# A PRIMER ON USING MONTE CARLO SIMULATIONS TO EVALUATE RESEARCH METHODOLOGY IN STRATEGIC MANAGEMENT

**S. TREVIS CERTO**
W.P. Carey School of Business
Arizona State University
PO Box 874006
Tempe, AZ 85287
trevis.certo@asu.edu

**JOHN R. BUSENBARK**
Mendoza College of Business
University of Notre Dame
South Bend, IN
jbusenba@nd.edu

**CHI HON LI**
Tilburg School of Economics and Management
Tilburg University
Tilburg, Netherlands
c.h.li@tilburguniversity.edu

**MICHAEL WITHERS**
Mendoza College of Business
University of Notre Dame
South Bend, IN
mwithers@nd.edu

Forthcoming in *Strategic Management Review*

# A PRIMER ON USING MONTE CARLO SIMULATIONS TO EVALUATE RESEARCH METHODOLOGY IN STRATEGIC MANAGEMENT


## ABSTRACT

Strategic management scholars increasingly rely on diverse research methodologies to investigate the field's central questions. In this paper, we advance knowledge by demonstrating how Monte Carlo simulations (MCSs) can serve as an essential tool for evaluating and enhancing these techniques. To do so, we first overview strategy research that has employed MCSs to adjudicate statistical or econometric techniques. We then describe the intuition and main steps for specifying and employing MCSs, which involve defining the research question, designing a data generation (DGP) process, analyzing data, and summarizing/communicating results. To facilitate the implementation of MCSs, we developed a custom generative pre-trained transformer (GPT) that enables researchers to produce the code (R, Stata, and/or Excel) necessary to design and execute MCSs. We close with a research agenda for strategy scholars interested in using MCSs to advance their own methodological pursuits.



**Keywords**: Monte Carlo simulations, custom GPT, data-generating process, statistical properties, research methodology

## INTRODUCTION

Since its inception, the field of strategic management has long focused on the firm as the key unit of analysis, seeking to understand how companies create and sustain competitive advantage in dynamic and uncertain environments (Nag et al., 2007; Drnevich et al., 2020). To this end, Rumelt et al. (1994) identified four fundamental questions central to strategy scholarship—how firms behave, why firms differ, what value headquarters provide, and what determines success in international competition—that continue to define the field's core imperatives (Leiblein and Reuer, 2020). But just as these questions represent the bedrock of strategic management research, they also underscore the inherent complexity researchers encounter when seeking to isolate organizational decisions and corresponding outcomes (Duhaime et al., 2021). Firms exist in a nexus of ever-evolving internal and external factors, including managerial judgment, organizational talent, internal competition for resources, competitive dynamics, and socio-institutional considerations, among so many others that ultimately drive heterogeneity in performance across businesses (Hamilton and Nickerson, 2003; Ghemawat and Levinthal, 2008; Leiblein et al., 2018; March and Sutton, 1997).

Stated plainly, then, the statistical models researchers employ cannot always capture the complex interdependencies often present in strategy research (Bettis and Blettner, 2020). Accentuating these complexities, empirical research in strategic management is also often messy (Natividad, In press; Lee, 2025). Instead of using carefully randomized controlled experiments in laboratories, strategy scholars commonly adopt archival data from sources to test hypotheses involving managerial decision-making and firm strategy. This reliance on secondary data stems from the fact that randomly assigning firms or top managers to treatment conditions is rarely feasible in strategic management research. Firms in these databases, however, vary dramatically

on a range of dimensions (Henderson et al., 2012) in ways that may confound hypothesis testing, causal inference, and knowledge accumulation to build and test theories (Hill et al., 2021). Taken together, these factors create substantial empirical challenges for researchers estimating causal relationships implied by strategy theory (Natividad, In press).

To help account for these empirical challenges, strategy researchers increasingly rely on advanced analytical and econometric techniques to test their hypotheses, but there remain concerns about the credibility and robustness of empirical findings in the field (Lee, 2025). It is, therefore, increasingly vital for scholars to understand the effectiveness of complex statistical techniques. Historically speaking, methodologists have often employed sophisticated mathematical equations to evaluate methodological approaches. More recently, however, strategy scholars have employed Monte Carlo simulation (hereafter MCS)[1] studies to "approximate difficult to track mathematical and statistical modeling problems" (Chalmers and Adkins, 2020: 248). This small (but growing) methods-oriented research in strategy has leveraged MCSs to illustrate and reveal how empirical protocols unfold in the presence of data rather than mathematical theory. For instance, strategy scholars have adopted MCSs to examine instrumental variables (e.g., Semadeni et al., 2014), sample selection (e.g., Certo et al., 2016), limited dependent variables (e.g., Zelner, 2009), omitted variables (e.g., Busenbark et al., 2022), and collinearity (e.g., Kalnins, 2018), among many others.

Simply stated, MCSs can be used to help researchers better understand how different data structures and properties influence the effectiveness of the estimators they employ. Using MCSs, researchers first build data with known data generation processes (DGPs) and then examine the

[1] Our use of the term "simulation" does not include other types of simulations, such as those used in computational modeling (Posen et al., 2024). We recognize that advancing methodology is only one application of Monte Carlo simulations, but it is sufficiently novel and valuable that we focus our research on that particular imperative.

effectiveness of different methodological techniques used to analyze the data. To this end, Chalmers and Adkins (2020: 248) suggest MCSs allow researchers "to independently repeat a computational experiment many times," which Lohmann et al. (2022: 2) note can inspire recommendations that "often reach seminal status." In this way, MCSs allow researchers to show—and not just tell—readers how certain techniques or estimators perform in different scenarios germane to a given domain or discipline.

Despite the benefits of research on the efficacy of statistical techniques, most organizational scholars are not trained in how to design or interpret MCSs (Sturman, 2023). After all, researchers publishing papers with MCSs often write to appeal to readers/reviewers who are familiar with the technique and less so for those not as acquainted with the method, which makes these papers difficult for some to understand (Lohmann et al., 2022) and obviates many of the benefits MCSs offer relative to econometric theory. To this end, many users of the insights from these methods studies—that is, empirical researchers seeking to test their hypotheses—may not be well-positioned to fully grasp the nuances of the corresponding MCSs and could therefore potentially misappropriate or misinterpret the message.

The objective of our work is to enhance empirical research in strategic management by providing a primer that explains the intuition underlying MCSs and demonstrates how they can serve as an essential tool for evaluating statistical techniques that strategy scholars may employ. While researchers in statistics (Sigal and Chalmers, 2016), epidemiology (Lohmann et al., 2022), and psychology (Chalmers and Adkins, 2020) have provided reviews of MCSs in different contexts and for different purposes, our primer elaborates how strategy scholars have employed the technique to advance research methodologies germane to organizational research. We also explain how researchers might design MCSs to mimic empirical contexts confronted by strategic

management scholarship, a pursuit we facilitate by developing and describing a custom generative pre-trained transformer (GPT) that enables researchers with almost any degree of methodological or coding ability to employ MCSs.

The structure of our paper includes three sections. In the first section, we provide background by explaining the reasons why strategy researchers might employ MCSs. We specifically highlight how strategy researchers primarily adopt MCSs to evaluate the effectiveness of statistical or econometric techniques. To help further emphasize the scholarly usage and importance of MCSs, we also review strategic management studies that have employed them to investigate methodological issues. Demonstrating the increasing relevance of MCSs, we notice that half of the papers in our review have appeared since 2020.

In the second section of our paper, we describe the basic steps of the simulation process. Specifically, we discuss how scholars develop research questions, design DGPs, summarize data, and interpret the results. We also overview the impact of both effect size and statistical power when designing MCSs. To help clarify these four steps, we illustrate how Semadeni et al. (2014) used each of the four steps in their investigation of endogeneity and instrumental variables.

When implementing the four steps of MCSs, one of the largest obstacles for researchers involves writing code, which is why various simulation-specific commands and packages have been written in different programming languages. To aid in the coding process, our custom GPT facilitates simulation protocols and makes them more accessible for researchers interested in employing the technique to investigate their own methodological inquiries. While this GPT can develop code for a variety of contexts and different software packages, we demonstrate its usefulness by providing a simple example illustrating how researchers can use this tool to generate multilevel data (an empirical reality for most strategy work). In contrast to other

research on MCSs that uses specific programs or packages that may become outdated, our dynamic GPT ensures scholars can utilize the best and newest coding practices for MCSs.

In our final section, we provide an agenda for scholars to use MCSs to advance strategy research, offering several ideas for future work. We highlight that almost all methodology scholarship adopting MCSs creates cross-sectional datasets with normally distributed variables (Certo et al., 2020; Busenbark et al., 2022; Semadeni et al., 2014), whereas strategy data are increasingly non-normal and exhibit nested characteristics such as a panel or multilevel structure (Certo et al., 2024b; Bliese et al., 2020). We also describe how researchers can employ MCSs as a means of generating data to mirror their own endeavors.

In sum, the evolving complexity of data and hypotheses in strategy research necessitates the use of advanced empirical methods. Yet, if these methods are applied inconsistently or incorrectly, even with the best of intentions, they can lead researchers astray from gaining true insight and can gradually stifle the growth of knowledge in our discipline. We are hopeful our contributions will aid strategy researchers in critically evaluating and enhancing the efficacy of the research methods we employ.

