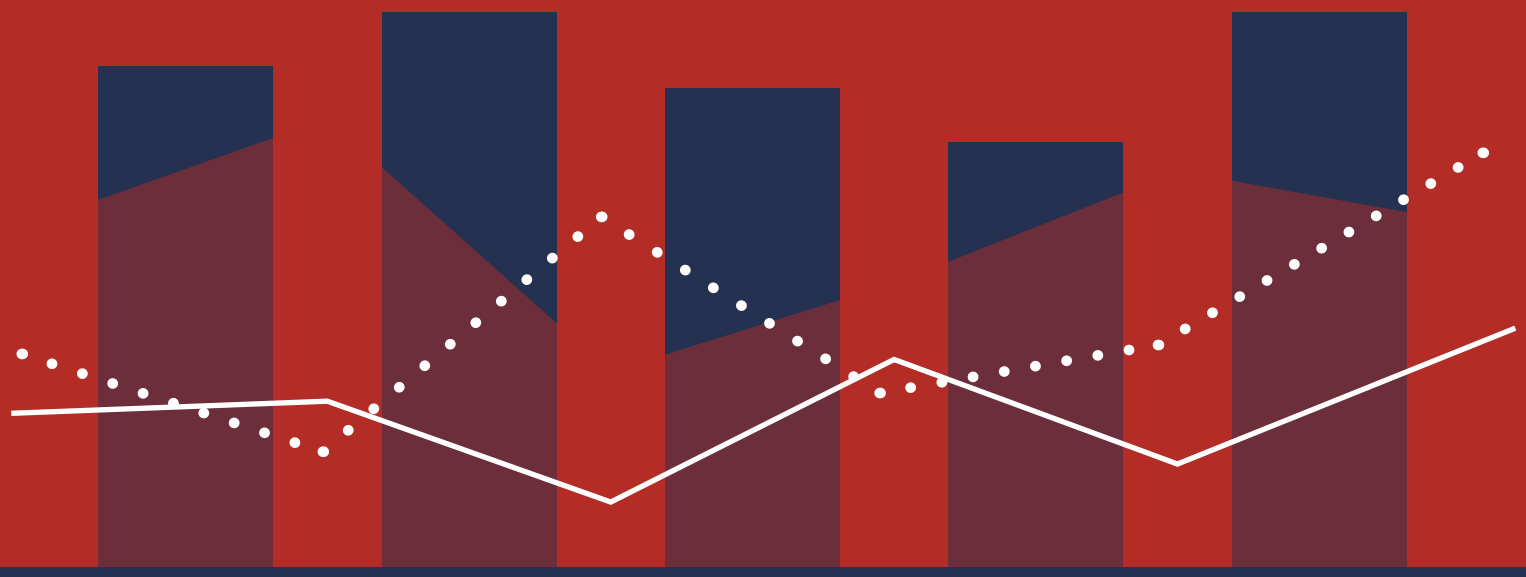
JOURNAL *of* MARKETING RESEARCH

# IMPACT@JMR: CURATIONS

VOLUME 1

## Identification-Related Challenges and Recommendations

EDITED BY
Kapil Tuli, Coeditor, *Journal of Marketing Research*
Rebecca Hamilton, Editor in Chief, *Journal of Marketing Research*



AMA | AMERICAN MARKETING ASSOCIATION

Printed in the United States of America

ISBN-13: 978-0-87757-025-7

# Foreword

**Kapil Tuli and Rebecca Hamilton**

Although randomized experimental studies are the gold standard for making causal inferences in scientific research, in several critical contexts, randomized experiments are neither feasible nor ethical (see, for example, Narang, Shankar, and Narayanan 2025; Shaik et al. 2025; Wu, Watson, and Faraji-Rad 2025). Fortunately, developments in econometrics and statistics present an array of methods that can allow marketing researchers to address identification challenges presented by observational data and present findings that are rigorously derived (see Angrist and Pischke 2010). Accordingly, this compendium combines three issues of Impact@JMR that provide a detailed discussion of three methodological approaches, their underlying assumptions, and decision points. This compendium can serve as a useful starting point for authors interested in making causal inferences when working with nonexperimental observational data.

The first piece, written by Li, Luo, and Pattabhiramaiah (2025) presents a detailed discussion of key aspects for scholars to consider when using quasi-experimental data. The authors present a detailed decision tree that outlines the specific methods and underlying assumptions that scholars should consider based on the nature of the exogenous shocks they are exploring and the data available for them. They conclude their discussion by identifying developments in machine learning methods that can be used to complement extant methods.

Grewal and Orhun (2025) present the second piece, which focuses on the classical instrumental variables (IV) approach to mitigate identification concerns related to omitted variables bias. The authors start by discussing the five key assumptions of the IV approach and then illustrate the importance of these assumptions by presenting two detailed examples from recent research. Drawing on the examples, the chapter also provides readers with a table outlining a novel way in which scholars can assess the rigor of their instruments based on the assumptions underlying the IV approach.

While quasi-experimental design and IV are traditional approaches to address identification concerns, an increasing number of studies in marketing and other disciplines (e.g., management and organizational behavior) are leveraging copulas as a statistical tool. Copulas, multivariate cumulative distribution functions for which each variable has a uniform marginal distribution on the interval [0, 1], can be used to model correlations among variables. Using the copula approach requires a clear understanding of the underlying assumptions (see for example, critiques by Becker, Proksch, and Ringle 2022; Eckert and Hohberger 2022). The third piece in this collection, by Park and Gupta (2025), outlines the assumptions of the traditional copula approach and identifies the different types of copulas authors should use based on the assumptions applicable to their specific context.

# References

Angrist, Joshua D., and Jörn-Steffen Pischke, (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24 (2), 3–30.

Becker, Jan-Michael, Dorian Proksch, and Christian M. Ringle (2022), "Revisiting Gaussian Copulas to Handle Endogenous Regressors," *Journal of the Academy of Marketing Science*, 50, 46–66.

Eckert, Christine and Jan Hohberger (2022), "Addressing Endogeneity Without Instrumental Variables: An Evaluation of the Gaussian Copula Approach for Management Research," *Journal of Management*, 49 (4), 1460–95.

Narang, Unnati, Venkatesh Shankar, and Sridhar Narayanan (2025), "Cross-Channel Effects of Failure in a Retailer's Mobile App," *Journal of Marketing Research* (published online January 30), https://doi.org/10.1177/00222437251318660.

Shaik, Muzeeb, John P. Costello, Mike Palazzolo, Adithya Pattabhiramaiah, and Shrihari Sridhar (2025), "How Fatal School Shootings Impact a Community's Consumption," *Journal of Marketing Research* (published online June 4), https://doi.org/10.1177/00222437251350150.

Wu, Yuechen, Jared Watson, and Ali Faraji-Rad (2025), "Who Will I Be With(out) You? Consequences of Perceived Romantic Relationship Status Stability on Product Rentals," *Journal of Marketing Research* (published online July 1), https://doi.org/10.1177/00222437251358160.

# Contributors

Rajdeep Grewal is Townsend Family Distinguished Professor and Area Chair of Marketing, Kenan-Flagler Business School, University of North Carolina, Chapel Hill, NC.

Sachin Gupta is Henrietta Johnson Louis Professor of Management and Professor of Marketing, Cornell University, USA.

Kathleen T. Li is Assistant Professor of Marketing, University of Texas at Austin, USA.

Lan Luo is Professor of Marketing, University of Southern California, USA.

Yesim Orhun is Associate Professor of Marketing and Michael R. and Mary Kay Hallman Fellow, Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI.

Sungho Park is Korbit Chaired Professor of Marketing, Seoul National University, South Korea.

Adithya Pattabhiramaiah is Associate Professor of Marketing, Georgia Institute of Technology, USA.

# Causal Inference with Quasi-Experimental Data

**Kathleen T. Li, Lan Luo and Adithya Pattabhiramaiah | 11.13.2023**

In this article, we attempt to overview the methodological toolkit available to empirical researchers who are interested in making causal inference using quasi-experimental data. In particular, Figure 1 provides an overview of the type of data available to researchers (e.g., randomized treatment, rich or constrained availability of observables, small or large number of time periods or treatment units) and describes corresponding suitable approaches, along with some pros and cons involved in their tactical use.

In marketing, randomized experimentation represents the gold standard for making causal inference using empirical data. In an ideal setting, we would randomly assign participants to different groups to receive varying types or levels of treatment. A rich history of research, including work published in the Journal of Marketing Research, has used randomized experimental designs for causal inference (see Ghose et al. [2024] and Cao, Chintagunta, and Li [2023] for recent examples). Nonetheless, there are many marketing-relevant settings where researchers do not have access to experimental data, or where running such experiments is too expensive or infeasible. Additionally, ethical considerations frequently preclude randomly assigning treatments, such as instances where it could lead to harm or deprive participants of necessary care, as in the case of life-saving treatments or medications. When randomization of treatment is not possible, one may have to rely on enhanced "statistical rigor" to compensate for the deficiencies in "design rigor."

In the remainder of this article, we review the common challenges pertaining to making causal inference with quasi-experimental data and discuss recent advances in helping alleviate them. Specifically, we focus on discussing methods that emphasize matching units based on the outcome variable (Y). Then, we provide an overview of methods that focus on matching on observable covariates (X). Lastly, we conclude by reflecting on our recommendations and discussing future research in related areas.

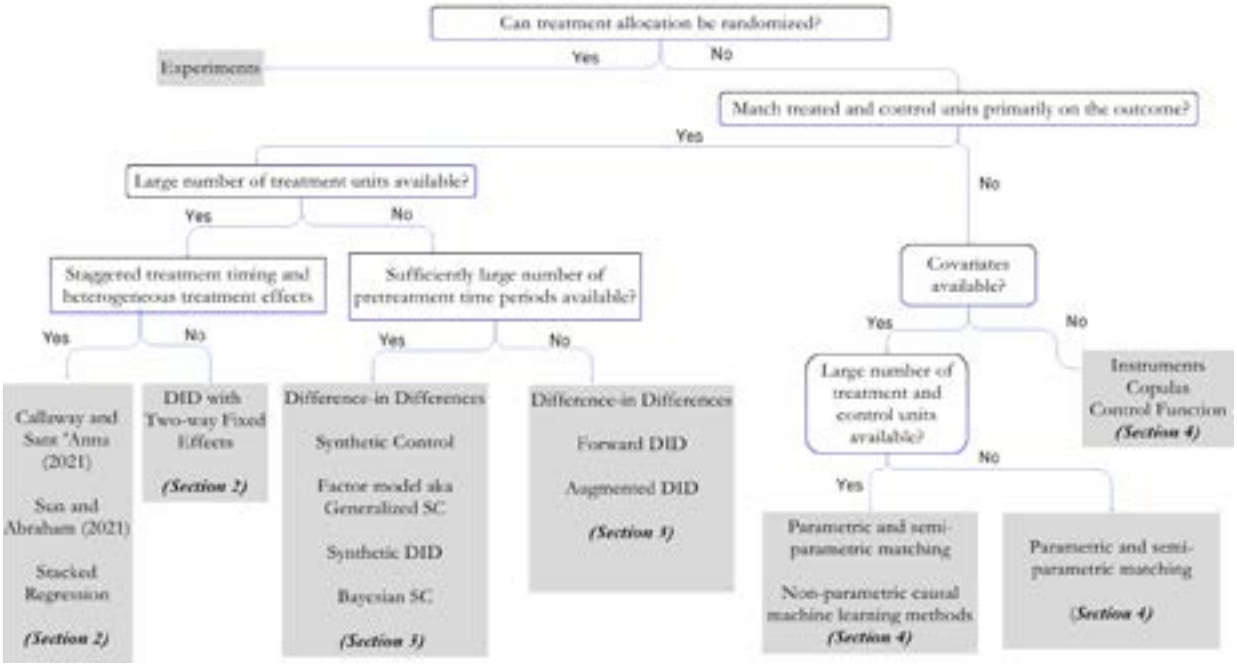
## Difference-in-Differences

In (the many) situations where researchers do not have the luxury of assigning units into treatment and control groups, they can still understand causal effects by leveraging quasi-experimental methods. The difference-in-differences (DID) method is the most widely used quasi-experimental method. It can be used in data settings with treatment and control units and pre- and post-treatment time periods.

Here, we begin by describing the simplest DID design where all treatment units are treated at the same time. We observe treatment and control units over time so yit is the outcome for unit i at time t. The DID model can be estimated using the following regression model:

$$y_{it} = \beta_1 + \beta_2 \text{Treat}_i + \beta_3 \text{Post}_t + \beta_4 \text{Treat}_i \text{Post}_t + x_i\tilde{\beta} + \epsilon_{it}. \quad (1)$$

where $\text{Treat}_i$ is a treatment indicator that takes a value of 1 if unit i belongs to the treatment group and 0 if it belongs to the control group, Postt is a posttreatment time period indicator that takes a value of 1 if time period t is in the post-treatment time period and 0 otherwise, $x_i$ is a k-dimensional vector of time invariant observable covariates, $\tilde{\beta} = (\beta_5, \ldots, \beta_5 + k)'$, and




$\epsilon_{it}$ is an error term. The coefficient β4 is the causal effect of interest, which is the average treatment effect (ATE) or average treatment effect on the treated (ATT).

