: A Monte Carlo study. *PLoS One, 8*(12), Article e81045.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge: MIT Press.

Zaefarian, G., Kadile, V., Henneberg, S. C., & Leischnig, A. (2017). Endogeneity bias in marketing research: Problem, causes and remedies. *Industrial Marketing Management, 65*, 39–46.

Zhang, X. F. (2006). Information uncertainty and stock returns. *The Journal of Finance, 61*(1), 105–137.

Zhang, Y., & Tong, T. W. (2020). How vertical integration affects firm innovation: Quasi-experimental evidence. *Organization Science, 32*(2), 455–479.

Zhou, Z., & Wan, X. (2022). Does the sharing economy technology disrupt incumbents? Exploring the influences of mobile digital freight matching platforms on road freight logistics firms. *Production and Operations Management, 31*(1), 117–137.

Zou, H., Xie, E., & Mei, N. (2024). Political connections and firms' trade credit in emerging economies. *Journal of Business & Industrial Marketing, 39*(3), 633–650.