## USING MONTE CARLO SIMULATIONS

MCSs were developed in Los Alamos during the Manhattan Project in World War II to solve problems related to nuclear physics (Gill, 2015). MCSs involve using computers to generate repeated draws of random numbers with pre-specified properties to provide insights into complex problems (Lohmann et al., 2022). The codename Monte Carlo was chosen by the scientists because the random draws and probability sampling in MCSs reminded them of casino gambling (Gill, 2015). MCSs are based on the law of large numbers and the central limit theorem (for a review, see Dunn and Shultis, 2012). The law of large numbers states that the sample mean

converges to the population mean as the number of trials or observations increases. Relatedly, the central limit theorem indicates the mean of a random variable will approximately follow a normal distribution in sufficiently large samples (Zhang et al., 2023), such that MCSs "depend on this critical theorem" (Dunn and Shultis, 2012: 45).

**MCSs in Social Sciences**

Researchers have used MCSs for a variety of purposes beyond investigating problems in nuclear physics, but these disparate applications can make it difficult to understand their underlying intuition (Astivia, 2020). In the social sciences, especially strategic management, researchers often use MCSs to assess the effectiveness of statistical methods (Astivia, 2020). As Hallgren (2013: 44) states, "Although, in general, statistical questions can be answered directly through mathematical analysis rather than simulation, the complexity of some statistical questions makes them more easily answered through simulation methods." Stated bluntly, MCSs allow methodological scholars to investigate the efficacy of empirical models under a variety of conditions that may not be readily accessible (or are too esoteric) for mathematical theory. For instance, social scientists have used MCSs to evaluate the performance of estimators (Kenny et al., 2015), implications of violated model assumptions (Arnau et al., 2013), and impact of variables with irregular characteristics (Astivia, 2020). To illustrate the value of MCSs, Astivia (2020) reports that during a focal year, every article in the three leading journals in quantitative psychology incorporated MCSs in some way.

When examining the effectiveness of statistical methods, researchers assess three central properties of any empirical estimation technique—bias, efficiency, and consistency (Greene, 2018; Kennedy, 2008). *Bias* refers to the extent to which the estimated coefficient deviates from its true value (Semadeni et al., 2014). Scholars might examine bias, for example, by studying the

extent to which an estimator (e.g., OLS) reports coefficients that deviate from the true values specified in the simulation (we will describe this in more detail). *Efficiency* is evaluated with the precision of coefficient estimates, which scholars often evaluate by comparing how standard errors differ across different estimators. In the context of statistical theory, lower (higher) standard errors typically indicate higher (lower) levels of efficiency. Finally, *consistency* assesses whether the model produces stable and reliable results as the sample size increases (Wooldridge, 2010). Estimators are consistent when a parameter (e.g., coefficient) becomes closer to its true value as the sample size in the simulation increases. Researchers can use all three of these properties to evaluate the effectiveness of estimators or data properties for a given model.

**MCSs in Strategy Research**

Taking a cue from the various social science domains we briefly summarized in the preceding section, strategy scholars have also adopted MCSs (and increasingly so) to evaluate statistical properties or techniques. We now turn toward offering an overview of the strategy-leaning studies that have employed MCSs. We specifically review those that sought to evaluate bias, efficiency, and/or consistency across various conditions of statistical properties (e.g., variable distributions, nested data) and/or analytic techniques (e.g., OLS, probit regression).

To locate strategy research that has adopted MCSs for these purposes, we reviewed articles published from 1997 to 2024 in *Strategic Management Journal*, *Organization Science*, *Organizational Research Methods*, *Journal of Management*, and *Strategic Organization*, collectively representing the top management outlets that consistently publish methodological research. We searched Google Scholar using the keyword "simulation(s)" and then manually scrutinized each to ensure a focus on MCSs rather than other types, such as agent-based MCSs or N-K modeling. We excluded articles that only mentioned the term *Monte Carlo simulation*

without providing sufficient details, such as the number of iterations, sample size, or outcome measures, as these details are essential for a comprehensive understanding of its application. In the end, our protocol identified 18 strategic management studies that used MCSs, with half of these articles published since 2020.

After identifying these studies and evaluating their content, it became apparent that there exist four broad categories or applications (as it pertains to advancing research methodologies) of MCSs, all of which involve important issues in empirical research in strategic management. Specifically, as we delineate in the remainder of this section, we classify the studies as those that employ MCSs to examine variable distributions, endogeneity/omitted variables, self/sample selection, and nested data. We offer a summary of our review in Table 1.

*--Insert Table 1 about here--*

**Variable distributions**. Several papers in strategic management have used MCSs to understand how the probability distributions of variables may influence statistical analyses. For example, Henkel (2009) uses MCSs to examine how skewness in return distributions can confound the risk-return association between the mean and variance of firms' returns. Certo et al. (2024b) similarly explore the efficiency of modeling techniques such as OLS, winsorization, log-transformation, robust standard errors, bootstrapping, and quantile regression in the context of non-normally distributed dependent variables. Also involving the distribution of dependent variables, Zelner (2009) assesses the statistical significance of predicted probabilities associated with the changes in predictors specified for logit and probit models, both of which are necessitated by a binary outcome indicator. Along these same lines, Woo et al. (2023) examine how rare binary event rates influence coefficient estimates, standard errors, statistical power, and model convergence failures. Scholars have also adopted MCSs to investigate the veracity of

estimators when variables take the form of ratios. Certo and colleagues (2020), for example, study the potential issues of spuriousness and statistical power in the estimation of ratio variables, an issue Villadsen & Wulff (2021) extend by adjudicating whether fractional response modeling estimates more consistent parameters when a dependent variable is a proportion.

**Endogeneity and omitted variables**. Scholars in strategic management have used MCSs to explore the effects of endogeneity or omitted confounders on parameter estimates and model performance. In one of the first such studies, Semadeni et al. (2014) use MCSs to investigate the prevalence of bias from endogeneity and the degree to which various strengths of instrumental variables can ameliorate the bias. Building directly on this simulation, Busenbark et al. (2022) employ MCSs to assess whether certain magnitudes of omitted variables produce more bias, as well as the inefficiency of two-stage modeling, a concept Eckert and Hohberger (2023) also extend by exploring the consistency of Gaussian Copula estimation. Adopting a slightly more subtle treatment of omitted variables, Kalnins (2018) uses MCSs to examine the effectiveness of variance inflation factors (VIF) to detect multicollinearity driven by unobserved common factors. Along these lines, Frake et al. (2024) use MCSs to examine how model choices affect collider bias in executive compensation.

**Selection bias**. Scholars in strategy have also used MCSs to explore and address selection bias, whether it occurs in the form of sample selection or self-selection. Certo et al. (2016) employ MCSs to examine different conditions under which sample selection bias may occur, including the correlation between the error terms in the first and second stages of the model, a notion Wolfolds and Siegel (2019) further elaborate to examine the role of exclusion restrictions in Heckman's two-step estimation. Clougherty et al. (2016) also show via their simulation that self-selection biases parameter estimates in OLS, indicating the importance of

using appropriate techniques such as the Heckman selection model. Whereas those studies tend to focus on selection into the sample, Balasubramanian et al. (2024) employ MCSs to investigate how assumptions about selection on the independent variable influence model results.

**Nested data**. Strategy researchers have also used MCSs to study nested (e.g., multi-level, longitudinal) data structures. Certo and Semadeni (2006), for instance, employ MCSs to explore the influence of contemporaneous correlation, heteroskedasticity, and autocorrelation on panel data analysis, particularly when using the OLS regression model. Parker and Witteloostuijn (2010) use MCSs in the context of moderation to test their proposed General Interaction approach against extant fit estimations, including fit-as-moderation, fit-as-deviation, and fit-as-system perspectives. Similarly, Sharapov et al. (2021) assess the performance of the Shapley Value approach for variance decomposition across levels in nested data, using MCSs to compare the Shapley Value approach to traditional techniques like ANOVA, multilevel models, and variance component analysis.

## DESIGNING MONTE CARLO SIMULATION STUDIES

In the previous section, we reviewed how strategy scholars have employed MCSs to study a variety of statistical issues with the central intent of advancing methodological pursuits. In this section, we review the primary steps for designing MCSs. Broadly speaking, MCSs involve generating data and variables with pre-specified properties (e.g., distributions, sample size, variable relationships) that the researcher can manipulate to examine the efficacy of different empirical models. As such, MCSs share some characteristics with experiments in the sense that scholars can manipulate conditions to investigate their research question or methodological focus (e.g., Beisbart and Norton, 2012). To do so, MCSs typically feature four interrelated elements: defining the research question, designing the DGP, analyzing data, and

summarizing model results. In the following sections, we review these four steps. To aid in understanding, we briefly review how Semadeni et al. (2014) implemented each of these steps to investigate endogeneity and instrumental variables. We also present these steps in Table 2.