## *Assumptions*

Before choosing the DID identification strategy, researchers first need to assess whether the identifying parallel trends assumption holds. The DID parallel trends assumption states that the treatment unit would have followed a path parallel to the control units in the absence of treatment. We make two observations about this assumption. First, the DID method can be interpreted as a method that primarily matches on outcomes. Although covariates can be included in the DID regression model, the main goal is to use the control units' outcomes to match the treatment unit's outcome during the pretreatment period, and then predict the treatment counterfactual and the ATT. Second, since the parallel trends assumption is a statement about the treatment counterfactual, we cannot directly test the parallel trends assumption. However, what we can do is test whether the treatment and control units followed parallel trends in the pretreatment period (parallel pretrends assumption). This is essentially the testable part of the parallel trends assumption. There are two popular approaches to check the testable part of the parallel trends assumption: (1) visual inspection and (2) statistical tests. Visual inspection involves plotting the treatment and control trends in the pretreatment period and inspecting whether they look parallel. Statistical tests "formalize" this evaluation somewhat, by testing whether the difference in mean outcomes of the treatment and control group, for each pretreatment period, are statistically different from a constant. However, statistical tests often have low power (i.e., they may fail to reject the null hypothesis of no difference even when the parallel pretrends assumption is violated), implying that visual inspection can be the only viable way of assessing whether the parallel pretrends assumption holds.

If the parallel pretrends assumption of the DID method is violated, researchers can attempt to use matching methods (described in more detail in the "Selection on Variables" section ) to first identify a subset of control units that are

more similar to the treatment unit on covariates and then apply the DID method. Alternatively, if the number of treatment units is not very large (e.g., less than 100), researchers can apply the synthetic control or related methods (described in more detail in the "Synthetic Control and Related Methods" section).

Another assumption underlying the use of DID methods (and many causal inference methods) is the stable unit treatment value assumption (SUTVA). SUTVA can be decomposed into two parts. The first part of SUTVA (no interference) requires that treatment applied to one unit does not affect the outcome of other units. To understand this better, let us consider the case where a state experiences a treatment (e.g., enactment of a local tax law). This should mean that the treatment should not affect outcomes in the control states or other treatment states (or vice versa). Researchers can use logical argumentation based on institutional knowledge to justify this assumption (e.g., geographical variation in treatment and control units makes interference unlikely). To the extent that researchers have access to more granular data, they can also check and confirm that there is no movement of individuals across treatment and control states, patterns that support the case for no interference. The second part of SUTVA (no hidden variations of treatment) requires that for each unit, there are no different forms or versions of each treatment level that may lead to different potential outcomes. Researchers can use institutional knowledge about the treatment itself as justification for this assumption.

## *Staggered Treatment Timing*

As researchers, we are commonly faced with treatments that apply to different units at different times. This is referred to as differential timing of treatment or staggered treatment timing. When treatment effects are homogeneous over time, the following regression equation can be used to estimate the DID model:

$$y_{it} = \beta_1 + \beta_2 Treat_i Post_t + x_i\tilde{\beta} + FE_i + FE_t + \epsilon_{it}. \quad (2)$$

This model is called a two-way fixed effects (TWFE) model because it contains fixed effects for both unit (FEi) and time (FEt). However, the homogeneity in treatment response can be a restrictive assumption, as it is common that the treatment effect may change over time or differ across treatment units.

Goodman-Bacon (2021) identified that for the case of heterogeneous treatment effects and staggered treatment timing, the conventional TWFE model breaks down (i.e., yields a biased average treatment effect). The intuition is that the conventional TWFE model uses a weighted average of all potential DID estimates obtained using different combinations of groups of treatment and control units (treated, not yet treated, and never treated groups). However, a group that has already been treated can only be used as a control group for a group that is treated later, in the case of time-invariant treatment effects. In other words, the already treated group is not a "clean" control when the treatment effect varies over time (we also refer to this as heterogeneous or time-variant treatment effects). Because the standard TWFE model uses control units that are not "clean," it is biased when the treatment effect varies over time. While Goodman-Bacon (2021) helps the researcher identify the specific sources of bias in their setting by decomposing the standard DID estimator into different underlying comparisons (e.g., early vs. late treated, treated vs. never treated), this paper does not provide a single, unbiased estimate of the treatment effect.

To solve the problem identified in Goodman-Bacon (2021), scholars have proposed many estimators. We discuss three proposed solutions. The first solution is an estimator proposed by Callaway and Sant'Anna (2021). It is perhaps the most popular and widely used. The main idea is to first estimate the ATT for each treatment group cohort and then use a weighted average of those ATTs. Specifically,

ATT(g) is the ATT for the treatment group cohort that first receives treatment at time period g. To estimate ATT(g), we can use a simple DID setup, where the treatment group is the treatment group cohort g and the control group are units that are not yet treated or never treated (excluding units that are already treated). Then, the overall ATT is a weighted average of all the ATT(g)'s estimated over different time periods. Callaway and Sant'Anna (2021) can accommodate an outcome-regression-based estimator (Heckman, Ichimura, and Todd 1997), an inverse probability weighted estimator (Abadie 2005), and a doubly robust estimator (Sant'Anna and Zhao 2020).

The second solution, which is closely related to Callaway and Sant'Anna (2021), is an estimator proposed by Sun and Abraham (2021). This estimator is an extension of Callaway and Sant'Anna (2021) to the case of dynamic treatment effects, where researchers are interested in estimating separate treatment effects for each posttreatment time period. This setting experiences the same problem that already treated units are not "clean" controls when the treatment effect is heterogeneous. If researchers need to calculate a treatment effect estimate for every posttreatment period, then they should use Sun and Abraham (2021).

The third solution is the "stacked regression" proposed by applied researchers (Cengiz et al. 2019; Gormley and Matsa 2011). The main idea is to create separate clean datasets of treatment groups and "clean" control groups, stack them by aligning the intervention time period, and use the DID TWFE regression with dataset and time fixed effects. While in practice, this solution is the simplest to implement, one may need to be cautious, as the sample average ATT may be inconsistent (Baker, Larcker, and Wang 2022).

Given the discussions above, we provide the following recommendations. First, if all treatment units receive treatment at the same time (no staggered timing), use the standard DID TWFE model. However, if there is staggered treatment timing, use one of the proposed solutions and justify the choice of clean controls. Finally, as a robustness check, researchers can separately estimate a DID TWFE for each treatment cohort group using clean controls (e.g., the never treated group) (Baker, Larcker, and Wang 2022). Note that each of these separate analyses does not suffer from the issues that affect staggered DID with heterogeneous treatment effects because each analysis only examines one treatment cohort at a time (no staggered treatment timing). While this robustness check does not result in an aggregated treatment effect, it may still be informative to show what the separate ATTs are for each treatment group cohort.[1]

For a more detailed discussion of these three solutions among many others, we refer readers to two review papers: "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometric Literature" by Roth et al. (2023) and "How Much Should We Trust Staggered Difference-in-Differences Estimates?" by Baker, Larcker, and Wang (2022). Note that all of the methods discussed in these two articles deal with settings characterized by a large number of treatment and control units and relatively short time periods.

## Synthetic Control and Related Methods

While the DID method is the most popular quasi-experimental method, it is often not

[1]Researchers looking to overcome issues related to heterogeneous treatment effects in staggered treatment settings and interpret the composite treatment effect across all cohorts might also consider the synthetic control method that we discuss in the next section. Synthetic control methods do not suffer from this problem, as they estimate a separate ATT for each treatment group and use "clean" controls (units that have never been treated) for computing the ATT for the entire sample.

viable, due to obvious observed violations in the parallel pretrends assumption. To overcome this, there has been a recent surge in flexible alternative estimators that are more widely applicable than DID. The most famous of such methods is the synthetic control method proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010). The synthetic control method has been called "arguably the most important innovation in the evaluation literature in the last fifteen years" by Susan Athey and Guido Imbens (the latter is winner of the 2021 Nobel Prize in Economics for making "methodological contributions to the analysis of causal relationships").

The synthetic control method uses a weighted average of the control units (instead of a simple average used in the DID) to predict the treatment counterfactual and the ATT. The synthetic control method, like DID, primarily matches on outcomes. It achieves this by using a weighted average of control units' outcomes to match the treatment unit's outcomes during the pre-treatment period. This approach better matches the treatment outcomes in the pre- treatment period, which then improves the prediction of the counterfactual and consequently, the estimate of the average treatment effect on the treated (ATT). However, the guidance for conducting proper inference using the synthetic control method is not clear. Previously, researchers had to rely on placebo tests, which entail making restrictive assumptions that are often violated. Li (2020) developed the inference theory for the synthetic control method, which allows the calculation of confidence intervals and quantifying uncertainty using a subsampling procedure.

While the synthetic control method is a very powerful new tool, it still has some associated restrictions. First, it is unable to easily accommodate settings involving a large number of treated units. Second, it is less suited for handling situations wherein the treatment and control units are very different from one another (e.g., situations where the outcome for the treatment unit is outside the range of that for the control units). Such settings call for more flexible methods, the most popular of which is the factor model.

In many marketing contexts, researchers have access to a large number of control units. The factor model in particular has gained traction in marketing due to its ability to elegantly handle such situations via dimension reduction. Conveniently, the dimension reduction of the factor model also serves as implicit regularization to prevent overfitting. The factor model is also known as generalized synthetic control (Xu 2017) or interactive fixed effects model (Chan and Kwok 2016; Gobillon and Magnac 2016).

Marketing researchers have applied this method in a variety of settings ranging from policy evaluation (Guo, Sriram, and Manchanda 2020; Pattabhiramaiah, Sriram, and Manchanda 2019) to advertising effects measurement (Lovett, Peres, and Xu 2019). How should researchers quantify uncertainty when using the synthetic control in combination with the factor model? Past research has recommended the use of a bootstrap procedure (Xu 2017), which can be restrictive. Li and Sonnier (2023) show that the bootstrap procedure provided in Xu (2017) often results in biased confidence intervals that are either too narrow or too wide, leading to false precision or false imprecision. False precision may lead researchers to erroneously conclude that they detected a true effect, whereas false imprecision may lead researchers to erroneously conclude that there was no detectable true effect. Both mistakes—false positives and false negatives—can lead to incorrect business decisions. Following the inference theory in Li and Sonnier (2023), researchers can correctly quantify uncertainty of causal effects to make more informed business decisions.

The synthetic DID method proposed by Arkhangelsky et al. (2021) is another flexible

quasi-experimental method that has gained traction in marketing. Marketing scholars have used synthetic DID to study the effect of TV advertising on online browsing and sales (Lambrecht, Tucker, and Zhang 2024) and the effect of soda tax on marketing effectiveness (Keller, Guyts, and Grewal 2024). The synthetic DID method proposes a general framework that uses both individual weights and time weights for additional flexibility. To conduct inference and compute standard errors, Arkhangelsky et al. offer three alternative procedures: (1) block bootstrap, (2) jackknife, and (3) permutation. Block bootstrap and jackknife require a large number of treatment units, without which the estimated confidence intervals may be unreliable (Clarke et al. forthcoming). Permutation does not have any restriction on the number of treatment units, but it requires a moderate to large number of control units and requires that the treatment unit and control units' variance be similar.

Next, we overview two additional quasi-experimental methods. The first is the ordinary least squares (OLS) method proposed by Hsiao, Ching, and Wan (2012) (which is also called the HCW method). The OLS method can be used when the number of control units is (much) smaller than the pretreatment time periods. Due to the increased flexibility of both the OLS method and synthetic DID methods, it is even more important when using these methods to check for overfitting using the backdating exercise described the last paragraph of this section. Another method is the matrix completion method (for additional details, see Athey and Imbens [2019] and Bai and Ng [2021]). The matrix completion method imputes missing values in a panel data setup to estimate counterfactuals when potential outcomes have a factor structure to estimate the ATT. In other words, this approach estimates missing values in a dataset by using principal components analysis, which can be used to estimate the effects of a treatment when some of the potential outcomes are missing (for a recent marketing application of this method, see Bronnenberg, Dubé, and Sanders [2020]).

All the methods described so far are frequentist methods. However, each of these methods can be estimated using a Bayesian framework. Kim, Lee, and Gupta (2020) propose a Bayesian synthetic control method that uses Bayesian shrinkage priors to solve the sparsity problem and conduct inference. Pang, Liu, and Xu (2022) propose a Bayesian factor model that uses a Bayesian shrinkage method for model searching and factor selection.

The synthetic control method and related flexible alternative methods (e.g. factor model method, synthetic DID method, OLS method, matrix competition method) we have discussed thus far require access to a sufficiently long pre-treatment time window (e.g. at least ten pretreatment time periods). However, what if researchers need a flexible alternative to DID but do not have access to a sufficient number of time periods before the treatment occurs? To fill this gap, researchers can consider using the augmented DID or forward DID methods. Specifically, if the outcome for the treatment unit is outside of the range of that of the control units, researchers can use the augmented DID method, which uses a scaled average of the control units to construct the treatment counterfactual (Li and Van den Bulte 2023). On the other hand, if the outcome for treatment unit is within the range of that of the control units, researchers can consider using the forward DID method, which uses a forward selection algorithm to select a relevant subset of control units and then applies the DID method (Li 2024).

Building on recent advances in the literature studying flexible alternatives to DID, we recommend following two best practices when implementing the synthetic control and related methods. First, after applying the method, visually inspect whether the parallel trends assumption of the corresponding method holds in the pretreatment window by

plotting the outcome variable corresponding to the treatment unit(s) and that of the fitted in-sample curve, which is created using the control units. If the parallel pretrends assumption does not hold, do not adopt the method. If the parallel pretrends assumption does hold, then continue to conduct a backdating (out-of-sample prediction) exercise to check for overfitting (Abadie 2021; Li 2020; Li and Sonnier 2023). We recommend only using the methods that satisfy both best practices.

Our discussion thus far has focused on methods aimed primarily at matching treated and control units on the outcome variable (Y). In the next section, we discuss approaches focused on matching on observable covariates (X).

## Selection on Observables (and Consequently, on Unobservables)

As we have noted, one main challenge in estimating causal effects from observational studies is the presence of confounding factors that simultaneously affect the treatment status and the outcome of interest. In some observational studies, the allocation of treatment can be presumed to resemble random assignment (e.g., a policy change determined independently from the outcomes being measured). However, such situations are rare because forces such as consumer incentives, firm objectives, or regulation can threaten such a pure treatment exogeneity argument. For such reasons, researchers often find themselves in two states of the world—one where covariates are not available, and one where they are. In situations where suitable covariates may not be available, researchers may consider methods such as instruments, copulas, and control function to infer causality using observational data. We refer readers to Wooldridge (2019), Petrin and Train (2010), Park and Gupta (2012), and Danaher and Smith (2010) for additional details.

However, in our information-rich era, researchers generally have abundant covariates at hand. Oftentimes, researchers observe many/most confounding factors that could simultaneously affect the treatment status and the outcome of interest. For example, Ellickson, Kar, and Reeder (2023) consider the case of observational studies using data from targeted marketing campaigns, where, following the definition of the targeting rule, the treatment assignment is determined based on some observable demographic or behavioral covariates. Thus, there are many observational studies where the unconfoundedness (or selection on observables) assumption is satisfied, and researchers can adopt methods to adjust for the observed confounders.

### *Matching on Covariates*

Matching methods, in simple terms, aim to pair/match units with similar covariates but different treatment statuses to estimate the treatment effects by comparing their outcomes. Some well-known traditional methods aimed at adjusting for observed confounders are parametric matching, propensity scores, and weighting methods. All these methods require that the researcher has knowledge about which covariates are important a priori. The identification of the observed confounders and the selection of the variables that represent them is usually based on economic theories, institutional knowledge, or intuition (e.g., targeting ads depends on consumer engagement on a website).

Propensity score matching has been a commonly used matching method in marketing for decades, although their viability has recently been called into question due to the technique's sensitivity to parametric assumptions (Athey and Imbens 2017). These methods start with the estimation of the propensity score (i.e., the probability of receiving the treatment conditional on covariates ($e_i = P(T_i = 1|X_i)$), which can be later combined with matching, stratification,

inverse probability weighting, or covariate adjustment (Austin 2011). Another popular method to estimate treatment effects under the unconfoundedness assumption is the augmented inverse probability weighting approach (Robins, Rotnitzky, and Zhao 1994), which combines regression models to estimate the potential outcomes with inverse propensity score weighting methods. One attractive property of this estimator is its robustness to bias or misspecifications in either the potential outcome estimate or the propensity score estimate (Bang and Robins 2005). See Gordon et al. (2019) for a recent example in marketing that illustrates such applications in the context of online advertising.

Despite their popularity, matching methods have limitations, especially when the number of covariates is very large. In such cases, conventional methods such as exact-matching might become infeasible, and nearest-neighbor matching can result in a biased estimate of the ATT (Abadie and Imbens 2006). To obtain a flexible specification of the propensity score when the number of covariates is large, we can apply variable reduction methods such as Lasso (e.g., Gordon et al. 2019) or penalized logistic regression (e.g., Eckles and Bakshy 2021). Additionally, we can use machine learning (ML) methods to reduce the dimension of the covariate space. For example, Li et al. (2016) illustrate the usefulness of using linear dimensionality reduction ML algorithms such as principal component analysis (PCA), locality preserving projections (LPP), and random projections before matching the treatment and control units. Ramachandra (2018) explores the use of auto-encoders as a dimensionality reduction technique prior to neighbor matching on simulated data. In a similar vein, Yao et al. (2018) develop a method based on deep representation learning that jointly preserves the local similarity information and balances the distributions of the control and the treated groups.

Additionally, Diamond and Sekhon (2013) propose GenMatch, a multivariate matching method based on genetic algorithm to iteratively check and improve covariate balance between the treated and the control groups. Zubizarreta (2015) also proposes a weighting method that allows researchers to prespecify the level of desired balance between the treated and the control groups. One advantage of this weighting method is that it runs in polynomial time, so large datasets can be handled quickly.

## *Causal Machine Learning Methods on Flexible Matching*

While matching on covariates can be highly powerful in many research settings, the methods discussed in the previous section often require that the researchers have prior knowledge about which covariates are important and which functional form is the most suitable for capturing their influence on the outcome variables. However, when working in high-dimensional settings, it might become difficult for researchers to identify which specific covariates are important (e.g., number of clicks, time spent on different sections of the website) or which functional form is appropriate for modeling their influence on outcomes (linear, quadratic, or more flexible specifications). Meanwhile, including all the covariates or allowing for flexible functional forms may reduce the power available in the dataset for learning about the treatment effect of interest (Chernozhukov et al. 2018). In such cases, researchers can benefit from adopting causal ML methods for flexible matching. Next, we discuss some commonly used causal ML methods often used on observational data for inferring causality. These methods are especially helpful in settings that involve high-dimensional covariates and/or when the relationship between them cannot be satisfactorily modeled in a parametric way. In such cases, ML methods will arguably provide a better specification of the propensity score

and outcome models than more traditional methods.

One such example is the doubly robust estimator (Bang and Robins 2005) that leverages ML methods for predicting both the propensity score and the potential outcome variables. The method then allows a doubly robust estimation of the potential outcome preserving the favorable statistical properties that permit rigorous causal inference. Another recent development in the estimation of treatment effects under the unconfoundedness assumption is the use of ML methods to directly make inference about the parameters using the double ML approach (Chernozhukov et al. 2018). This method involves using ML methods to residualize any potential impact that the covariates may have had on both the treatment and the outcome variables. The double ML framework can be combined with doubly robust estimators. It can also be readily extended to estimate heterogeneous treatment effects.

We can also use tree-based ML approaches such as causal forest (Wager and Athey 2018) for estimating both average and heterogeneous treatment effects in observational studies where the unconfoundedness assumption is satisfied. As discussed in Athey and Imbens (2016) and Wager and Athey (2018), the causal forest model can be particularly suitable for inferring treatment effects from rich observational data containing a large number of covariates. In contrast with conventional propensity score matching, causal forests utilize a flexible nonparametric data-driven approach to determine similarity across observations. Additionally, the estimation of traditional propensity score methods is often sensitive to the model specification (Fong, Hazlett, and Imai 2018), especially when the treatment variable is continuous. The causal forests are immune to such problems because the building of an honest tree (the building block of causal forests) does not rely on any particular functional form. Some recent examples in marketing of the use of causal forests for inferring causality from observational data are Guo, Sriram, and Manchanda (2021), Ellickson, Kar, and Reeder (2023), Pattabhiramaiah, Overby, and Xu (2022), and Zhang and Luo (2023).

For any of the quasi-experimental methods discussed above, researchers can also conduct sensitivity analyses, which involve assessing the extent of unobserved confounding necessary for nullifying the causal effect (Altonji, Elder, and Taber 2008; Imbens 2003; Rosenbaum and Rubin 1983). Liu, Kuramoto, and Stuart (2013) provide a nice introduction to sensitivity analysis. Recent work (e.g., Cinelli and Hazlett 2020; Oster 2019) has expanded on this idea to formally bound the strength of unobserved confounders by comparing them with observed covariates. Oster (2019) argues that the robustness of estimates to omitted variable bias can be examined by observing movements in (1) the coefficient of interest and (2) model $R^2$ from specifications that either include or exclude control variables in a regression. Masten and Poirier (2022) point out that unobserved confounders can either drive baseline estimates to zero or reverse their sign, with the latter actually being easier. They recommend several best practices for sensitivity assessment and even offer a companion Stata package to help researchers adopt these tools.

Last but not least, in addition to using sophisticated statistical/econometric/ML methods for mitigating such concerns, we can also consider using field or lab experiments to complement causal conclusions drawn from field data, especially for forming a deeper understanding of the underlying mechanisms (for some recent examples, see Nickerson et al. [2023] and Anderson et al. [2024]).

## Conclusion

The main purpose of this article is to offer some guidance to help marketing researchers

choose the most appropriate method for understanding causal relationships from quasi-experimental data. We begin with the basic DID method that is widely used in settings with treatment and control units and pre- and posttreatment periods. We thereafter discuss advances in using DID for contexts characterized by staggered treatment timing and heterogeneous treatment effects. We then explore flexible alternatives to DID, such as the synthetic control method, which is predicated on the researcher's access to a relatively large number of pre-treatment periods, and other alternative methods that do not require a large number of pretreatment periods. We cover how to estimate causal effects, conduct inference, and recommend best practices for these alternatives. Additionally, we review quasi-experimental methods with covariates, such as matching, and recent advances in causal ML methods for flexible matching. Given the rapid development of new methods, this article is not meant to be an exhaustive review of the literature on causal inference using observational data, but rather a useful starting point. Recent research has introduced novel ways for combining ML with instrumental variables (e.g., Hartford et al. 2017; Singh, Hosanagar, and Gandhi 2020) and incorporating natural language processing techniques within a causal framework (Feder et al. 2022) with the goal of improving causal inference. Thus, the researcher's methodological toolkit is ever expanding. We hope that this writeup helps guide researchers identify the right set of tools for answering causal research questions based on the data characteristics of their problem.

## AUTHORS

**Kathleen T. Li** is Assistant Professor of Marketing, University of Texas at Austin, USA

**Lan Luo** is Professor of Marketing, University of Southern California, USA.

**Adithya Pattabhiramaiah** is Associate Professor of Marketing, Georgia Institute of Technology, USA.

## CITATION

.Li, Kathleen T., Lan Luo, and Adithya Pattabhiramaiah (2024), "Causal Inference with Quasi-Experimental Data" IMPACT at JMR (November 13, 2024), https://www.ama.org/marketing-news/causal-inference-with-quasi-experimental-data/

## REFERENCES


Abadie, Alberto (2005), "Semiparametric Difference-in-Differences Estimators," Review of Economic Studies, 72 (1), 1–19.

Abadie, Alberto (2021), "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," Journal of Economic Literature, 59 (2), 391–425.

Abadie, Alberto, Alexus Diamond, and Jens Hainmueller (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, Alberto and Javier Gardeazabal (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country." American Economic Review, 93 (1), 113–32.

Abadie, Alberto and Guido W. Imbens (2006), "Large Sample Properties of Matching Estimators for Average Treatment Effects," Econometrica, 74 (1), 235–67.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2008), "Using Selection on Observed Variables to Assess Bias from Unobservables When Evaluating Swan-Ganz Catheterization," American Economic Review , 98 (2), 345–50.

Anderson, Eric, Chaoqun Chen, Ayelet Israeli, and Duncan Simester (2024), "Canary Categories," Journal of Marketing Research, 61 (5), 872–90.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager (2021), "Synthetic Difference-in-Differences," American Economic Review , 111 (12), 4088–4118.

Athey, Susan and Guido Imbens (2017), "The State of Applied Econometrics: Causality and Policy Evaluation," Journal of Economic Perspectives, 31 (2), 3–32.

Athey, Susan and Guido W. Imbens (2019), "Machine Learning Methods That Economists Should Know About," Annual Review of Economics, 11, 685–725.

Austin, Peter C. (2011), "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies," Multivariate Behavioral Research, 46 (3), 399–424.

Bai, Jushan and Serena Ng (2021), "Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data," Journal of the American Statistical Association, 116 (536), 1746–63.

Baker, Andrew C., David F. Larcker, and Charles C.Y. Wang (2022), "How Much Should We Trust Staggered Difference-In-Differences Estimates?" Journal of Financial Economics, 144 (2), 370–95.

Bang, Heejung and James M. Robins (2005), "Doubly Robust Estimation in Missing Data and Causal Inference Models," Biometrics, 61 (4), 962–73.

Bronnenberg, Bart J., Jean-Pierre Dubé, and Robert E. Sanders (2020), "Consumer Misinformation and the Brand Premium: a Private Label Blind Taste Test," Marketing Science, 39 (2), 382–406.

Callaway, Brantly and Pedro H.C. Sant'Anna (2021), "Difference-in-differences with Multiple Time Periods," Journal of Econometrics, 225 (2), 200–230.

Cao, Jingcun, Pradeep Chintagunta, and Shibo Li (2023). "From Free to Paid: Monetizing a Non-Advertising-Based App," Journal of Marketing Research, 60 (4), 707–27.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer (2019), "The Effect of Minimum Wages on Low-Wage Jobs," Quarterly Journal of Economics, 134 (3), 1405–54.

Chan, Marc K. and Simon C.M. Kwok (2016). "Policy Evaluation with Interactive Fixed Effects," working paper, University of Sydney.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018), "Double/Debiased Machine Learning for Treatment and Structural Parameters," Econometrics Journal , 21 (1), C1–C68.

Cinelli, Carlos and Chad Hazlett (2020), "Making Sense of Sensitivity: Extending Omitted Variable Bias,"

Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82 (1), 39–67.

Clarke, Damian, Daniel Pailanir, Susan Athey, and Guido Imbens (forthcoming), "On Synthetic Difference-in-Differences and Related Estimation Methods in Stata," Stata Journal.

Danaher, Peter J. and Michael S. Smith (2010). "Modeling Multivariate Distributions Using Copulas: Applications in Marketing," Marketing Science, 30 (1), 4–21.

Diamond, Alexis and Jasjeet Sekhon (2013). "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies," Review of Economics and Statistics, 95 (3), 932–45.

Eckles, Dean and Eytan Bakshy (2021), "Bias and High-Dimensional Adjustment in Observational Studies of Peer Effects," Journal of the American Statistical Association, 116 (534), 507–17.

Ellickson, Paul B., Wreetabrata Kar, and James C. Reeder III (2023), "Estimating Marketing Component Effects: Double Machine Learning from Targeted Digital Promotions," Marketing Science, 42 (4), 704–28.

Feder, Amir, Katherine A. Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, et al. (2022). "Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation and Beyond," Transactions of the Association for Computational Linguistics, 10, 1138–58.

Fong, Christian, Chad Hazlett, and Kosuke Imai (2018). "Covariate Balancing Propensity Score for a Continuous Treatment: Application to the Efficacy of Political Advertisements," Annals of Applied Statistics, 12 (1), 156–77.