*--Insert Table 2 about here--*

**Step 1: Defining the Research Question(s)**

The first step in using MCSs entails defining an appropriate research question. While empirical research in strategy often tests hypotheses based on theory, as we emphasized in the previous section, MCSs can be employed to examine questions about the efficacy of analytical techniques and/or the implications of data structures. In contrast to conceptual work, effective research questions in the simulation context are typically formed after reviewing practices in the literature (e.g., Woo et al., 2023), which can also highlight inconsistencies and/or the scope of a problem (e.g., Kalnins, 2018). Examples of broad research questions include: "How do various effect sizes influence the impact of endogeneity?", "How can I use MCSs to better understand the best ways to analyze panel data?", or "How does the impact of sample selection change as the variables are (non)normally distributed to different degrees?", among many other potential lines of inquiry. In their study of endogeneity, Semadeni et al. (2014) defined their research objective to "provide a series of MCSs to illustrate the consequences of endogeneity and the robustness of the techniques prescribed to circumvent these consequences" (p. 1071).

**Step 2: Designing the Data Generation Process (DGP)**

Simulation design is fundamentally predicated on modeling a DGP. Whereas empirical researchers typically collect data and then use models to estimate the relationships (e.g., coefficients) between the variables, MCSs require the inverse approach. Specifically, scholars specify the distributions and relationships between the variables (e.g., correlations, coefficients)

and then use those specifications to generate the data.

**Variable distributions**. The first step in designing a DGP involves specifying the distributions of the variables of interest. Researchers can indicate the distributions of the dependent, independent, and control variables (e.g., continuous, count, dichotomous). In this stage, scholars can then define the means and standard deviations (or other moments depending on the distributions) of all variables in the DGP (e.g., Eckert and Hohberger, 2023). In addition to specifying variables with normal distributions, researchers might also generate variables following non-normal distributions (e.g., gamma, Cauchy). When possible, decisions in this regard should match the research context the researcher is trying to study. Although simulating non-normally distributed data is often challenging, given the dearth of pre-existing commands in statistical packages allowing users to generate random variables that follow other distributions, the custom GPT that we describe later makes this process remarkably more accessible.

**Relationships between variables.** Researchers must also stipulate the relationships between the variables of interest. As a simple example, researchers might use a DGP that results in groups with different means due to a treatment effect. As another example, researchers may simply generate a series of variables with pre-specified correlations (Certo et al., 2020). Alternatively, researchers might use a slope-based approach, whereby an algebraic equation generates the dependent variable as a function of independent variables and an error term (e.g., Busenbark et al., 2022). The slope-based process randomly draws a set of regressors (i.e., independent variables and an error term) and then calculates the dependent variable as a function of those variables and corresponding coefficients. Many simulation studies generate covariates and error terms with specified correlations or means in the first step and then use a slope-based process in the second step to generate the dependent variable as a function of those variables and

error term (Kalnins, 2018; Semadeni et al., 2014). Based on the reviewed studies contained in Table 1, using the slope-based DGP represents the most common approach.

**Specifying effect sizes**. One of the key components of any DGP involves specifying appropriate effect sizes, which "refers to the magnitude of the relation between the independent and dependent variables" (Funder and Ozer, 2019: 156). Establishing the desired relationship between the independent and dependent variables in a simulation depends on the type of DGP the researcher uses. A correlation-based DGP combines a randomly drawn $y$ that has a pre-specified correlation with $x$ (e.g., Certo et al., 2020; Kim et al., 2022), making the effect size (at least in terms of correlations) explicit. For the slope-based DGP, altering the coefficient or changing the relative variance of the independent variable or error term in any way will change the correlation between $y$ and $x$. One challenge with the slope-based approach is researchers must calculate the correlation between the variables after generating the data and then iterate model conditions to achieve the desired effect size, which can require some trial and error. When simulating differences in group means, researchers might use variations of Cohen's $d$ as a measure of effect size (Kallogjeri and Piccirillo, 2023). Importantly, these various effect size measures can be converted back and forth using different formulas (Kelley and Preacher, 2012).

**Determining statistical power**. After determining the appropriate effect size, researchers must also focus on statistical power, which is the probability a scientific investigation or statistical test would lead to a statistically significant result (Cohen, 1992). In the context of MCSs, researchers typically determine statistical power by specifying the percentage of iterations that report a statistically significant result (at any desired threshold), with higher percentages representing higher levels of statistical power. As we will describe later, deviations from the specified level of statistical significance—which can take any value between 0 and 1—

would result in Type I or Type II errors. Research in psychology suggests scholars should aim for power levels of approximately 0.80 (Correll et al., 2020), meaning 80 percent of the iterations in a simulation would exhibit a statistically significant estimate. In the domain of MCSs, it is important to set baseline power levels that allow for fluctuations in both directions across conditions (e.g., around 65 to 80 percent).

Turning back to Semadeni et al. (2014), these scholars designed the DGP whereby the dependent variable ($y$) was a function of an independent variable ($x$) multiplied by a coefficient, which they set to .1, as well as an error term ($e$). To investigate endogeneity, they varied the correlation between $x$ and $e$ (i.e., 0, .1, and .3). Based on their effect sizes and endogeneity levels, they used sample sizes of 100, 500, and 1000 to investigate statistical power.

**Step 3: Analyzing Data and Creating Outcome Measures**

Researchers using MCSs can run a statistical model on each generated dataset and save the corresponding estimates, which represents the central advantage of MCSs over mathematical theory. To do so, scholars iterate the DGP and the empirical models many times over, storing the results from each iteration. Generally speaking, there are two categories of simulation outcomes. The first type involves the characteristics of the dataset. For instance, researchers might save the moments (e.g., means, standard deviations, etc.) of $x$ and $y$, the correlations between the variables, or any other data characteristic that appears relevant. The second category involves model parameter estimates. In the context of OLS regression, for example, researchers might save the coefficient, standard error, and p-value associated with each independent variable, as well as model-level outcomes such as R-squared or the F-statistic. Of course, these estimates vary depending on the type of model employed (e.g., OLS vs. logit).

Extending our running example, Semadeni et al. (2014) ran OLS and 2SLS models on the

generated data and recorded parameter estimates (e.g., betas, standard errors) for each iteration. As we will discuss next, this enabled them to compare both bias and efficiency across each type of estimator and various simulated conditions. Given the nature of the empirical question at hand—involving the impact of endogeneity—Semadeni and colleagues also retained the statistical significance of Durbin-Wu-Hausman tests for each condition.

**Step 4: Summarizing Results**

After running the statistical models of interest across all the simulation iterations, researchers can then use the parameter estimates from the analytic model to evaluate the three central properties of any empirical estimation technique: bias, efficiency, and consistency (Kennedy, 2008; Greene, 2018). In Table 3, we depict several summarized results scholars can consult to quantify the extent to which the simulation exhibits bias and efficiency. As Semadeni et al. (2014) describe, researchers can examine the average or median value (and/or its standard deviation) of the estimated coefficient across all the simulated iterations. Bias exists if that value deviates from the specified coefficient value. Whereas inferential statistics are aimed to help approximate the frequency the model would estimate the true value in repeated hypothetical samples, MCSs allow scholars to examine how often this occurs in actual repeated samples across all the simulation iterations.

In addition to examining bias, it is possible to also assess the efficiency of each model. Efficiency is evaluated with the precision of coefficient estimates. Typically, researchers work to triangulate efficiency by consulting either the standard error (Certo et al., 2024b) or statistical significance (Semadeni et al., 2014) of the estimates. In the case of MCSs, researchers can scrutinize the average (and standard deviation) of the standard errors across all the simulated conditions. Alternatively, they could evaluate the percentage of simulated iterations in which a

coefficient is statistically significant at any desired threshold. All else being equal, and assuming no bias that results in Type I or Type II errors, models with a higher (lower) percentage of statistically significant estimates would indicate higher (lower) levels of efficiency.

Finally, researchers can scrutinize the consistency of different specifications or modeling approaches. Doing so entails noting how simulation results change as the sample size increases, enabling scholars to ascertain if the model converges on the true properties (i.e., the specified relationships between variables) as the number of observations grows. Intriguing, our review of MCSs in strategy research suggests researchers are more interested in bias and efficiency than consistency, as scholars tend to emphasize the degree to which coefficients deviate from their specified value and exhibit Type I or Type II errors rather than how much estimates converge as a function of increases in the sample size. This preference may stem from the fact that empirical strategy studies tend to have large sample sizes, which may dampen concerns about consistency.

Returning to our example, Semadeni et al. (2014) summarized the parameter estimates in their MCSs by creating median values across the 1,000 iterations for each condition. In other words, Semadeni and colleagues simulated observations for a given condition, employed OLS and 2SLS models, retained the estimates from those models, performed that process 1,000 times over, and then ultimately reported the median values across those 1,000 iterations for each condition. The authors then compared the median estimates across conditions and to the true values (which they specified in their DGP via Step 2) to draw conclusions about the effects of endogeneity on OLS and 2SLS at various different strengths of instruments. Comparing the median betas to the true values enabled them to evaluate bias, and comparing standard errors allowed them to evaluate efficiency.

*---Insert Table 3 about here---*

## A PRE-TRAINED GENERATIVE TRANSFORMER (GPT) FOR MCSs

In addition to delineating the steps and intuition involved in creating MCSs (i.e., our focus thus far), we now work to make MCSs in strategy research more accessible by describing a custom GPT we specifically designed to assist researchers seeking to employ MCSs via various software packages (namely R, Stata, and/or Excel). Our custom GPT guides users through the entire simulation process from conceptualization to implementation and interpretation. The custom GPT can be accessed via: https://tinyurl.com/simcustomgpt (the GPT link is located at the Wiki section on the home page). Our approach to developing the custom GPT follows the recent efforts in other disciplines to leverage custom GPTs for specialized academic and professional tasks (e.g., Aykut and Sezenoz, 2024; Masters et al., 2024).

Custom GPTs build upon the foundational knowledge of ChatGPT's large language model, GPT-4o, but involve more idiosyncratic and tailored information from the designer than would be available in the broader ChatGPT platform outside of the custom GPT. Indeed, custom GPT designers can enhance these for specific applications by incorporating domain-specific information and providing detailed instructions to guide user interactions (Sevgi et al., 2024). By providing domain-specific information, the custom GPT can provide more accurate and user-friendly responses to end-user prompts (Collins et al., 2024). A designer of the custom GPT can provide detailed guidance in terms of how the custom GPT interacts with end users by inputting specific instructions in the custom GPT's configuration window.

For our custom GPT, we provided specific open-access information related to simulation design and execution for R, Stata, and Excel. We also provided detailed instructions related to its interaction with an end user. We set it up to follow a structured, step-by-step approach, asking one question at a time and waiting for user responses before proceeding. In asking these

questions, the custom GPT covers various aspects of the simulation process, including initial assessment, research question clarification, simulation design, code development, result interpretation, and visualization. We also directed the custom GPT to tailor its language to the user's experience level, offer clear explanations, and emphasize methodological rigor and reproducibility. The instructions were set up to avoid having the custom GPT make assumptions about the user's prior knowledge or provide overly technical language without explanation.

**Using the GPT to Employ MCSs**

Following a similar approach to extant research, our custom GPT was developed to help scholars better understand answers to general questions about MCSs, articulate simulation objectives, and translate complex methodological concepts into actionable steps by providing detailed code. An important benefit of our GPT is that it can provide explanations tailored to the user's level of expertise, especially by allowing users to ask multiple questions or to solicit the GPT to rephrase a response. In this section, we describe the overall application of our GPT.

**General questions about MCSs**. The most basic use for our custom GPT involves helping researchers understand the intuition underlying simulation techniques. Although we previously highlighted the major steps in MCSs, researchers can engage the GPT to dive into specific questions. For instance, scholars might ask the GPT, "What types of DGPs are best for MCSs investigating models of firm performance?" or "Explain how MCSs help to understand statistical bias and efficiency." Inquiries like these will prompt high-level responses from the GPT that the user can then iterate to gain as much (or as little) insight as desired.

**Articulation of simulation objectives**. The custom GPT engages a researcher in an iterative dialogue to refine the overall objectives of the study. This interaction can help a researcher understand specific methodological issues and/or apply empirical techniques

(Megahed et al., in press). In the process of clarifying research objectives, it also directs the researcher to consider the key methodological considerations relevant to the focal methodological question. Drawing from its knowledge of simulation techniques, informed both by our design and the information otherwise available on the ChatGPT platform, the custom GPT suggests appropriate approaches to help tailor the simulation design to specific research or learning needs. It can also assist in refining the scope of the simulation study by having discussions about the trade-offs between complexity and interpretability. In doing so, it can assist the researcher in determining the optimal number of parameters to vary and the range of conditions to examine with the simulation.

**Coding**. In our view, the most useful aspect of the GPT is its ability to translate research questions into code that researchers can reproduce in R or Stata. Although other reviews of MCSs provide relevant code for their particular lines of inquiry (e.g., Certo et al., 2016; Balasubramanian et al., 2024), these papers can become outdated, and they provide code that readers must learn, interpret, and modify for applications other than the focal study at hand. This is particularly true as programming languages and software continue to develop, as there may be more updated statistical packages that enhance analytical efficiency. By contrast, our GPT allows users to generate customized code matching the researcher's objective and draws from the most current statistical resources available, which dramatically lowers the bar to entry to create MCSs.

For example, our GPT can aid a researcher in specifying the distributions of variables in the DGP. Recognizing many variables in strategy research deviate from the normal distribution (Certo et al., 2024b), the custom GPT can guide users in generating non-normally distributed variables by offering specific coding packages/commands or requisite formulas that extend

beyond those commands. It, in turn, provides code snippets and explanations for creating variables with specific distributional properties, such as skewness or kurtosis.

Similarly, for researchers working with panel or multilevel data, which are common in strategic management, the custom GPT can offer guidance on creating nested data structures in the DGP. Users can specify within-firm and between-firm variance components, ensuring the simulated data accurately represents the complexities of longitudinal or hierarchical strategy-focused datasets. It can also assist in incorporating more advanced features into the DGP, including endogeneity, selection bias, extreme outliers, and heteroskedasticity, which are often present in strategy data. The customized GPT can also suggest diagnostics to verify that the DGP meets the intended specifications and provide guidance on adjusting parameters to achieve desired characteristics. This iterative process ensures the final DGP produces data that closely aligns with the researcher's intended scenario and reflects realistic strategy phenomena.

**An Illustration Using the GPT to Create Simulation Code**

In this section, we illustrate how to employ a simulation in Stata by using prompts provided to (and then received from) our custom GPT. Our goal is to create simulation codes for a DGP with firm-year panel data, employing a random-effects regression model, and then ultimately retaining the estimated coefficients and standard errors. We include the screenshots of the conversations with the custom GPT (Figures 1 to 5).

**Employing a simulation with the GPT**. To help researchers, we include four broad interactable options from the onset of the GPT initial screen: designing a simulation in Stata, designing a simulation in R, helping interpret the results from our existing MCSs, and providing the general steps for a simulation (Figure 1). It is important to highlight that these options are example prompts and are not required for using the GPT. Instead, they represent a starting point

for users who may feel uncomfortable initiating an idiosyncratic conversation with the GPT, whereas more sophisticated researchers are able to bypass them completely and start by entering their own inquiries. For example, a researcher could get started by indicating, “Create a simulation examining the effectiveness of random effects models.” Combined, these options give researchers the flexibility to ask specific simulation-related questions for their studies. Table 4 includes a number of prompt examples researchers might use when interacting with the GPT.

*--Insert Figures 1 to 5 and Table 4 about here--*

In this illustration, we aim to obtain simulation codes for a regression model, so we select the option: “I am designing a simulation in Stata. Can you help?” After submitting our inquiry, we go through the model specification where the assistant responds by asking us to describe the type of simulation we want to design—whether we are working on a regression model, generating a random dataset, or testing a specific hypothesis (Figure 2). We reply that we are working with a regression model. The assistant then asks for more details, including the type of regression model (e.g., linear or complex) and the number of iterations we plan to run. We indicate our intention to run a panel regression with 1,000 iterations. Next, we specify the model and number of observations, explaining that we need a dependent variable $y$, a random continuous independent variable $x$ with a coefficient of 0.1, a random binary variable $w$ with a coefficient of 0.5, and a normally distributed error term. We also require a panel dataset consisting of 1,000 units (e.g., firms) over a span of 10 years (i.e., 10,000 total observations in the sample, which then gets iterated 1,000 times over).

It is imperative to emphasize all of the values depicted here are simply for demonstration purposes and are not necessarily best practices for designing a simulation. We suggest that researchers start with a thorough literature review when designing MCSs to get a sense for

reasonable effect sizes, sample sizes, and variable distributions. Researchers can use the GPT, for example, to ask about realistic parameters by employing prompts such as "What are typical sample sizes in strategy research?" or "What type of distribution should I use for the dependent variable?" or "How much between-firm variance should I include when generating firm performance as a dependent variable?" Scholars might also consider downloading data directly from databases to better understand the properties of their variables. Certo et al. (2024b), for example, downloaded data from Compustat to examine the non-normality of performance variables and designed a DGP to mimic these variables.

The custom GPT then provides us with the steps for creating the simulation code in Stata (Figures 3 and 4). The code begins by setting the seed to ensure reproducibility. Next, it sets up the program, named "*mysim*," to specify the panel data structure, with each firm having a unique ID over 10 years. The code also generates the firm-level error term for the random effects. As requested, the independent variable *x* is created using a normal distribution, while the binary variable *w* follows a binomial distribution with equal probabilities of 0 and 1. With the full regression model, the code returns the coefficients and standard errors from each iteration. After running the simulation 1,000 times within the "*mysim*" program, we can obtain the final outputs, including the summary statistics for the average coefficients and standard errors, along with their means and standard deviations over iterations. The assistant also provides explanations of the codes for each step (Figure 5). Overall, the custom GPT offers user-friendly and detailed guidance, making it accessible for researchers to conduct simulation analyses and meet their research needs, even if those researchers have little or no simulation training.