Ghose, Anindya, Heeseung Andrew Lee, Kihwan Nam, and Wonseok Oh (2024), "The Effects of Pressure and Self-Assurance Nudges on Product Purchases and Returns in Online Retailing: Evidence from a Randomized Field Experiment," Journal of Marketing Research, 61 (3), 517–35.

Gobillon, Laurent and Thierry Magnac (2016), "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," Review of Economics and Statistics, 98 (3), 535–51.

Goodman-Bacon, Andrew (2021), "Difference-in-Differences with Variation in Treatment Timing," Journal of Econometrics, 225 (2), 254–77.

Gordon, Brett R., Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky (2019), "A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook," Marketing Science, 38 (2), 193–225.

Gormley, Todd A. and David A. Matsa (2011), "Growing Out of Trouble? Corporate Responses to Liability Risk," Review of Financial Studies, 24 (8), 2781–2821.

Guo, Tong, Srinivasaraghavan Sriram, and Puneet Manchanda (2020), "'Let the Sunshine In': The Impact of Industry Payment Disclosure on Physician Prescription Behavior," Marketing Science, 39 (3), 516–39.

Guo, Tong, Srinivasaraghavan Sriram, and Puneet Manchanda (2021), "The Effect of Information Disclosure on Industry Payments to Physicians," Journal of Marketing Research, 58 (1), 115–40.

Hartford, Jason, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy (2017), "Deep IV: A Flexible Approach for Counterfactual Prediction," in Proceedings of the 34th International Conference on Machine Learning, Vol. 70, D. Precup and Y.W. The, eds. PMLR, 1414–23.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," Review of Economic Studies, 64 (4), 605–54.

Hsiao, Cheng, H. Steve Ching, and Shui Ki Wan (2012), "A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China," Journal of Applied Econometrics, 27 (5), 705–40.

Imbens, Guido W. (2003), "Sensitivity to Exogeneity Assumptions in Program Evaluation," American Economic Review, 93 (2), 126–32.

Keller, Kristopher O., Jonne Y. Guyt, and Rajdeep Grewal (2024), "Soda Taxes and Marketing Conduct," Journal of Marketing Research, 61 (3), 393–410.

Kim, Sungjin, Clarence Lee, and Sachin Gupta (2020), "Bayesian Synthetic Control Methods," Journal of Marketing Research, 57 (5), 831–52.

Lambrecht, Anja, Catherine Tucker, and Xu Zhang (2024), "TV Advertising and Online Sales: A Case Study of Intertemporal Substitution Effects for an Online Travel Platform," Journal of Marketing Research, 61 (2), 248–70.

Li, Kathleen T. (2020), "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods," Journal of the American Statistical Association, 115 (532), 2068–83.

Li, Kathleen T. (2024), "Frontiers: A Simple Forward Difference-in-Differences Method," Marketing Science, 43 (2), 267–79.

Li, Kathleen T. and Garrett P. Sonnier (2023), "Statistical Inference for Factor Model Approach to Estimate Causal Effects in Quasi-Experimental Settings," Journal of Marketing Research, 60 (3), 449–72.

Li, Kathleen T. and Christophe Van den Bulte (2023), "Augmented Difference-in-Differences," Marketing Science, 42 (4), 746–67.

Li, Sheng, Nikos Vlassis, Jaya Kawale, and Yun Fu (2016), "Matching via Dimensionality Reduction for Estimation of Treatment Effects in Digital Marketing Campaigns," in Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI). IJCAI, 3768–74.

Liu, Weiwei, S. Janet Kuramoto, and Elizabeth A. Stuart (2013), "An Introduction to Sensitivity Analysis for Unobserved Confounding in Nonexperimental Prevention Research," Prevention Science, 14, 570–80.

Lovett, Mitchell J., Renana Peres, and Linli Xu (2019), "Can Your Advertising Really Buy Earned Impressions? The Effect of Brand Advertising on Word of Mouth," Quantitative Marketing and Economics, 17, 215–55.

Masten, Matthew A. and Alexandre Poirier (2022), "The Effect of Omitted Variables on the Sign of Regression Coefficients," arXiv, https://doi.org/10.48550/arXiv.2208.00552.

Nickerson, Dionne, Michael Lowe, Adithya Pattabhiramaiah, and Alina Sorescu (2023), "The Impact of Corporate Social Responsibility on Brand Sales: An Accountability Perspective," Journal of Marketing, 87 (1), 5–28.

Oster, Emily (2019), "Unobservable Selection and Coefficient Stability: Theory and Evidence," Journal of Business Economic Statistics, 37 (2), 187–204.

Pang, Xun, Licheng Liu, and Yiqing Xu (2022), "A Bayesian Alternative to Synthetic Control for Comparative Case Studies," Political Analysis, 30 (2), 269–88.

Park, Sungho and Sachin Gupta (2012), "Handling Endogenous Regressors by Joint Estimation Using Copulas," Marketing Science, 31 (4), 567–86.

Pattabhiramaiah, Adithya, S. Sriram, and Puneet Manchanda (2019), "Paywalls: Monetizing Online Content," Journal of Marketing, 83 (2), 19–36.

Pattabhiramaiah, Adithya., Eric Overby, and Lizhen Xu (2022), "Spillovers from Online Engagement: How a Newspaper Subscriber's Activation of Digital Paywall Access Affects Her Retention and Subscription Revenue," Management Science, 68 (5), 3528–48.

Petrin, Amil and Kenneth Train (2010), "A Control Function Approach to Endogeneity in Consumer Choice Models," Journal of Marketing Research, 47 (1), 3–13.

Ramachandra, Vikas (2018), "Deep Learning for Causal Inference," arXiv.org, https://doi.org/10.48550/arXiv.1803.00149.

Robins, James, Andrea Rotnitzky, and Lue Ping Zhao (1994), "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," Journal of the American Statistical Association, 89 (427), 846–66.

Rosenbaum, P.R. and D.B. Rubin (1983), "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome," Journal of the Royal Statistical Society: Series B (Methodological), 45 (2), 212–18.

Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe (2023), "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," Journal of Econometrics, 233 (2), 345–66.

Sant'Anna, Pedro H.C. and Jun Zhao (2020), "Doubly Robust Difference-in-Differences Estimators," Journal of Econometrics, 219 (1), 101–22.

Singh, Amandee, Kartik Hosanagar, and Amit Gandhi (2020), "Machine Learning Instrument Variables for Causal Inference," in Proceedings of the 21st ACM Conference on Economics and Computation (EC '20). ACM, 400–417

Sun, Liyang and Sarah Abraham (2021). "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," Journal of Econometrics, 225 (2), 175–99.

Wager, Stefan and Susan Athey (2018), "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests," Journal of the American Statistical Association, 113 (523), 1228–42.

Wooldridge, Jeffrey M. (2019). Introductory Econometrics: A Modern Approach, 7th ed. Cengage Learning.

Xu, Yiqing (2017), "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models," Political Analysis, 25 (1), 57–76.

Yao, Liuyi, Sheng Li, Yaliang Li, Mengdi Huai, Jing Gao, and Aidong Zhang (2018), "Representation Learning for Treatment effect estimation from Observational Data," in Advances in Neural Information Processing Systems. NeurIPS, 2633–43.

Zhang, Mengxia and Lan Luo (2023), "Can Consumer-Posted Photos Serve as a Leading Indicator of Restaurant Survival? Evidence from Yelp," Management Science, 69 (1), 25–50.

Zubizarreta, José R. (2015), "Stable Weights That Balance Covariates for Estimation with Incomplete Outcome Data," Journal of the American Statistical Association, 110 (511), 910–22.

## About JMR

The *Journal of Marketing Research* delves into the latest thinking in marketing research concepts, methods and applications from a broad range of scholars. It is included in both the *Financial Times* top 50 siness journals and the University of Texas Dallas research rankings journal list.

## About Impact at JMR

Impact at JMR highlights important and actionable content published in the *Journal of Marketing Research*. Executives can leverage this content in their organization. Professors can use it in their classrooms. Researchers can learn about distinct scholarly insights for their next big idea. We invite everyone to read, enjoy, learn, think and act.

# Unpacking the Instrumental Variables Approach

**Rajdeep Grewal and Yeşim Orhun | 8.23.2024**

Marketing scholars and practitioners frequently encounter causal questions related to strategic marketing decisions. Examples of such decisions include pricing, advertising, market entry, product development, brand positioning, contractual choices, and distribution decisions, to name a few. The factors and rules shaping strategic marketing decisions are often not fully observed by the researcher. When unobserved factors are also associated with the outcome of interest, this confounding relationship, known as the omitted variables or common causes issue, hinders the identification of the causal relationship of interest. For example, when estimating a causal impact of advertising spending on demand, researchers may find that both advertising spend and sales of a product are driven by the product's inherent market potential. Unobserved confounders create an identification challenges that empirical marketing scholars and practitioners often face when answering causal questions related to marketing decisions.

While randomized controlled experiments may be viable in some cases to address this issue, for many important strategic decisions, experimentation may not be feasible or ethical. In such cases, researchers use quasi-experimental approaches. Goldfarb, Tucker, and Wang (2022) provide an excellent overview and detailed guidance for a variety of quasi-experimental methods, including the instrumental variable (IV) approach. Papies, Ebbes, and Feit (2023) review the history of marketing literature on the use of various quasi-experimental methods and identify instrumental variables as the most common method.

Despite the broad use of the IV approach, the nature of confoundedness in a particular empirical context and how the proposed IV addresses it is, unfortunately, not always as clear as it could be. In this piece, we build on existing work to provide marketing scholars and practitioners with a resource that we hope will help them (1) identify confoundedness concerns in their empirical context, (2) transparently discuss the assumptions that need to hold for the IV approach to be valid in uncovering the causal effect, and (3) evaluate the plausibility of these assumptions.

We hope researchers use this piece not as a checklist but to critically assess whether the IV approach is appropriate given their research question and data. The IV approach has been used across markedly different empirical contexts and to answer a wide range of causal questions in the marketing literature. For example, recent articles using the IV approach examine the impact of review variance on demand (Lee, Bollinger, and Staelin 2023), the role of television advertising in satellite operators' commercial success (Yang, Lee, and Chintagunta 2021), the influence of pictures on the engagement with social media posts (Li and Xie 2020), and whether money-back guarantees can serve as a signal of quality (Yu, Ghosh, and Viswanathan 2022), to name a few. Each of these questions and empirical contexts presents a unique identification challenge, and the validity of the IV approach in each case has to be established based on a unique set of

facts and arguments. Papies, Ebbes, and Feit (2023, p. 281) note, "One of the main lessons from history is that there are no easy turn-key solutions to an endogeneity problem. Each of the methods used to address endogeneity in observational data relies on assumptions. It is critical that researchers using these methods carefully assess these assumptions in the context of their research question and data." We hope that this piece will aid readers in this regard.

## The Setup

Figure 1 presents a directed acyclic graph (DAG) demonstrating the omitted variable bias problem arising from unobserved confounders.[1] The causal impact we want to identify is the influence of treatment D on outcome Y (D → Y ). For example, we might be interested in the impact of price (D) on sales (Y ) for diet soda. The observed common causes W and the unobserved common causes U (unobservability to the researchers is represented by dashed lines) impact both D and Y . These sets of variables are referred to as confounders for the direct causal relationship of interest, because they lead to an association between D and Y even in the absence of a causal relationship. For example, we might imagine that a manager's sales expectations may be associated with both the pricing decision and the sales outcomes. Typically, expectations are unknown (an example of U). The price also responds to input costs, which may also be associated with demand (an example of W if observed). For example, in the case of diet soda, aspartame prices may go up when the demand for the diet soda category increases. It is relatively straightforward to deal with confounding due to W by conditioning on W, since these variables are observable.[2] The main identification challenge to uncover the causal impact D → Y arises from the existence of unobserved confounders U.

**Figure 1: DAG Representation of the Omitted Variable Bias and IV Approach**

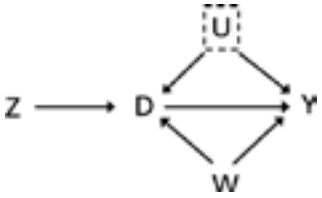



Figure 1 also includes a variable Z, which has a mediated pathway from Z to Y. In the case of a randomized controlled trial, Z indicates the random experimental assignment to treatment and control arms. Alternatively, Z can refer to an IV, which is not experimentally randomized, but is "as good as random" for the purposes of identifying D → Y. Either way, if five assumptions, which we detail below, are satisfied on the causal chain from Z to Y, Z can be used to identify the D → Y relationship.

## Assumptions

For Z to be a valid IV for identifying the causal effect of D → Y, five assumptions need to be satisfied: (1) the independence assumption, (2) the stable unit treatment value assumption (SUTVA), (3) the inclusion restriction, (4) the exclusion restriction, and (5) the monotonicity assumption. The plausibility of these assumptions determines whether the IV approach is suitable for a specific research question within a given institutional context. Next, we explain these assumptions and detail their implications in the context of an example to build intuition. For more technical coverage, we recommend readers consult econometrics textbooks and articles, such as Angrist, Imbens,

1 In a directed graph, each node is a random variable, and the edges are directed, indicating causal associations in the direction of the arrow. In acyclic graphs, causality runs in one direction. To learn more about DAGs and their usefulness, see Pearl (2009) and Imbens (2020).

2 Care must be taken when controlling for the W → Y relationship. A model with strong functional form assumptions, such as a linear regression model, may not be appropriate. Approaches such as matching, nonparametric methods, and machine learning methods can be used instead.

and Rubin (1996), Angrist, and Krueger (2001), Angrist and Pischke (2009), Cunningham (2021), and Murray (2006), among others.