**Revising the simulation and error messages**. In our experience, designing MCSs with this GPT is an iterative process. At times, the codes provided by the GPT sometimes result in

error messages when tested in R or Stata, with the latter receiving slightly more errors than the former owing to its proprietary (rather than open source) nature. However, it is easy to simply copy error messages and paste them directly into the GPT as prompts. When doing so, the GPT will search for alternative codes or approaches, a process that can be applied repeatedly.

## DISCUSSION AND CONCLUSION

Our primary objective was to describe the intuition and processes used to conduct MCSs with the intention of enhancing the quality of methodological procedures in strategic management. Despite the scholarly impact of multiple strategy articles adopting MCSs, many strategy scholars lack familiarity with the technique. We have sought to address this tension by providing a comprehensive overview of the simulation process, introducing a custom GPT to assist scholars with simulation research, and outlining future directions for simulation-based research in strategy to continue to better investigate the efficacy of empirical techniques.

Our study offers several methodological contributions to strategic management. First, we provide a comprehensive overview of MCSs. We explain the intuition of MCSs and review the reasons why scholars in the social sciences use them. We then review how strategy researchers have adopted MCSs to examine the overall effectiveness of a variety of statistical techniques. Our overview serves not only to help researchers interpret published simulation studies but also to encourage broader adoption of MCSs in strategy research.

Second, to make MCSs more accessible, we also developed, implemented, and introduced a custom GPT designed specifically to aid researchers in conducting and interpreting MCSs. This GPT guides users step-by-step through the process of crafting MCSs as they go from idea conceptualization to implementation to interpretation of the results. By providing tailored assistance and generating executable code for R and Stata, the custom GPT has the potential to

lower the barriers for scholars interested in creating MCSs. We hope our GPT tool will help to accelerate the adoption and further advancement of simulation techniques, which can potentially lead to new methodological insights in strategy research.

Our custom GPT offers several advantages to strategy researchers who seek to run MCSs, increasing both accessibility and rigor. Many strategy researchers, while experts in their content areas, may lack the advanced programming or statistical skills required to develop complex MCSs. Our custom GPT bridges this knowledge gap by providing step-by-step guidance and code generation that enables a broader range of researchers to use these techniques. By helping to refine research questions and design an appropriate DGP, the GPT assists researchers in simulating data that better represent the complexities of strategy research. Scholars can also learn about different aspects of simulation design and analysis through their interactions with the custom GPT, potentially enhancing their methodological skills. It also has the potential to help promote reproducibility in research by providing clear, documented code for each step of the simulation process. Further, by streamlining the simulation process, researchers are able to focus more on the methodological implications and interpretations of their results rather than becoming mired in technical details, resulting in richer methodological contributions from simulation studies.

Finally, in the following section, we offer an agenda for future research applying MCSs. With this research agenda, we emphasize the need for MCSs that better reflect the realities of strategy data. Most simulation-based strategy research has assumed variable normality and the absence of a nesting structure (Semadeni et al., 2014; Busenbark et al., 2022). Yet recent work suggests strategy research often features some type of nesting structure (typically panel data) and extreme non-normality in the distribution of key variables (Certo et al., 2017; Certo et al.,

2024b). Accordingly, we assist in rectifying the misalignment between MCSs and reality by offering a research agenda and tool that can aid researchers in examining these dynamics. We also highlight several opportunities to leverage MCSs to further the field's understanding of statistical concepts—including systems of equations, probability distributions of outcomes, and statistical power—that are emerging as important methodological considerations. We hope our proposed research agenda will direct future simulation studies to align more closely with the empirical contexts strategy researchers often encounter.

**Future Directions and Applications for MCSs**

We believe advances in computing power and artificial intelligence have the potential to dramatically increase the use of MCSs in strategy research. Historically, coding expertise has created an entry barrier for scholars interested in using MCSs, an issue almost entirely resolved by AI. Whether it is our customized GPT or other AI interfaces (e.g., DeepSeek, Gemini), these algorithms can create customized MCSs in seconds that would require days from even the most experienced coders. Accordingly, we believe these tools will not only open the door to researchers who are new to coding, but they will also help researchers who are familiar with coding to conduct such research more quickly and effectively.

With these new tools in hand, we now shift our attention to elaborating future directions or applications of the simulation procedures we have delineated thus far. In the following sections, we detail what we think are some of the most exciting new directions for simulation research in strategic management. As we describe, these directions can help researchers understand some of the unique characteristics of data used in strategy research.

**Nested (or panel) data.** Despite the notable methodological advancements from the literature that uses MCSs to examine empirical techniques—please see Table 1 for a summary of

these studies—research in the area has relied almost exclusively on cross-sectional data. Part of the reason for this focus is likely because multilevel or nested data introduce new empirical challenges that detract from the core message of a given study (Kennedy, 2008) and also because simulating such data is somewhat difficult in its own right (a challenge we help alleviate with our GPT). At the same time, strategy scholars routinely describe how the broader domain relies to a large degree on panel (firms nested within years) and multilevel (firms nested within industries, etc.) data (Certo et al., 2017; Shaver, 2021), as well as the unique empirical challenges associated with this type of structure (McNeish and Kelley, 2019; Bliese et al., 2020).

We surmise that simulation-based methodological research would benefit from considering empirical inquiries in the context of both cross-sectional and nested data to determine (a) whether the focal models or estimation issues manifest uniquely in both data structures and (b) if there exist any underexplored techniques that may help scholars navigate the challenges associated with nested data. Similarly, perhaps it is the case that unexplained heterogeneity—a topic that has garnered attention from several studies employing MCSs (e.g., Semadeni et al., 2014; Busenbark et al., 2022; Certo et al., 2016)—impacts outcomes differently depending on the level of the data. Along these same lines, scholars have adopted MCSs to examine probabilistic models (Zelner, 2009; Woo et al., 2023), and nested data could also help illustrate idiosyncrasies associated with conditional (i.e., fixed effects) logistic modeling. Although these are just a few examples of methodological areas that could stand to benefit from fuller incorporation of nested data in MCSs, many existing MCSs could be extended to examine the potential effects of nested data structures.

**Non-normally distributed variables.** With the exception of research on binary dependent variables (Zelner, 2009; Woo et al., 2023) and one study specifically about non-

normality (Certo et al., 2024b), most studies that adopt MCSs tend to generate normally distributed variables. Much as was the case for nested data, this focus is likely due to the fact that non-normality can induce a host of empirical issues unrelated to the question at hand, as well as because simulating non-normally distributed variables is challenging, to say the least (an issue we again help resolve with our custom GPT). Yet, scholars indicate that many variables germane to strategy research depart heavily from a normal distribution and thus undermine the assumptions of most parametric estimators (Henderson et al., 2012; Certo et al., 2024b). Accordingly, methodological work adopting MCSs could benefit a great deal from coupling the issues at the heart of any particular study with the implications of the non-normal data.

One area that could especially advance extant research methods knowledge involves marrying the empirical issues associated with non-normality with those stemming from nested data. To the best of our knowledge, it remains unclear whether the efficiency challenges from non-normally distributed dependent variables (Wooldridge, 2010) would manifest in a panel (or nested) data situation owing to the increased power in such samples and the diffusion of variance across multiple levels. Similarly, it remains unclear whether the distribution of observed and omitted variables amplifies or attenuates bias from unexplained variance. Perhaps it is the case that correlations between specified regressors and the error term are driven by outlying observations or a portion of the distribution that gets adjusted by variable transformations or modeling procedures. In the end, essentially all of the research that has adopted MCSs could be replicated with various distributions of the key parameters, if for no other reason than to ensure the central conclusions remain intact.

**Systems of equations.** With the exception of multi-stage instrumental variable techniques geared toward reducing bias from unexplained heterogeneity, most strategy research

involves estimating relationships in a single structural model (Shaver, 2021; Ketchen et al., 2008). In other words, scholars tend to focus on a single equation that estimates a marginal impact on the dependent variable for a unit change in an independent variable, sometimes as being contingent or contextualized on a third parameter (Aguinis et al., 2017). At the same time, more micro-oriented management research routinely recognizes that a structural equation (and the marginal effect on the dependent variable) is often path dependent on preceding variables or initial stage specifications, an idea often addressed via structural equation modeling or mediation (Kline, 2015; Aguinis et al., 2017). In a structural equation or mediation model, the change in a dependent variable ($y$) is a function of the impact on an independent or mediating variable ($x$) and a preceding variable ($z$) that works to predict both $y$ and $x$.

Broadly speaking, strategy research has largely overlooked this actuality in no small part due to the complexities of macro-oriented archival data and the pursuit of addressing the several model violations we summarized previously. MCSs, however, allow scholars to craft a known universe of data with all other assumptions upheld. So, we envision such an approach as being fruitful for revisiting several traditional macro-oriented econometric topics—e.g., limited dependent variables (e.g., Long and Freese, 2014), panel data (e.g., Certo et al., 2017), difference-in-difference (e.g., Roth et al., 2023), endogeneity (e.g., Semadeni et al., 2014), and many others—in the context of systems of equations and/or mediation. Perhaps it is the case that lower-order equations in the system can exacerbate bias via their own violated assumptions, or maybe properly specified lower-order equations can help alleviate estimation challenges in a structural model by offering more insight into the focal variables. There is no doubt that MCSs can provide a great deal of insight into these methodological challenges.

**Using MCSs as customized power calculators.** Strategy researchers have emphasized

the challenges associated with null hypothesis significance testing (Bettis et al., 2016), one of which is that it is difficult to ascertain whether a "significant" estimate represents a statistical anomaly or is representative of a true population effect (Goldfarb and King, 2016). Stated differently, after researchers obtain data and run empirical models, it often remains unclear whether the null hypothesis was rejected (or failed to get rejected) because of actual relationships or artifacts of the data. To offer more insight in this regard, we propose that scholars can employ MCSs to get a better sense of the likelihood their model *should* reject the null hypothesis based on the underlying properties of the data.

In many ways, these ideas extend the seminal statistical power tables produced by Cohen (1992), which are certainly valuable but perhaps not as effective for researchers using non-normal variables and more complex research designs. To help better contextualize statistical power for data and empirical settings that strategy scholars confront, researchers can specify the sample size of a simulation using the number of observations in their data, the correlation(s) between a vector of independent variables and their dependent variable using those correlations from those data, and then their estimator(s) of interest. Researchers can then compute the percentage of iterations with a statistically significant result following the procedure we outlined previously, which can help contextualize the findings from their own data.

Although not infallible, using simulations as customized power calculators can provide crucial insight to help guide empirical strategies. In one scenario, a simulation might reveal a relatively high percentage of significance, such that insignificant findings from the actual data may signal an irregularity, or significant findings may indicate consistency with the underlying data. In another scenario, a relatively low percentage of significance may indicate that a significant finding from the actual data demonstrates the novelty of the theory and sample, or it

could point to an anomalous outcome. Alternatively, MCSs illustrating a low percentage of significance might also help researchers reconsider their data approach or empirical context when their own findings are insignificant rather than assume the hypothesis is unsupported.

In many ways, our proposal here extends a nascent line of inquiry that has begun using MCSs to examine what relationships between particular conceptual/theoretical variables *should* exist absent certain model assumption violations. For instance, Frake et al. (2024) adopt MCSs to revisit the relationship between having a female CEO and the career outcomes of other women in the organization by simulating a scenario without supposed collider bias. In doing so, they suggest that findings in the literature may represent statistical artifacts rather than true associations between the variables. Similarly, Balasubramanian et al. (2024) incorporate MCSs to examine employee financial imperatives in the presence of employment restrictions that are rarely observed in practice. Using MCSs, they explore elements of the employee and agreement selection process that archival data simply do not reveal but that confer important implications for the models that test relationships in this area.

**Limitations**

Despite the potential for MCSs to improve empirical research in strategic management, it is important to note the limitations of MCSs. Perhaps the most important limitation is that MCSs involve the generation and analysis of artificial data that simplifies reality. Although the most effective MCSs will attempt to generate data that mimic the research context—an imperative that we help amplify and enable with our custom GPT—it is impossible to replicate reality perfectly. As such, authors, reviewers, and editors must recognize no simulation is perfect, but the most effective MCSs will attempt to mimic reality.

We should also note that, in our experience, some reviewers may lean too heavily on this

criticism when evaluating papers using MCSs. Specifically, reviewers may call for "real" data instead of simulated data to evaluate statistical techniques. While we appreciate this sentiment, it belies the benefits of MCSs, namely that researchers never know the "true" DGP of data collected from the real world. For instance, researchers may evaluate the efficacy of two different models using real data, but it is impossible to determine which is "true" because the real-world data may include other issues the researcher is not controlling (e.g., endogeneity, reverse causality). In contrast, MCSs allow researchers to specify the exact data generation process, allowing for a more controllable context in which to study bias, efficiency, and consistency. Physicist and Nobel laureate Richard Feynman once famously wrote, "What I cannot create, I do not understand" (Way, 2017: 2941). Using MCSs—unlike "real" data—allows researchers to understand true relationships because they created every aspect.

Another limitation of MCSs involves computing power. Although most machines are able to accommodate MCSs with great expediency, even the most powerful processors can sometimes encounter challenges with MCSs involving sophisticated algorithms like machine learning (Stolfi and Castiglione, 2021) or Bayesian analyses (Certo et al., 2024a). Given that such techniques advance in parallel with improvements in computing power, we anticipate this is a challenge that may require scholars to use supercomputers when seeking to adopt MCSs in these contexts.

We should also acknowledge the limitations of using GPTs and similar tools to generate simulation code. Large language models operate as black boxes and are known to occasionally produce incorrect information. In our experience, we have found that GPTs also produce errors when generating computer code. But we have also found that these tools are able to interpret error messages and easily adjust code, making them useful for iterative debugging. Relatedly, just as drop-down menus and point-and-click user interfaces have allowed researchers to run

sophisticated statistical models in software packages without fully understanding the associated complexities (e.g., assumptions, options), there are dangers associated with authors blindly accepting AI-generated code. AI—whether our customized GPT or other tools—may not generate the code an inexperienced author believes it is providing. Importantly, this issue of code not representing the intent of an author has always plagued empirical research—even before AI. As such, it is imperative for researchers to understand the myriad issues surrounding both DGPs and statistical models.

**Conclusion**

In this paper, we provide an overview of MCSs aimed at improving the methodological rigor in strategy research. Our review highlights important characteristics that are unique to strategy research and demonstrates how our custom GPT can help accelerate the process required for researchers to learn and employ MCSs. Coupled with our research agenda, we hope that scholars find MCSs a powerful tool for advancing strategy research.

## REFERENCES


Aguinis H, Edwards JR and Bradley KJ (2017) Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods* 20(4): 665-685.

Arnau J, Bendayan R, Blanca MJ and Bono R (2013) The effects of skewness and kurtosis on the robustness of linear mixed models. *Behavioural Research* 45: 873-879.

Astivia OLO (2020) Issues, problems and potential solutions when simulating countinous, non-normal data in the social sciences. *Meta-Psychology* 4: 1-16.

Aykut A and Sezenoz AS (2024) Exploring the potential of code-free Custom GPTs in ophthalmology: An early analysis of GPTs store and user-creator guidance. *Ophthalmology and Therapy*.

Balasubramanian N, Starr E and Yamaguchi S (2024) Employment restrictions on resource transferability and value appropriation from employees. *Strategic Management Journal*.

Beisbart C and Norton JD (2012) Why Monte Carlo simulations are inferences and not experiments. *International Studies in the Philosophy of Science* 26(4): 403-422.

Bettis RA and Blettner D (2020) Strategic reality today: Extraordinary past success, but difficult challenges loom. *Strategic Management Review* 1(1): 75-101.

Bettis RA, Helfat CE and Shaver JM (2016) The necessity, logic, and forms of replication. *Strategic Management Journal* 37(11): 2193-2203.

Bliese PD, Schepker DJ, Essman SM and Ployhart RE (2020) Bridging methodological divides between macro-and micro research: Endogeneity and methods for panel data. *Journal of Management* 46(1): 70-99.

Busenbark JR, Yoon HE, Gamache D and Withers M (2022) Omitted variable bias: Examining management research with the impact threshold of a confounding variable (ITCV). *Journal of Management* 48(1): 17-48.

Certo ST, Albader LA, Raney KE and Busenbark JR (2024a) A Bayesian approach to nested data analysis: A primer for strategic management research. *Strategic Organization* 2(22): 241-268.

Certo ST, Busenbark JR, Kalm M and LePine JA (2020) Divided we fall: How ratios undermine strategic management research. *Organizational Research Methods* 23(2): 211-237.

Certo ST, Busenbark JR, Woo H-S and Semadeni M (2016) Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal* 37(13): 2639-2657.

Certo ST, Raney K, Albader L and Busenbark JR (2024b) Out of shape: The implications of (extremely) nonnormal dependent variables. *Organizational Research Methods* 27(2): 195-222.

Certo ST and Semadeni M (2006) Strategy research and panel data: Evidence and implications. *Journal of Management* 32(3): 449-471.

Certo ST, Withers M and Semadeni M (2017) A tale of two effects: Using longitudinal data to compare within- and between-firm effects. *Strategic Management Journal* 38(7): 1536-1556.

Chalmers RP and Adkins MD (2020) Writing effective and reliable Monte Carlo simulations wiht the SimDesign package. *The Quantitative Methods for Psychology* 16: 248-280.