### *Five Assumptions*

The first assumption, called the independence assumption, also known as the ignorability or exchangeability assumption, stipulates that the assignment of different levels of Z to different units (e.g., firms, individuals) is as good as random. In our DAG, this means that Z is not associated with U or W (indicated by the fact that no edge is drawn between Z and these variables in Figure 1). We can weaken this assumption to conditional independence. For example, if there were an association between Z and W, conditional independence could be achieved by controlling for W.[3] One can alternatively think of this assumption as stipulating that either there is no selection into Z, and if there is selection, it is only on observables. The researcher needs to defend the plausibility of this assumption, which is most easily done when the instrument Z is indeed randomized. Otherwise, why believe the conditional independence of Z, but not D? The answer will depend on the institutional context, as we discuss below.

The second assumption, the stable unit treatment value assumption (SUTVA), states that the value of unit i's instrument or treatment does not affect other units' potential outcomes (i.e., there are no unmodeled spillovers or interference). This assumption permits us to write Di(Z) = Di(Zi) and Yi(Z, D(Z)) = Yi(Zi, Di(Zi)). This is not a trivial assumption, and researchers should contemplate its plausibility carefully. In practice, SUTVA violations can occur due to many reasons, including general equilibrium effects, anticipation, contagion, information spillovers, social comparisons, externalities, and network effects, among others. Even in the context of a randomized experiment, SUTVA may be violated. For example, if an instrument (or experimental manipulation) varies the prices of a subset of firms, it can also have an impact on the prices of untreated firms through competitive pricing (see, e.g., Holtz et al. 2024).

When the SUTVA and independence assumptions are satisfied, the IV is referred to as being exogenous. Many marketing papers focus on the independence assumption and fail to consider SUTVA in discussing the validity of their IV. Moreover, it is common for marketing papers to refer to the independence assumption as the exclusion restriction—a confusion that may be driven by the necessity of the independence assumption in establishing exogeneity. We discuss the exclusion restriction below and highlight the need to discuss the plausibility of all five assumptions if the objective is to identify the causal effect D → Y.

It is important to note that the independence and SUTVA assumptions are sufficient for Z → Y to be interpreted causally. In the IV literature, this causal effect is called the reduced form effect. Sometimes, the reduced form effect suffices when the research question can be satisfactorily answered by identifying the causal effect of Z on Y.[4] However, if the causal effect of interest is D → Y, three additional assumptions we discuss below are necessary for the identification of the local average treatment effect (LATE), which is defined as the causal effect D → Y among units for which Z had an effect on D. The IV approach can only identify the "local" effect of D on Y for units whose treatment status can be

3 With all other assumptions satisfied, the two-stage least square (2SLS) with covariates estimates a weighted average of the covariate-specific local average treatment effects (LATEs). Abadie (2003) shows how to estimate the overall LATE using a weighting approach based on a "propensity score" for the instrument.

4 With all other assumptions satisfied, the two-stage least square (2SLS) with covariates estimates a weighted average of the covariate-specific local average treatment effects (LATEs). Abadie (2003) shows how to estimate the overall LATE using a weighting approach based on a "propensity score" for the instrument.

manipulated by the instrument, and not the average treatment effect (ATE) among all units. This has implications for generalizability of the identified causal effect, which we return to in our concluding remarks.

The third assumption, called the inclusion restriction, also referred to as the first-stage assumption or relevance assumption, is that the instrument Z must influence D. This assumption is the only one among the five assumptions that can be empirically verified. It is expected that papers provide empirical evidence for the association between Z and D and assess the strength of that relationship by testing the coefficient on the instrument in the so-called first-stage regression of the treatment on the instrument. We return to the issue of power of the first stage and inference with weak instruments in our concluding remarks.

The fourth assumption, the exclusion restriction assumption, indicates that any effect of the instrument Z on outcome Y is exclusively through its effect on D. This assumption is depicted by Figure 1 indicating that there is no other impact of Z on Y once we condition on the value of D. If the model is overidentified (i.e., there are more instruments than endogenous variables), then an overidentification test (e.g., Sargan–Hansen test) can be performed to test whether all instruments are uncorrelated with the 2SLS residuals. However, these tests require a constant effects assumption that is often difficult to defend, and therefore overidentification tests are not commonly used (see, e.g., Kiviet and Kripfganz 2021). Instead, researchers establish logical support for this assumption using institutional details and falsification tests, as we discuss in the next section.

As we noted previously, marketing studies sometimes confuse the exclusion restriction with the independence assumption. Thus, it is important to highlight the differences. While the independence assumption is about the instrument (Z) not being correlated with unobserved confounders (U ), the exclusion restriction states that the instrument does not impact the dependent variable (Y ) except through its effect on the endogenous independent variable (D). For example, in using a cost shifter as an IV for price in estimating the impact of prices on demand, the cost shifter satisfies the exclusion restriction if the only effect that the cost shifter has on demand operates through its effect on price. Thus, not all cost shifters will satisfy the exclusion restriction, a point we return to when discussing the role of the institutional context in assessing the plausibility of the assumptions.

The fifth assumption, called the monotonicity assumption, is the final assumption necessary in the identification of the LATE when D → Y is heterogeneous across units (Heckman, Urzua, and Vytlacil 2006; Heckman and Vytlacil 2005, 2000; Imbens and Angrist 1994).[5] The monotonicity assumption requires that a hypothetical change in Z either has no impact on a unit's treatment status D or changes it in the same direction as it does for all other units on which it has an impact. Let's consider a binary Z and a binary D for illustrative purposes. In the language of the potential outcomes model (Rubin 2005), compliers are units whose behavior is impacted by the instrument. For compliers, assume D = 1 when Z = 1 and D = 0 when Z = 0. Defiers are those for whom the instrument has the opposite effect as compliers; for them, then, D = 1 when Z = 0 and D = 0 when Z = 1. Always- takers and never-takers are not impacted by the instrument; for always-takers, D = 1, and for never takers, D = 0, regardless of Z. Thus, always-takers and never-takers do not inform the IV estimate.

5 If the homogeneity of D →Y assumption can be upheld, then (1) we do not need the monotonicity assumption, and (2) the causal effect can be generalized to the population at large, giving us the ATE instead of the LATE. However, the homogeneity assumption is unlikely to hold in many marketing applications; therefore, we discuss the monotonicity assumption in the main text.

The monotonicity assumption indicates that we can have either compliers or defiers, but not both.[6] The monotonicity assumption would be violated, for example, when a nudge (e.g., antismoking ads) works in the expected direction for some but causes a backlash reaction for others. Without making further assumptions, it is not possible to empirically verify the monotonicity assumption.[7]

### *A Demonstration*

Authors need to translate the implications of these five assumptions into their empirical context and convince the reader of their plausibility. To demonstrate, we discuss the example offered by Calder-Wang and Gompers (2021), who study the impact of employee gender diversity on venture capital firm performance. To identify this causal effect, the authors use the sex of venture capital partners' children as an instrument for the decision to hire a woman. The inclusion restriction conjecture is that partners who have more daughters are more likely to hire women. The authors discuss conceptual reasons for this relationship and also present empirical evidence of the first stage, which we recommend all papers using the IV approach do.

In this context, SUTVA requires that the hiring decisions and financial performance of a partner are impacted only by the number of daughters the partner has, and not by the gender of the children of other partners. This assumption can be violated in several ways. For example, if the supply of qualified female employees is extremely tight, then the increased interest in hiring female employees due to the gender composition of one firm's partners could impact the hiring and/or financial performance of a competing firm.

The independence assumption is equivalent to assuming that whether partners have sons or daughters (conditional on having children) is as good as random. If certain parents (e.g., those who hold different gender views) employed a gender-based stopping rule (e.g., keep having children until they have at least one son), the independence assumption would be violated. To defend against this particular concern, the authors provide evidence that a first-born daughter does not predict the total number of children, which constitutes an example of a falsification test—a concept we discuss below.

The exclusion restriction in this example necessitates that the genders of partners' children do not have an impact on the venture capital firm's performance except through the impact on gender diversity in the firm. The authors recognize that if having more daughters directly improves a partner's skills in a way that increases their ability to source or close deals, this assumption would be violated. They provide evidence that venture capital partners with more daughters do not have more successful deals, which is another example of a falsification test.

Finally, the monotonicity assumption necessitates that we can only have partners for whom having more daughters would either increase their likelihood of hiring women or not impact it, but we cannot have partners for whom having more daughters would decrease the likelihood of hiring women. This assumption would be violated, for example, if, for a minority of partners, having daughters reinforces sexist views of the workplace.

We summarize this discussion in Table 1. The table also includes Sinkinson and Starc (2019) as another working example

6 This nomenclature is the reason why the LATE is also referred to as the complier average causal effect (CACE). See De Chaisemartin (2017) for inference in the IV approach without the monotonicity assumption.

7 Frandsen, Lefgren, and Leslie (2023) offer a test of the monotonicity assumption in a particular IV design by making additional assumptions on the ATE among those who violate the monotonicity assumption.

to demonstrate the assumptions required in using political advertising cycles as an instrument for advertising spend. For a more detailed discussion of potential violations of the exclusion restriction and monotonicity assumption in using political cycles as instruments, and the important role time and market fixed effects play, see Moshary, Shapiro, and Song (2021). As these examples make clear, the plausibility of the identifying assumptions needs to be defended with institutional details and supporting empirical patterns. If the assumptions cannot be credibly defended, researchers should not use the IV approach.

## Evaluating and Defending the Plausibility of Assumptions

None of the identifying assumptions we discussed above, except for the inclusion restriction, can be empirically validated. Therefore, they must be logically established and defended based on common sense, subject matter arguments, and institutional details. Goldfarb, Tucker, and Wang (2022, p. 5) suggest that "the objective for the authors is to pursue projects only when they can convince themselves (and their readers) that the causal interpretation is more plausible than other possible explanations. It is impossible to prove the validity of a quasi-experiment. … The credibility of any quasi-experimental work therefore relies on the plausibility of the argument for causality rather than on any formal statistical test." In our assessment, many of the published papers in marketing using the IV approach do not offer a detailed-enough discussion of the implications and plausibility of the required assumptions in the empirical context they study. When they do, the discussion tends to focus mainly on the relevance and independence assumptions. In addition to providing a discussion of other assumptions, we suggest that researchers treat identification as a central part of the manuscript's narrative, using institutional details and theory to tie together elements that make the research question important and the identification valid. Thus, in this section, we discuss approaches researchers can take to evaluate the plausibility of all the assumptions in the IV framework to develop a cohesive and convincing story.

### *Institutional Details*

Subject matter arguments based on institutional knowledge are paramount to judging whether the required assumptions are plausible and, consequently, whether causal inference can be achieved. In many quasi-experimental papers, identifying assumptions are justified solely by subject-matter arguments that use institutional details. The clarity with which the authors translate the identifying assumptions to their context and provide detailed discussion of institutional details that help the reader evaluate these assumptions go a long way in convincing the reader of a causal relationship.

Institutional context makes or breaks an instrument. An instrument that satisfies an assumption naturally in some contexts may blatantly violate it in others. To illustrate, let's consider cost-based instruments, which are commonly used in marketing and industrial organization to obtain the causal impact of prices on demand. Do input costs as an instrument satisfy the exclusion restriction? The answer depends on the institutional details. For example, consider using orange wholesale prices as an instrument for orange juice prices when estimating demand for orange juice in Michigan. Imagine that a drought in Florida pushed up orange wholesale prices across the nation. It might be relatively straightforward to defend that this input cost variation is plausibly exogenous to demand conditions for orange juice in Michigan. Now, instead, consider the cost of steel as an input for automobile manufacturing. The automotive industry accounts for 10%–15% of global steel use, and automobile production levels are known to impact steel prices. Therefore, the

**Table 1:** Summary of Assumptions in the IV Approach

| Assumption | Critical Question | Calder-Wang and Gompers (2021) | Sinkinson and Starc (2019) |
|---|---|---|---|
| **Independence assumption:** The instrument Z does not share common causes with the outcome Y | Are there any omitted variables that determine Z and Y? What allows us to claim that Z is as good as random (when X is not)? | Gender of partner's children is determined by nature, and thus independent of firm performance. | The variation in political ad spending is independent of the variation in statin demand conditions. |
| **SUTVA:** A unit's response to its own value of the instrument $Z_i$ does not depend on the value of the instrument for other units $Z_{-i}$. | Is it possible that there are spillovers or interference among different units? | A partner's hiring decisions and financial performance is not impacted by the gender of other partners' children. | Advertising and sales of a pharmaceutical company in a market-month does not respond political ad spending in other markets and months. |
| **Inclusion restriction:** The instrument Z must influence the treatment D, either in a positive or negative manner. | Why do we expect D to respond to Z? (Note: this is the only assumption for which we can provide empirical evidence) | Partners who have more daughters are more likely to hire women. | Increases in political advertising spending displace other types of advertising. |
| **Exclusion restriction:** The effect of instrument Z on outcome Y operates only through the effect of Z on D. | Can Z influence Y through other channels (direct or indirect) that are not through D? | Gender of partners' children do not have an impact on firm performance except through its impact on hiring. | Political advertising cycle has no other effect on pharmaceutical demand except through its impact on advertising decisions. |
| **Monotonicity:** The impact of the instrument Z on D across units of analysis is (weakly) in the same direction. | Do compliers and defiers coexist? | Having more daughters would (weakly) increase the likelihood of hiring women for all partners. | All statin manufacturers (weakly) decrease advertising when political ad spending increases. |

**Disclaimer/Warning:** The goal of this table is to summarize the discussion in the "Assumptions" section. It should not be used as a template or a checklist. It is not intended to support, not replace, critical engagement with the necessary identifying assumptions in a given empirical context.

demand for automobiles may have an impact on steel prices. Alternatively, both steel prices and demand for automobiles may be impacted by the strength of the economy. In this context, it may be hard to refute that steel prices have no association with consumer demand for automobiles except through their impact on car prices.