Clougherty JA, Duso T and Muck J (2016) Correcting for self-selection based endogeneity in management research: Review, recommendations and simulations. *Organizational Research Methods* 19(2): 286-347.
Cohen J (1992) A power primer. *Psychological bulletin* 112(1): 155-159.
Correll J, Mellinger C, McClelland GH and Judd CM (2020) Avoid Cohen's 'small', 'medium', and 'large' for power analysis. *Trends in Cognitive Science* 24: 200-207.
Drnevich PL, Mahoney JT and Schendel D (2020) Has Strategic Management Research Lost Its Way? *Strategic Management Review* 1(1): 35-73.
Duhaime IM, Hitt MA and Lyles MA (2021) *Strategic management: State of the field and its future.* Oxford, UK: Oxford University Press.
Dunn WL and Shultis JK (2012) *Exploring Monte Carlo Methods.* Oxford, UK: Academic Press.
Eckert C and Hohberger J (2023) Addressing endogeneity without instrumental variables: An evaluation of the Gaussian Copula approach for management research. *Journal of Management* 49(4): 1460-1495.
Frake J, Hagemann A and Uribe J (2024) Collider bias in strategy and management research: An illustration using women CEO's effect on other women's career outcomes. *Strategic Management Journal*.
Funder DC and Ozer DJ (2019) Evaluating effect size in psycologica research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science* 22: 156-168.
Ghemawat P and Levinthal D (2008) Choice interactions and business strategy. *Management Science* 54(9): 1638-1651.
Gill J (2015) *Bayesian Methods: A social and behavioral sciences approach.* Boca Raton, FL: CRC Press.
Goldfarb B and King AA (2016) Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal* 37(1): 167-176.
Greene WH (2018) *Econometric Analysis.* New York: Pearson.
Hallgren KA (2013) Conducting simulation studies in the R programming environment. *Tutorials in Quantitative Methods for Psychology* 9: 43-60.
Hamilton BH and Nickerson JA (2003) Correcting for endogeneity in strategic management research. *Strategic Organization* 1(1): 51-78.
Henderson AD, Raynor ME and Ahmed M (2012) How long must a firm be great to rule out chance? Benchmarking sustained superior performance without being fooled by randomness. *Strategic Management Journal* 33(4): 387-406.
Henkel J (2009) The risk-return paradox for strategic management: disentangling true and spurious effects. *Strategic Management Journal* 30(3): 287-303.
Hill AD, Johnson SG, Greco LM, et al. (2021) Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research. *Journal of Management* 47: 105-143.
Kallogjeri D and Piccirillo JF (2023) A simple guide to effect size measures. *JAMA Otolaryngology–Head & Neck Surgery* 149(5): 447-451.
Kalnins A (2018) Multicollinearity: How common factors cause type 1 errors in multivariate regression. *Strategic Management Journal* 39(8): 2362-2385.
Kelley K and Preacher KJ (2012) On effect size. *Psychological Methods* 17(2): 137-152.
Kennedy P (2008) *A Guide to Econometrics.* Oxford, UK: Blackwell.
Kenny DA, Kaniskan B and McCoach Db (2015) The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research* 44: 486-507.

Ketchen DJ, Boyd BK and Bergh DD (2008) Research methodology in strategic management: Past accomplishments and future challenges. *Organizational Research Methods* 11(4): 643-658.
Kim YM, Busenbark JR, Jeong S-H and Lam SK (2022) The performance impact of marketing dualities: A response surface approach to resolving empirical challenges. *Journal of the Academy of Marketing Science* 50: 915-940.
Kline RB (2015) *Principles and practice of structural equation modeling.* New York - London: Guilford Publications.
Lee GK (2025) Surprises, conflicting findings, or questionable research practices? A methodology for evaluating cumulative empirical analyses and replication studies. *Strategic Management Review* 6(3): 273-292.
Leiblein MJ and Miller DJ (2003) An empirical examination of transaction-and firm-level influences on the vertical boundaries of the firm. *Strategic Management Journal* 24(9): 839-859.
Leiblein MJ and Reuer JJ (2020) Foundations and Futures of Strategic Management. *Strategic Management Review* 1(1): 1-33.
Leiblein MJ, Reuer JJ and Zenger T (2018) What makes a decision strategic? *Strategy Science* 3(4): 558-573.
Lohmann A, Astivia OLO, Morris TP and Groenwold RHH (2022) It's time! Ten reasons to start replicating simulation studies. *Frontiers in Epidemiology*. 1-9.
Long JS and Freese J (2014) *Regression models for categorical dependent variables using Stata.* College Station, TX: Stata press.
March JG and Sutton RI (1997) Organizational performance as a dependent variable. *Organization Science* 8(6): 698-706.
Masters K, Benjamin J, Agrawal A, et al. (2024) Twelve tips on creating and using custom GPTs to enhance health professions education. *Medical Teacher* 46(6): 752-756.
McNeish D and Kelley K (2019) Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods* 24(1): 20.
Megahed FM, Chen Y-J, Ferris JA, et al. (in press) How generative AI models such as ChatGPT can be (mis)used in SPC practice, education, and research? An exploratory study. *Quality Engineering*. 1-29.
Nag R, Hambrick DC and Chen MJ (2007) What is strategic management, really? Inductive derivation of a consensus definition of the field. *Strategic Management Journal* 28(9): 935-955.
Natividad G (In press) Econometrics of strategy: A review. *Strategic Management Review*.
Parker SC and Witteloostuijn Av (2010) A general framework for estimating multidimensional contingency fit. *Organization Science* 21(2): 540-553.
Posen HE, Wang Mz, Chen JS and Elfenbein DW (2024) In defense of diversity in theory-building approaches. *Academy of Management Review* 49(1): 199-205.
Roth J, Sant'Anna PH, Bilinski A and Poe J (2023) What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*.
Rumelt RP, Schendel D and Teece DJ (1994) *Fundamental issues in strategy: A research agenda.* Boston, MA: Harvard Business School Press.

Semadeni M, Withers MC and Certo ST (2014) The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations. *Strategic Management Journal* 35(7): 1070-1079.
Sevgi M, Antaki F and Keane PA (2024) Medical education with large language models in ophthalmology: custom instructions and enhanced retrieval capabilities. *British Journal of Ophthalmology*.
Sharapov D, Kattuman P, Rodriguez D and Velazquez FJ (2021) Using the SHAPLEY value approach to variance decomposition in strategy research: Diversification, internationalization, and corporate group effects on affiliate profitability. *Strategic Management Journal* 42(3): 608-623.
Shaver JM (2021) Evolution of quantitative research methods in strategic management. In: Duhaim IM, Hitt MA and Lyles MA (eds) *Strategic management: State of the field and its future*. Oxford, UK: Oxford University Press.
Sigal MJ and Chalmers RP (2016) Play it again: Teaching statistics with Monte Carlo simulation. *Journal of Statistics Education* 24: 136-156.
Stolfi P and Castiglione F (2021) Emulating complex simulations by machine learning methods. *BMC Bioinformatics* 22: 1-14.
Sturman MC (2023) Real research with fake data: A tutorial on conduction computer simulation research and training. *Organizational Research Methods*.
Villadsen AR and Wulff JN (2021) Are you 110% sure? Modeling of fractions and proportions in strategy and management research. *Strategic Organization* 19(2): 312-337.
Way M (2017) What I cannot create, I do not understand. The Company of Biologists Ltd, 2941-2942.
Wolfolds SE and Siegel J (2019) Misaccounting for endogeneity: The peril of relying on the Heckman two-step method without a valid instrument. *Strategic Management Journal* 40(3): 432-463.
Woo H-S, Berns JP and Solanelles P (2023) How rare is rare? How common is common? Empirical issues associated with binary dependent variables with rare or common event rates. *Organizational Research Methods* 26(4): 655-677.
Wooldridge JM (2010) *Econometric analysis of cross section and panel data.* Cambridge, MA: MIT press.
Zelner BA (2009) Using simulation to interpret results from logit, probit, and other nonlinear models. *Strategic Management Journal* 30(12): 1335-1348.
Zhang X, Astivia OLO, Kroc E and Zumbo BD (2023) How to think clearly about the central limit theorem. *Psychological Methods* 28: 1427-1445.

**Table 1**. Literature review on Monte Carlo simulation in strategy research

| I. Studies | II. Simulation Purpose | III. Iterations | IV. Samples Size | V. Outcome Measures | VI. Findings |
|---|---|---|---|---|---|
| Certo and Semadeni (2006) | To study the impacts of contemporaneous correlation in panel data in strategic management research. | 1,000 | 200 | Overconfidence (the ratio of the standard deviation of the 1,000 estimates to the root mean square average of the 1,000 standard errors) and 95% confidence interval. | Contemporaneous correlation and heteroskedasticity affect the outcome measures, while autocorrelation does not. |
| Henkel (2009) | To show how skewness can affect the measurement of risk-return relationship, and how it can be addressed. | 100 | 1,000 | Spurious contribution to the measured covariance, correlations, and standard deviations. | The spurious effect of skewness reversed the signs of the estimated correlations. |
| Zelner (2009) | To capture the changes in predicted probabilities given the change in the value of the predictors. | 1,000 | 469 (from the original Leiblien and Miller's (2003) study) | Estimated coefficients, 95% confidence interval, and difference in predicted probabilities. | To showcase the use of Monte Carlo simulation for easier statistical and graphical interpretation. |
| Parker and Witteloostuijn (2010) | To compare the General Interaction approach with different contingency fit estimating techniques. | 10,000 | 18, 39, 90, and 163 observations, depending on the models. | Detection of true interaction effects, measurement error, statistical robustness under small sample conditions, and multicollinearity. | The General Interaction approach outperforms traditional fit methods, especially when handling interactions between multiple contingency variables. |
| Semadeni et al. (2014) | 1) Study the implications of endogeneity in OLS, 2) the strength and exogeneity of instruments, and 3) how both affect the test to detect endogeneity. | 1,000 | 500 | Median coefficients, median standard errors, 95% confidence interval, and percentage significance. | Low endogeneity can bias OLS results. Both weak and moderator instruments can provide more accurate estimates but with lower statistical power. Strong and exogenous instruments can help detect endogeneity. |