Another illustration of the institutional context mattering for the validity of an instrument comes from examiner designs (also called judge fixed-effect design, or leniency design). In these designs, there is an examiner who has discretion in determining the outcomes (e.g., a judge in a hearing, a grader in a class, or a consumer in responding to a satisfaction survey), and there is systematic heterogeneity in their judgments (e.g., some judges being systematically more lenient than others, some consumers being generally more grumpy). Similar designs have been adapted to study marketing questions (e.g., Lee, Bollinger, and Staelin 2023; Li and Xie 2020). In cases where the assignment of the examiner is as good as random, we can consider the identity of the examiner as an instrument for the treatment whose effect we are trying to examine. For example, consider being interested in the impact of review valence on product sales, and assume that certain consumers are systematically more negative in their review behavior and that the arrival of consumer types (in terms of their overall negativity) is random. We could imagine using the identity of the consumer as an instrument for review valence. In this context, is the monotonicity assumption satisfied? It depends. Monotonicity holds whenever any product that would have received a 4-star rating from a generally negative consumer receives a 4-star or 5-star rating from a happy-go-lucky consumer. It is violated if the happy-go-lucky consumer sometimes rates products worse than the generally negative consumer. In this context, the researcher may argue that the assumption is more likely to hold within a product category, rather than across categories. However, even within a product category, consumers may vary in what makes them unhappy. For example, the happy-go-lucky consumer might be generally positive, except in cases where the product arrives damaged, in which case they switch and become even more negative than the generally negative consumer. The institutional context matters greatly in the researcher's ability to make a case for (or refute) the likelihood of this scenario.[8]

There are many papers that bring institutional details expertly into the narrative and use them to lay out the rationale for the plausibility of the identifying assumptions. For example, consider Bruhn et al. (2022), who examine whether veterans who have experienced more combat exposure are more likely to have negative life outcomes postdeployment (education, financial heath, suicide, incarceration, etc.). They clearly explain the institutional process of brigade assignments in the U.S. Army to support the relevance of their instrument and defend the (conditional) independence assumption. In making a case for the plausibility of the exclusion restriction against one potential criticism, Sinkinson and Starc (2019) point out that detailing spending levels are set at the annual level and therefore cannot be quickly adapted at the market level in response to TV ad spending declining. Similarly, in examining peer effects on salesperson quitting behavior using the IV approach, Sunder et al. (2017) argue that the management's evaluation of the salesperson in the first month of them joining the firm satisfies the independence assumption due to its private nature. We recommend that all authors think through the institutional details when specifying a causal model and picking an instrument, and communicate these

8 For more details on inference in examiner designs, we refer the interested reader to Chyn, Frandsen, and Leslie (2024). Chapter 7.8.2 of Cunningham (2021) also provides a useful discussion of the plausibility of the independence, exclusion restriction, and monotonicity assumptions in examiner designs.

details to their readership in the context of the five identifying assumptions we have discussed.

### *Falsification Tests*

An important benefit of specifying the identifying assumptions of a causal research designs is that these assumptions often have falsifiable implications. Although identifying assumptions cannot be verified empirically, researchers can conduct tests for these implications to check whether they can be empirically refuted (Angrist and Krueger 1999). These tests are called falsification tests. Failed falsification attempts do not prove the assumption they are designed to falsify, but can help build the case for the plausibility of the identifying assumptions. In this section, we aim to build intuition for coming up with falsification tests by sharing examples of tests used by researchers across a variety of areas.

In devising a falsification test for the independence assumption, researchers often check for balance in observables across levels of Z.[9] The idea here is that if Z is as good as random, then we would not expect any systematic differences in the means or distributions of pretreatment covariates across different levels of Z. At other times, the researchers evaluate a plausible threat to the independence assumption. We discussed one such example in the case of Calder-Wang and Gompers (2021). Another example refuting a threat to the independence assumption comes from Gong et al. (2017). The authors conduct a field experiment to show that tweets by a media company on Sina Weibo about its TV shows increase the viewership of these shows. The experiment assigns a subset of shows to receive company tweets. The authors want to refute the possibility that the assignment of shows was somehow correlated with unobservables that drive show viewership (e.g., show popularity). To do so, they exploit the fact that tweeting happens in a single point in time, but the TV show airs at different times across geographic areas. Under the independence assumption, we would not expect any impact of treatment when tweeting happened after the show airs in a geography. If the assignment correlated with unobservables, however, we would expect a higher viewership of shows even if tweeting happened after the show airs. This constitutes a clever falsification test of the independence assumption in the paper's institutional context.

To create a falsification test for the exclusion restriction, researchers ask what empirical pattern might suggest that Z has an impact on Y that did not operate through D alone. The falsification test Calder-Wang and Gompers (2021) offer for the exclusion restriction is an example of evaluating a particular channel by which Z may have a direct effect on Y. In most cases, falsifying the exclusion restriction involves testing the reduced form effect of Z on Y in situations where it is impossible (or, extremely unlikely) for Z to influence D. If Z has an impact on Y in these situations, it would show that Z's impact on Y does not always operate through D. Sometimes, it may be possible to evaluate the reduced form effect among never-takers as a falsification test. For example, Erikson and Stoker (2011) use the Vietnam draft lottery based on birth dates as an instrument for vulnerability to military service to study whether this vulnerability impacts political attitudes. Under the exclusion restriction assumption, birth dates should not affect (1) men's attitudes who are exempt from the draft due to college deferrals, or (2) women's attitudes. Thus, the reduced form effect is expected to be zero for these two groups of never-takers, unless birth dates have a direct effect on political attitudes that does not only operate through vulnerability to military service.

9 Of course, if the assumption is conditional independence, the balance assessment is also conditional. When assessing balance across a number of variables, instead of running numerous independent balance tests, it is better to employ one omnibus balance test (e.g., Hansen and Bowers 2008).

In other cases, it might be possible to devise falsification tests based on the fact that the exclusion restriction would predict a null effect of the instrument among the always-takers. For example, in the context of medical research, a physician's general tendency to operate is used as an instrument for whether patients received surgery, in order to examine the impact of having surgery versus not having it on mortality. Yang et al. (2014) provide a falsification test that exploits the fact that certain subgroups of patients in the data are always operated on, and therefore their in-hospital mortality should not be affected by their physician's general tendency to operate (i.e., the reduced form effect among always-takers should be zero) if the exclusion restriction holds (see also Keele et al. 2019).

Falsification tests for other assumptions can also be devised based on the particulars of the empirical context (see, e.g., Burke, Bergquist, and Miguel 2019; Danieli et al. 2024). Overall, falsification tests can be a useful tool in making the case for the plausibility of the identifying assumptions, but as we cautioned above, should not be interpreted as providing proof for them.

## Concluding Remarks

In this piece, we provided a brief discussion of the identifying assumptions in the IV approach, focusing on the importance of assessing their plausibility in a given empirical context and possessing sufficient institutional knowledge to do so. While we focused on one approach, we hope that many of the insights are useful for researchers in thinking through other quasi-experimental approaches. As we conclude, we would like to draw the reader's attention to a few additional items.

First, we want to emphasize the importance of being clear on the nature of confounding relationships. This clarity helps determine whether the IV approach is necessary and, if necessary, helps identify a valid IV. More generally, clearly specifying the sources of confoundedness is a useful first step in figuring out what methods are useful to deal with the identification challenges they present. Simpler methods, with fewer assumptions, might be sufficient and preferred. For instance, if the confounders vary at a level higher than the variation in D or Y , the researcher may be able to control for U with fixed effects.

Second, we would like to highlight the issue of weak instruments. Although the inclusion restriction only requires that Z is associated with D, a large body of literature, which we do not cover here, shows that weak associations can mean that the 2SLS estimate is vulnerable to bias (e.g., Andrews, Stock, and Sun 2019; Rossi 2014). The weak instrument bias is often exacerbated by a large number of instruments. Therefore, one strong instrument is generally preferred over using many instruments, some of which are weak (Angrist and Kolesár 2024). This should also caution readers about including a large number of fixed effects as IVs in the model, especially if the nature of the confounder does not require them. To assess the strength of the instrument, researchers often use the rule of thumb that the F-statistic (on the null hypothesis that coefficients in the first stage are zero) should be 10 or larger, even though the original research this rule of thumb is predicated on offers more nuanced critical values (e.g., Staiger and Stock 1997; Stock, Wright, and Yogon 2002; Stock and Yogo 2002). A well-known issue with using this rule of thumb is that homoskedasticity was a key assumption in the literature that produced it. In the case of one endogenous regressor and linear models, Olea and Pflueger (2013) propose an effective first-stage F-statistic that corrects for nonhomoskedastic errors (e.g., clustering, autocorrelation). Lee et al. (2022) offer an inference approach for the single-instrument case that is robust to heteroskedasticity and clustering, which applies an adjustment factor to the 2SLS standard errors based on the first stage. This work shows that once violations of homoskedasticity

are considered, the necessary critical values are larger by an order of magnitude compared with the common rule of thumb. We should also highlight that in the case of a weak first stage, not all hope is lost: researchers can use weak instrument robust inference. We refer the interested reader to the econometrics literature on weak instruments (e.g., Angrist et al. 1999; Andrews and Stock 2005; Andrews, Stock, and Sun 2019; Chernozhukov and Hansen 2008) for further details.

Third, it is important for authors using the IV method to think carefully about the external validity of the results they obtain. As we have discussed, the IV approach can only identify the treatment effect among compliers. This is an unknown subset of the data, as treated units are a mix of always-takers and compliers. Furthermore, the complier group depends on the instrument used. Different instruments will lead to different estimands. So, given the instrument(s) the researcher is using, it is important to consider the specificity of the source of variation in D that is generated by Z and discuss how generalizable the results may be to other groups, situations, or times. Sometimes, the instrument only impacts the behavior of a narrow group of people who are likely to have different D → Y than the population of interest. At other times, we may not expect meaningful differences in the causal D → Y relationship between compliers and the general population. Research benefits from transparency. We recommend authors to openly discuss external validity issues and use institutional details to support any arguments of generalizability.

To conclude, strategic and nonrandom decisions by consumers, managers, firms, regulators, and other institutional actors permeate marketing problems and issues. For many such situations, the IV approach may be the right quasi-experimental approach to study research questions of interest. We hope this piece encourages the appropriate use of IVs as tools to provide valid theoretical and managerial insights.

## AUTHORS

**Rajdeep Grewal** is Townsend Family Distinguished Professor and Area Chair of Marketing at the Kenan-Flagler Business School, University of North Carolina, Chapel Hill, NC (Email: grewalr@unc.edu).

**Yesim Orhun** is Associate Professor of Marketing and Michael R. and Mary Kay Hallman Fellow at the Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI (Email: aorhun@umich.edu).

## CITATION

Grewal, Rajdeep and Yesim Orhun (2024), "Unpacking the Instrumental Variables Approach," *Impact at JMR* (August 23), https://www.ama.org/marketing-news/unpacking-the-

## REFERENCES


Abadie, Alberto (2003), "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113 (2), 231–63.

Andrews, Donald W.K. and James H. Stock (2005), "Inference with Weak Instruments," NBER Technical Working Paper 0313, https://www.nber.org/papers/t0313.

Andrews, Isaiah, James H. Stock, and Liyang Sun (2019), "Weak Instruments in Instrumental Variables Regression: Theory and Practice," *Annual Review of Economics*, 11, 727–53.

Angrist, Joshua and Michal Kolesár (2024), "One Instrument to Rule Them All: The Bias and Coverage of Just-ID IV," *Journal of Econometrics*, 240 (2), 105398.

Angrist, Joshua D., Guido W. Imbens, and Alan B. Krueger (1999), "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14 (1), 57–67.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91 (434), 444–55.

Angrist, Joshua D. and Alan B. Krueger (1999), "Empirical Strategies in Labor Economics," in Handbook of Labor Economics, O. Ashenfelter and D. Card, eds. Elsevier, 1277–1366.

Angrist, Joshua D. and Alan B. Krueger (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 15 (4), 69–85.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Bruhn, Jesse, Kyle Greenberg, Matthew Gudgeon, Evan K. Rose, and Yotam Shem-Tov (2022), "The Effects of Combat Deployments on Veterans' Outcomes," NBER Working Paper 30622, https://www.nber.org/papers/w30622.

Burke, Marshall, Lauren F. Bergquist, and Edward Miguel (2019), "Sell Low and Buy High: Arbitrage and Local Price Effects in Kenyan Markets," *Quarterly Journal of Economics*, 134 (2), 785–842.

Calder-Wang, Sophie and Paul A. Gompers (2021), "And The Children Shall Lead: Gender Diversity and Performance in Venture Capital," *Journal of Financial Economics*, 142 (1), 1–22.

Chernozhukov, Victor and Christian Hansen (2008), "The Reduced form: A Simple Approach to Inference with Weak Instruments," *Economics Letters*, 100 (1), 68–71.

Chyn, E., B. Frandsen, and E.C. Leslie (2024), "Examiner and Judge Designs in Economics: A Practitioner's Guide," NBER Working Paper 32348, https://www.nber.org/papers/w32348 .

Cunningham, Scott (2021), *Causal Inference: The Mixtape*. Yale University Press.

Danieli, Oren, Daniel Nevo, Itai Walk, Bar Weinstein, and Dan Zeltzer (2024), "Negative Control Falsification Tests for Instrumental Variable Designs," arXiv preprint, https://doi.org/10.48550/arXiv.2312.15624.

De Chaisemartin, Clément (2017), "Tolerating Defiance? Local Average Treatment Effects Without Monotonicity," *Quantitative Economics*, 8 (2), 367–96.

Erikson, Robert S. and Laura Stoker (2011), "Caught in the Draft: The Effects of Vietnam Draft Lottery Status on Political Attitudes," *American Political Science Review*, 105 (2), 221–37.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie (2023), "Judging Judge Fixed Effects," *American Economic Review*, 113 (1), 253–77.

Gerber, Alan S. and Donald P. Green (2012), *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton.

Goldfarb, Avi, Catherine Tucker, and Yanwen Wang (2022), "Conducting Research in Marketing with Quasi-Experiments," *Journal of Marketing*, 86 (3), 1–20.

Gong, Shiyang, Juanjuan Zhang, Ping Zhao, and Xuping Jiang (2017), "Tweeting as a Marketing Tool: A Field Experiment in the TV Industry," *Journal of Marketing Research*, 54 (6), 833–50.

Hansen, Ben B. and Jake Bowers (2008), "Covariate Balance in Simple, Stratified and Clustered Comparative Studies," *Statistical Science*, 23 (2), 219–36.

Heckman, James J., Sergio Urzua, and Edward Vytlacil (2006), "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics & Statistics*, 88 (3), 389–432.

Heckman, James J. and Edward Vytlacil (2005), "Structual Equations, Treatment Effects, and Econometirc Policy Evaluation," *Econometrica*, 73 (3), 669–738.

Heckman, James J. and Edward J. Vytlacil (2000), "The Relationship Between Treatment Parameters within a Latent Variable Framework," *Economics Letters*, 66 (1), 33–39.

Holtz, David, Felipe Lobel, Ruben Lobel, Inessa Liskovich, and Sinan Aral (2024), "Reducing Interference Bias in Online Marketplace Pricing Experiments," *Management Science* (published online April 5), https://doi.org/10.1287/mnsc.2020.01157.

Imbens, Guido W. (2020), "Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics," *Journal of Economic Literature*, 58 (4), 1129–79.

Keele, Luke, Qingyuan Zhao, Rachel R. Kelz, and Dylan Small (2019), "Falsification Tests for Instrumental Variable Designs with an Application to Tendency to Operate," *Medical Care*, 57 (2), 167–71.

Kiviet, Jan F. and Sebastian Kripfganz (2021), "Instrument Approval by the Sargan Test and Its Consequences for Coefficient Estimation," *Economics Letters*, 205, 109935.

Lee, David S., Justin McCrary, Marcelo J. Moreira, and Jack Porter (2022): "Valid t-Ratio Inference for IV," *American Economic Review*, 112 (10), 3260–90.

Lee, Nah, Bryan Bollinger, and Richard Staelin (2023), "Vertical Versus Horizontal Variance in Online Reviews and Their Impact on Demand," *Journal of Marketing Research*, 60 (1), 130–54.

Li, Yiyi and Ying Xie (2020), "Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement," *Journal of Marketing Research*, 57 (1), 1–19.

Moshary, Sarah, Bradley T. Shapiro, and Jihong Song (2021), "How and When to Use the Political Cycle to Identify Advertising Effects," *Marketing Science*, 40 (2), 283–304.

Murray, Michael P. (2006), "Avoiding Invalid Instruments and Coping with Weak Instruments," *Journal of Economic Perspectives*, 20 (4), 111–32.

Olea, José L.M. and Carolin Pflueger (2013), "A Robust Test for Weak Instruments," *Journal of Business & Economic Statistics*, 31 (3), 358–69.

Papies, Dominik, Peter Ebbes, and Elea M. Feit (2023), "Endogeneity and Causal Inference in Marketing," in *The History of Marketing Science*, 2nd ed., Russell S. Winer and Scott A. Neslin, eds. World Scientific, 253–300.

Pearl, Judea (2009), "Causal Inference in Statistics: An Overview," *Statistics Surveys*, 3, 96–146.

Rossi, Peter E. (2014), "Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications," *Marketing Science*, 33 (5), 655–72.

Rubin, Donald B. (2005), "Causal Inference using Potential Outcomes: Design, Modeling, Decisions," *Journal of the American Statistical Association*, 100 (469), 322–31.

Sinkinson, Michael and Amanda Starc (2019), "Ask Your Doctor? Direct-to-Consumer Advertising of Pharmaceuticals," *Review of Economic Studies*, 86 (2), 836–81.

Staiger, Douglas and James Stock (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65 (3), 557–86.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo (2002), "A Survey of Weak Instruments and

Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20 (4), 518–29.

Stock, James H. and Motohiro Yogo (2002), "Testing for Weak Instruments in Linear IV Regression," Technical Working Paper 0284, https://www.nber.org/papers/t0284.

Sunder, Sarang, V. Kumar, Ashley Goreczny, and Todd Maurer (2017), "Why Do Salespeople Quit? An Empirical Examination of Own and Peer Effects on Salesperson Turnover Behavior," *Journal of Marketing Research*, 54 (3), 381–97.

Yang, Fan, José R. Zubizarreta, Dylan S. Small, Scott Lorch, and Paul R. Rosenbaum (2014), "Dissonant Conclusions When Testing the Validity of an Instrumental Variable," *American Statistician*, 68 (4), 253–63.

Yang, Joonhyuk, Jun Youn Lee, and Pradeep K. Chintagunta (2021), "Commercial Success Through Commercials? Advertising and Pay-TV Operators," *Journal of Marketing Research*, 58 (5), 925–47.

Yu, Shan, Mrinal Ghosh, and Madhu. Viswanathan (2022), "Money-Back Guarantees and Service Quality: The Marketing of In Vitro Fertilization Services," *Journal of Marketing Research*, 59 (3), 659–73.

## About JMR

The *Journal of Marketing Research* delves into the latest thinking in marketing research concepts, methods and applications from a broad range of scholars. It is included in both the *Financial Times* top 50 business journals and the University of Texas Dallas research rankings journal list.

## About Impact at JMR

Impact at JMR highlights important and actionable content published in the *Journal of Marketing Research*. Executives can leverage this content in their organization. Professors can use it in their classrooms. Researchers can learn about distinct scholarly insights for their next big idea. We invite everyone to read, enjoy, learn, think and act.

# A Review of Copula Correction Methods to Address Regressor–Error Correlation

**Sungho Park and Sachin Gupta | 5.15.2024**

The omnipresent error term in regression models does not always receive careful attention by model builders. What factors are included in this error? Naturally, it would be ideal if the error were entirely due to random shocks. However, sometimes factors that should be explicitly incorporated in the model but cannot be observed or are unavailable to be used as explanatory variables are also present in the error. Worse, often our accumulated knowledge and theories indicate that the variables seeping into the error term are systematically related to the explanatory variables included in the model. This results in regressor–error correlation, which, if ignored, leads to biased estimates.

## Why Does Regressor–Error Correlation Arise?

As an example, Heitmann et al. (2020) investigate the effect of two key visual design decisions: brand typicality (similarity within the brand's range) and segment typicality (similarity to the competitive set) on consumer purchasing. In the car market, visual appearance is a vital determinant of success; hence, automakers track consumers' changing tastes and strategically incorporate these into design changes in their newest models. However, because researchers typically cannot observe these changing tastes, they are encompassed in the error term. As a result, empirical models intended to measure the impact of design changes suffer from the problem that the key regressor capturing design changes is correlated with the error. Put differently, the design change regressor is endogenous. If not corrected, the estimated impact of design changes will be biased. As our academic field matures, we continue to discover reasons for regressor–error correlations that were previously overlooked. Examples of such phenomena are (1) advertising endogeneity due to self-selection by consumers in advertising response and (2) pricing endogeneity, because firms and consumers know aspects of product quality that researchers do not see in the data. While additional information such as instrumental variables or exogenous shocks can help address the endogeneity issue, obtaining such information is often challenging. In such situations, the copula correction method provides an alternative approach.

## How Does the Copula Correction Work?

The copula correction method directly addresses the issue of regressor–error correlation by assuming a plausible relationship between the endogenous regressor and the error. This additional structure enables the

researcher to estimate the model parameters without bias. However, the crucial underlying condition is that the assumed relationship between the endogenous regressor and the error is appropriate. Park and Gupta's (2012) copula correction method (P&G method hereinafter) assumes a general and convenient Gaussian copula–based relationship between the regressor and the error. The various advantages of the Gaussian copula are well known (Danaher and Smith 2011). The Gaussian copula covers nearly the full (–1, 1) range in pairwise correlation, making it a general and robust copula for most applications. Additionally, its complexity increases at a much slower rate than other multivariate models as the number of dimensions increases.

The P&G method has been extensively used in marketing in diverse contexts such as addressing potential endogeneity of product design changes (e.g., Heitmann et al. 2020), advertising content decisions (e.g., Guitart and Stremersch 2021), and marketing-mix variables (Datta et al. 2022). Since publication of Park and Gupta (2012), various methods that directly model the regressor–error relationship to avoid bias have evolved through subsequent studies. Interestingly, recent developments in this area explore the assumptions of the P&G method and make meaningful improvements by either relaxing them or suggesting alternatives that offer methodological benefits. Accordingly, with the goal of assisting applied researchers interested in employing the copula correction method, in this paper we revisit each assumption of the P&G method and illustrate how new methods enhance them.

The P&G method makes the following assumptions: (1) the endogenous regressor (let's call it X_en) is nonnormal, (2) the error follows a normal distribution, and (3) the dependence between the endogenous regressor and the error can be captured by a Gaussian copula. The model may include exogenous regressors (let's call them X_ex) along with the endogenous one. An implicit assumption made in the P&G method is (4) there is no correlation between exogenous and endogenous regressors.

Assumption 1 is easily testable and is often satisfied in many cases (below, we discuss methods that relax this assumption). Assumption 2 is a plausible assumption, commonly used in likelihood-based models or Bayesian models, but it can be violated, and empirical testing can be challenging, especially in situations with regressor–error correlation. Similarly, Assumption 3 is a plausible assumption but it cannot be easily tested (we will also discuss a method that relaxes Assumptions 2 and 3). Fortunately, Assumption 4 is easily testable. If X_ex and X_en are highly correlated, it is necessary to appropriately incorporate this correlation when constructing the model, as bias may arise otherwise. Haschka (2022) proposes a likelihood-based estimation method for this situation by constructing the joint distribution of the error and all explanatory variables to carry out the estimation. We also note that a number of other recently proposed methods account for the correlation between X_ex and X_en: a nonparametric control function method (Breitung, Mayer, and Wied 2024), 2sCOPE model (Yang, Qian, and Xie 2023), and SORE

model (Qian and Xie 2023). Table 1 summarizes all the assumptions of the P&G method, indicates whether they are testable, and suggests methods to consider in case the assumption is violated or to enhance robustness.

Park and Gupta (2012) demonstrate that the copula correction method can be applied to discrete choice models as well. The crucial first step in applying the copula correction method is appropriately deriving the linear form of regressor–error dependence. For instance, in the analysis of aggregate sales data, prominent models such as BLP (Berry, Levinsohn, and Pakes 1995) include linear regressor–error dependence between price and common shocks (i.e., price endogeneity). Once we obtain this linear form of regressor–error dependence, applying a copula correction method can address estimation issues.

## Recent Developments in Copula Correction Methods

Table 2 summarizes the key strengths of recently proposed methods. The data used in empirical marketing analyses often have a panel structure. When panel data sets have numerous cross-sectional units and relatively few time periods per unit, the challenges of estimation are addressed through fixed-effect transformation. Haschka (2022) extends Park and Gupta's (2012) approach to panel data where fixed-effect transformation is necessary. The concern when applying fixed-effect transformation is the presence of nonspherical errors. After resolving the problem of nonspherical errors through a generalized least squares transformation, Haschka develops a copula correction method based on the joint distribution of the error and all explanatory variables.

Table 1. Assumptions of the P&G Method and Recent Developments

| Assumption of the P&G Method | Testable? | Methods to Consider if the Assumption Is Violated or to Enhance Robustness |
|---|---|---|
| The endogenous regressor is nonnormal | Yes | • Yang, Qian, and Xie (2023)—2sCOPE model |
| The error follows a normal distribution | No | • Breitung, Mayer, and Wied (2024)—nonparametric control function method |
| The dependence between the endogenous regressor and the error can be captured by a Gaussian copula | No | • Qian and Xie (2023)—SORE model<br>• Breitung, Mayer, and Wied (2024)—nonparametric control function method |
| There is no correlation between exogenous and endogenous regressors | Yes | • Haschka (2022)<br>• Breitung, Mayer, and Wied (2024)—nonparametric control function method<br>• Yang, Qian, and Xie (2023)—2sCOPE model<br>• Qian and Xie (2023)—SORE model |

The copula correction method obtains unbiased estimates of model parameters by modeling the relationship between the regressor and the error. Of course, the true relationship between the two is unknown. The P&G method provides a plausible starting point, and adding other options is naturally beneficial for empirical research. By considering models based on alternative relationships between regressors and errors, researchers can conduct more robust analyses.

In the P&G method, the assumed regressor–error correlation based on Gaussian copula allows us to decompose the error into (a) the part correlated with the endogenous regressor and (b) pure exogenous shocks that are unrelated with all the regressors. Part (a) is expressed as a nonlinear function of the endogenous regressor, and this part plays a role very similar to a control function (for an overview of control functions, see, e.g., Navarro [2010] and Wooldridge [2015]). Breitung, Mayer, and Wied (2024) propose a novel "nonparametric control function method." In this approach, the control function that constitutes Part (a) follows a normal distribution, and Part (b) is a mean-zero shock that does not necessarily have to be normal. Consequently, Assumption 2 of the P&G method is relaxed. Similar to the P&G method, which assumes nonnormality of the endogenous regressor for model identification, the Breitung, Mayer, and Wied model requires that specific assumptions related to the distribution of the endogenous regressor be satisfied. While this approach originates from the idea of the copula correction approach, it has the advantage of not assuming a specific copula. Furthermore, Breitung, Mayer, and Wied formally demonstrate the consistency, asymptotic normality, and validity of bootstrap standard errors for the model parameters.

We turn next to Assumption 1, which is that the endogenous regressor has a nonnormal distribution. The recently proposed "two-stage copula endogeneity correction" (2sCOPE)

## Table 2. Key Strengths of Recently Proposed Methods

| Method | Strength |
|---|---|
| Haschka (2022) | Provides fixed-effect transformation to handle data with numerous cross-sectional units but relatively few time periods per unit |
| Breitung, Mayer, and Wied (2024)—nonparametric control function method | Provides a robustness check of the P&G method in cases where researchers cannot justify that (1) the error follows a normal distribution, and/or (2) the dependence between the endogenous regressor and the error follows the Gaussian copula (e.g., previous studies may argue that the error deviates from normality) |
| Yang, Qian, and Xie (2023)— 2sCOPE model | Allows for the application of the copula correction method when X_en follows a normal distribution, X_en and X_ex are correlated, and X_ex deviates from normality |
| Qian and Xie (2023)—SORE model | Handles discrete endogenous regressors with only a few levels, such as binary regressors or count-valued regressors with small means |

method relaxes this requirement (Qian and Xie 2023). Additionally, like Haschka (2022), 2sCOPE assumes that the endogenous regressors, exogenous regressors, and errors are interrelated through a Gaussian copula. For estimation it employs a two-stage approach using control functions derived from the assumed model. An advantage of the method is that it allows for consistent parameter estimation even if the endogenous regressor follows a normal distribution, as long as one of the correlated exogenous regressors deviates from normality.

As noted, the essence of the copula correction approach lies in directly modeling the correlation between regressors and errors to estimate model parameters without bias. The semiparametric odds ratio (SOR) has often been used in applied research in marketing and related fields as a flexible method to capture dependence between variables (see, e.g., Chen 2007; Qian and Xie 2011). The semiparametric odds ratio endogeneity (SORE) model has recently been proposed as a method that utilizes SOR to capture regressor–error dependence (Qian and Xie 2023). One notable advantage of SOR is its ability to handle the association between discrete endogenous regressors and the error effectively. While the P&G method can be applied to discrete endogenous regressors, it does not handle endogenous regressors with only a few levels well; examples are binary regressors or count-valued regressors with small means. This limitation arises because the P&G method treats discrete endogenous regressors as realizations from underlying continuous latent variables and performs an inverse mapping from the cumulative distribution functions of endogenous regressors to the latent variables. The SORE model addresses this issue. However, this benefit comes at a cost: SORE constructs a conditional distribution from the odds ratio (OR) function and nonparametric baseline distribution functions. If the OR function is misspecified, it can lead to bias and/or issues of model nonidentification.

One of the primary reasons researchers may choose to use SORE is its ability to handle binary endogenous regressors. A more classical solution in such cases is to employ a Gaussian copula–based approach with a structure similar to the models proposed by Heckman (1976) or Lee (1983). These models assume a specific relationship between the binary endogenous regressor and the error based on Gaussian copula. In this scenario, researchers can estimate the model without bias using conditional likelihood instead of the reverse mapping proposed in Park and Gupta (2012).

The robustness of the P&G method has been stress-tested by multiple subsequent studies. In Park and Gupta (2012), the copula correction method's performance was demonstrated in a simple setting without an intercept. Becker, Proksch, and Ringle (2022) show that the performance of copula correction in a more general setting when an intercept is included is diminished when the sample size is small. However, Qian, Xie, and Koschmann (2024) find that the substantial bias identified in Becker, Proksch, and Ringle is primarily due to their method of constructing the empirical copula. Specifically, the correction term for the empirical

copula, which is based on a fixed-value percentile for the highest rank, can significantly distort the distribution of the copula correction terms, resulting in suboptimal performance of the copula correction method. When the P&G method is applied more precisely, as suggested by Qian et al., the bias in the coefficient estimate of the endogenous regressor becomes negligible when the sample size reaches 400, rather than 4,000. Becker, Proksch, and Ringle also carefully examine the nonnormality assumption and how this assumption affects the results. In a similar vein, Eckert and Hohberger (2023) investigate the performance of the P&G method when various assumptions are violated, especially in cases of near-normal endogenous regressors, nonnormal and skewed errors, and the regressor–error correlation based on non–Gaussian copulas and provide guidelines for such scenarios. Like all models, copula correction methods rely on assumptions and naturally their use requires significant caution, especially when the sample size is small. Fortunately, the series of recent papers that have extended the original P&G method address many of these situations. More specifically, the issue of nonnormality can be mitigated in the 2sCOPE method. Problems related to skewed or nonnormal errors can be addressed through the nonparametric control function method. Moreover, an advantage of both SORE and the nonparametric control function methods is their flexibility to consider relationships between regressors and errors that do not necessarily follow a Gaussian copula.

## Guidance for the Applied Researcher

To wrap up, we suggest the following three-step procedure for researchers interested in applying the copula correction method.

- Step 1—Check whether the endogenous regressor follows a nonnormal distribution. If it is near normal, researchers can try the 2sCOPE model. If the endogenous variable is discrete and has only a few levels, such as binary regressors or count-valued regressors with small means, one can apply the SORE model. If the endogenous regressor follows a nonnormal distribution, proceed to Step 2.

- Step 2—Check for correlations between X_en and X_ex. If the correlations are large, apply the 2sCOPE model. If the data set has a panel structure and requires fixed-effect transformation to handle numerous cross-sectional units and relatively few time periods, apply the method proposed by Haschka et al. (2022). If there is low correlation between X_en and X_ex, apply the P&G method.

- Step 3—As a robustness check, consider running the nonparametric control function method if the endogenous regressor is continuous. Unfortunately, Assumptions 2 and 3 of the P&G method are not easily testable. The nonparametric control function method does not require the normality of the error (Assumption 2) or assume a

specific copula structure between the endogenous regressor and the error (Assumption 3). However, it does require an alternative set of assumptions, and some of these assumptions are also difficult to test using data. We suggest the nonparametric control function method as a robustness check because, like the P&G method, it is relatively easy to apply. Finding consistent results between the P&G method and the nonparametric control function method provides greater assurance of validity.

Additionally, we provide below answers to some frequently asked questions regarding the use of the copula correction method in practice.

*Q1: Is it correct to use multiple copula correction terms for multiple endogenous variables in the same model?*

A: This is correct. One advantage of the copula correction method based on the Gaussian copula is that it can include multiple copula correction terms to handle multiple endogenous regressors.

*Q2: In estimating a model with higher-order terms (e.g., interaction and quadratic terms) of the endogenous variable, should we generate additional copula correction terms for them?*

A: Qian, Xie, and Koschmann (2022) addresses this issue formally. They show that once copula correction terms for the main effects of endogenous regressors are included as generated regressors, there is no need to include additional correction terms for the interaction terms or higher-order terms. This simplicity in handling higher-order endogenous regression terms is a merit of the copula correction approach. More importantly, adding these unnecessary correction terms has harmful effects and leads to suboptimal solutions of endogeneity bias.

*Q3: Is it acceptable to exclude nonsignificant copula correction terms from the final model when it involves multiple copula correction terms?*

A: This issue is similar to a common challenge encountered in statistical analysis for which the final answer is not clear-cut: Should you exclude or include nonsignificant regressors when building the final model? Considering factors such as model complexity, influence on other variables, theoretical implications, and model fit, researchers may choose to drop nonsignificant regressors or leave them in the model. If the copula correction term is not significant, removing nonsignificant regressors in the final model can have positive effects in terms of model simplicity, degrees of freedom, and multicollinearity. We suggest examining how sensitive the estimates of key variables are when removing nonsignificant copula correction terms. If the effects of key variables are not very sensitive, removal may be harmless.

*Q4: Is it acceptable to utilize the significance of the copula correction term as an indicator to determine whether endogeneity is a concern?*

A: If the P&G assumptions are correct, the nonsignificance of the copula correction term implies that there is no endogeneity caused by the regressor–error correlation. While the assumptions of P&G can serve as a

plausible starting point, one cannot conclusively determine the absence of endogeneity based on this result alone. Therefore, it is advisable to consider other methods as a robustness check (e.g., nonparametric control function method, 2sCOPE).

## Conclusion

The copula correction method has extended beyond marketing and is increasingly being introduced and widely used in various fields, including management, economics, and psychology. Open-source code is also becoming widely available to implement the method (Gui et al. 2023). Concurrently, there has been substantial additional research on the assumptions and weaknesses of the original P&G model, leading to its development and evolution. As we know, there is no free lunch. To be able to conduct analysis without instrumental variables or additional information, copula correction methods must make assumptions about the relationship between regressors and errors. Through further research, we need to understand the relationship between regressors and errors better, both theoretically and empirically, and leverage this additional knowledge to develop a copula correction model that captures the regressor–error correlation more completely.


## AUTHORS

**Sungho Park** is Korbit Chaired Professor of Marketing at SNU Business School, Seoul National University.

**Sachin Gupta** is Henrietta Johnson Louis Professor of Management and Professor of Marketing at the SC Johnson College of Business, Cornell University.



## CITATION

Park, Sungho, and Sachin Gupta (2024), "A Review of Copula Correction Methods to Address Regressor–Error Correlation," Impact at JMR, (May 15), Available at: https://www.ama.org/marketing-news/a-review-of-copula-correction-methods-to-address-regressor-error-correlation/.

## ACKNOWLEDGMENT

We would like to express our gratitude to Kapil Tuli and Rebecca Hamilton for their valuable feedback and numerous helpful suggestions during the review process, which have contributed to enhancing the utility of this article.

## REFERENCES


Becker, Jan-Michael, Dorian Proksch, and Christian M. Ringle (2022), "Revisiting Gaussian Copulas to Handle Endogenous Regressors," Journal of the Academy of Marketing Science, 50, 46–66.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," Econometrica, 63 (4), 841–90.

Breitung Jörg, Alexander Mayer, and Dominik Wied (2024), "Asymptotic Properties of Endogeneity Corrections Using Nonlinear Transformations," Econometrics Journal (published online January 24),, https://doi.org/10.1093/ectj/utae002.

Chen, Hua Yun (2007), "A Semiparametric Odds Ratio Model for Measuring Association," Biometrics, 63 (2), 413–21.

Danaher, Peter J. and Michael S. Smith (2011), "Modeling Multivariate Distributions Using Copulas: Applications in Marketing," Marketing Science, 30 (1), 4–21.

Datta, Hannes, Harald J. van Heerde, Marnik G. Dekimpe, and Jan-Benedict E.M. Steenkamp (2022), "Cross-National Differences in Market Response: Line-Length, Price, and Distribution Elasticities in 14 Indo-Pacific Rim Economies," Journal of Marketing Research, 59 (2), 251–70.

Eckert, Christine and Jan Hohberger (2023), "Addressing Endogeneity Without Instrumental Variables: An Evaluation of the Gaussian Copula Approach for Management Research," Journal of Management, 49 (4), 1460–95.

Gui, Raluca, Markus Meierer, Patrik Schilter, and René Algesheimer (2023), "REndo: Internal Instrumental Variables to Address Endogeneity," Journal of Statistical Software, 107 (3), 1–43.

Guitart, Ivan A. and Stefan Stremersch (2021), "The Impact of Informational and Emotional Television Ad Content on Online Search and Sales," Journal of Marketing Research, 58 (2), 299–320.

Haschka, Rouven E. (2022), "Handling Endogenous Regressors Using Copulas: A Generalization to Linear Panel Models with Fixed Effects and Correlated Regressors," Journal of Marketing Research, 59 (4), 860–81.

Heckman, James J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," Annals of Economic and Social Measurement, 5 (4), 475–92.

Heitmann, Mark, Jan R. Landwehr, Thomas F. Schreiner, and Harald J. Van Heerde (2020), "Leveraging Brand Equity for Effective Visual Product Design," Journal of Marketing Research, 57 (2), 257–77.

Lee, Lung-Fei (1983), "Generalized Econometric Models with Selectivity," Econometrica, 51 (2), 507–12.

Navarro, Salvador (2010), "Control Functions," in Microeconometrics, Steven N. Durlauf and Lawrence E. Blume, eds. The New Palgrave Economics Collection. Palgrave Macmillan, 2-–28.

Park, Sungho and Sachin Gupta (2012), "Handling Endogenous Regressors by Joint Estimation Using Copulas," Marketing Science, 31 (4), 567–86.

Qian, Yi and Hui Xie (2011), "No Customer Left Behind: A Distribution-Free Bayesian Approach to Accounting for Missing Xs in Marketing Models," Marketing Science, 30 (4), 717–36.

Qian, Yi and Hui Xie (2023), "Correcting Regressor-Endogeneity Bias via Instrument-Free Joint Estimation Using Semiparametric Odds Ratio Models," Journal of Marketing Research, (published online August 3), https://doi.org/10.1177/00222437231195577.

Qian, Yi, Hui Xie, and Anthony Koschmann (2022), "Should Copula Endogeneity Correction Include Generated Regressors For Higher Order Terms? No, It Hurts," NBER Working Paper 29978, http://www.nber.org/papers/w29978.

Qian, Yi, Hui Xie, and Anthony Koschmann (2024), "A Practical Guide to Endogeneity Correction Using Copulas," NBER Working Paper 32231, http://www.nber.org/papers/w32231.

Wooldridge, Jeffrey M. (2015), "Control Function Methods in Applied Econometrics," Journal of Human Resources, 50 (2), 420–45.

Yang, Fan, Yi Qian, and Hui Xie (2023), "Addressing Endogeneity Using a Two-Stage Copula Generated Regressor Approach," NBER Working Paper 29708, https://www.nber.org/papers/w29708.

## About JMR

The *Journal of Marketing Research* delves into the latest thinking in marketing research concepts, methods and applications from a broad range of scholars. It is included in both the *Financial Times* top 50 business journals and the University of Texas Dallas research rankings journal list.

## About Impact at JMR

Impact at JMR highlights important and actionable content published in the *Journal of Marketing Research*. Executives can leverage this content in their organization. Professors can use it in their classrooms. Researchers can learn about distinct scholarly insights for their next big idea. We invite everyone to read, enjoy, learn, think and act.