| | | | | | |
|---|---|---|---|---|---|
| Certo et al. (2016) | 1st study: When sample selection process bias results. 2nd study: How to use the Heckman model. 3rd study: How the Heckman model addresses endogeneity from sample selection. | 1,000 | 1,000 | Average coefficients, standard errors, percent significance, correlations, and Pseudo $R^2$. | Heckman model can address potential sample selection bias, requiring careful assessment of lambda. |
| Clougherty et al. (2016) | To show the presence of self-selection bias and various techniques associated with this issue. | 500 | 250 and 1,000 (endogenous treatment); 500 and 2,000 (endogenous switching) | Mean coefficients and mean standard errors. | When there are endogenous treatment and switching, OLS will provide biased estimates. |
| Kalnins (2018) | To test the implications of common-factor multicollinearity. | 10,000 | 1,000 | Average coefficients, variance inflation factor, and Condition Index. | When the controls share a common factor with the key predictors, this key variable will have a higher likelihood of suffering a Type 1 error. |
| Wolfolds and Siegel (2019) | To evaluate the performance of the selection models under different functional forms and variable assumptions. | NA. Only used for generated simulated data. | 10,000 | Coefficients and standard errors. | Heckman model performs better than OLS, regardless of the error term distribution, with valid instruments or selection on observables. |
| Certo et al. (2020) | To study whether ratio variables will produce inaccurate estimates. | 1,000 | 1,000 | Median coefficient and percentage significance. | When the predictor or outcome is a ratio, the predicted relationship will fluctuate when the dispersion of the denominator changes. Ratio as a control will also affect the estimated relationship. |

| | | | | | |
|---|---|---|---|---|---|
| Sharapov et al. (2021) | To assess whether the Shapley Value approach performs better than the ANOVA, multi-level, and VCA approaches. | 1,000 | 1,000 | Estimated proportion of total variance and root mean square errors. | The Shapley Value approach generates more accurate estimates of variance contributed compared with other techniques. |
| Villadsen and Wulff (2021) | To study the differential performance of various fractional outcome approaches. | 1,000 | 800 | Marginal effects at the mean and robust standard error. | Using the log-odds transformation or Tobit model will generate bias in estimating marginal effects. |
| Busenbark et al. (2022) | To examine how the omitted variables cause bias in a causal estimation. | 1,000 | 1,000 | Estimated coefficients and standard deviations. | Omitted variables need to have high correlations with the predictors to cause meaningful bias. |
| Eckert and Hohberger (2023) | To demonstrate the strengths and weaknesses of the Gaussian Copula approach. | 500 | 500 | Median coefficients, t-bias measure, median standard errors, percentage significance of the coefficients, and correction terms in the Gaussian Copula approach. | Gaussian Copula approach works as well as instrumental variable approaches as long as its assumptions are satisfied. |
| Woo et al. (2023) | To examine the empirical issue related to the use of rare and common event models. | 1,000 | 100 to 1,500 | Average estimated coefficients, standard errors, percentage significance, and percentage of model convergence. | More rare events will bias estimated coefficients and standard errors, causing inaccurate causal inferences. |
| Balasubramanian et al. (2024) | To examine whether the potential impacts of selection into restriction adoption influence the main empirical results. | 100 | 1,000 | The proportion of the variance explained by the predictors. | The causal treatment effects are not affected much by the selection effect. |
| Certo et al. (2024b) | To investigate the implications of the | 10,000 | 1,000 | Mean coefficients, standard errors, percentage | Nonnormal distribution of a dependent variable will |

| | | | | | |
|---|---|---|---|---|---|
| | nonnormality of the dependent variables. | | | significance, and the average number of outlier observations. | decrease the efficiency of OLS regression. |
| Frake et al. (2024) | To test in what conditions the collider bias is more serious. | 100 | 1,000 | Coefficients and p-values. | When there are TMT members with the wage premium paid to the individuals who will become CEOs or smaller TMT sizes, the bias increases. When the proportion of women on the TMT increases, the bias decreases and remains negative. |

**Table 2**. Steps for Designing Simulations

| Steps | Descriptions | Examples |
|---|---|---|
| 1. Defining the research questions | • What methodological issue am I probing (e.g., endogeneity, bias, sample selection)? | "How can I use simulations to illustrate the consequences of endogeneity and the robustness of the techniques prescribed to circumvent these consequences?" |
| 2. Designing the data generation process | • Choose variables & distributions.<br>- Dependent variable y (normal/skewed?)<br>- Key independent variable x (continuous/binary?)<br>- Error term e (normal/heteroskedastic?)<br>• Specify relationships (equation or correlation matrix).<br>• Set experimental conditions (sample size, iterations, parameters to vary). | • Equation: $y = 0.1 \times x + e$.<br>• Vary corr(x, e) = 0, 0.1, 0.3 to induce endogeneity.<br>• Sample = 500; iterations = 1,000. |
| 3. Analyzing data and creating outcome measures | • Select empirical model(s) to estimate (e.g., OLS, 2SLS, RE logit).<br>• Store parameter estimates & fit statistics for every iteration. | Run OLS and 2SLS on each synthetic dataset; save β, SE, p-value, $R^2$ for 1,000 iterations. |
| 4. Summarizing results | • Aggregate estimates across iterations (means, medians, SDs).<br>• Compute diagnostics:<br>- Bias = mean(β) – true β<br>- Efficiency = mean(SE)<br>- Consistency = trend as N increases<br>• Compare methods/conditions. | Averaged βs and SEs across iterations; compared to true values to judge bias and efficiency of OLS versus 2SLS. |

**Table 3.** Summarizing outcome variables from the simulation

| Summary Variable | Calculation | Purpose/Interpretation |
|---|---|---|
| Coefficient | Any given coefficient (e.g., $B\text{x}_1$) across all the iterations in the simulation. | Used to gauge bias/inaccuracy of the estimator. Mean coefficient values closer to zero than specified exhibit suppression bias, whereas those further from zero than specified exhibit inflation bias. |
| Standard error | Standard error for any given parameter (e.g., S.E. of $B\text{x}_1$) across all the iterations in the simulation. | Used to gauge efficiency. Mean standard error values smaller than specified represent more efficient models, whereas those higher than specified represent less efficient models. |
| Percent significant | Percentage of the simulation iterations in which the parameter estimate is statistically significant (at any desired threshold). | Used to gauge statistical power (and/or types of error). Percent significant values that exceed the specified level represent Type I error, meaning false positives. Percent significant values that fall short of the specified level represent Type II error, meaning false negatives. |
| Correlation[y, x] | Correlation between two variables across all the iterations in the simulation. | Used to gauge bias/inaccuracy of the DGP. When adopting the slope-based DGP, the mean/median correlation should remain consistent across all iterations when the effect size is held constant. Changes in these values can indicate issues with the DGP. |
| r-squared | Model r-squared (or adjusted r-squared) values across all the iterations in the simulation. | Used to gauge model-level effect sizes, which can provide insight into bias/inaccuracy. The model is upwardly biased when r-squared values exceed their specified value, and the model is downwardly biased when r-squared values fall short of their specified value. |
| Other model-level characteristics | Any relevant model-level output (e.g., MSE, f-statistic, log-likelihood) across all the iterations in the simulation. | Researchers are able to examine any model-level characteristics desired to determine whether changing a condition alters model estimates, or whether these appear to fluctuate within conditions (indicative of issues with the DGP). |

**Table 4.** Example Prompts to Use in Customized GPT

1. Variable distributions
    a. How do I change the distributions of the variables in the simulation?
    b. How do I make the dependent variable follow a skewed distribution?
    c. How can I compare different variable distributions in my simulation?
    d. How can I generate multilevel or longitudinal variables in my simulation?
    e. How do I create variables with values contingent on another indicator?

2. Data Generation Process (DGP)
    a. How do alter the effect size of the independent variable in my DGP?
    b. How might I examine the role of different effect sizes in my simulation?
    c. How do I incorporate control variables that are correlated with an independent variable in my DGP?
    d. How might I incorporate endogeneity in my DGP?
    e. Are there other ways to specify the DGP?

3. Outcome measures
    a. What are the best outcome measures based on my research objectives?
    b. How do I incorporate outcome measures in my simulation?
    c. How do I interpret and evaluate the results of the outcome measures?
    d. How can I store marginal effects for interaction effects and models that use maximum likelihood estimation?

4. Different models
    a. How might I add different types of models to my simulation?
    b. What types of models might I consider given my research objectives?
    c. How can I compare how different types of models perform in my simulation?
    d. How do I couple Bayesian estimation with [any given model]?

5. Sample size
    a. How would different sample sizes influence the results of the simulation?
    b. How can I add different conditions examining the role of sample size in my results?
    c. How can I determine the minimum sample size required to ensure robust and reliable simulation results?

6. Iterations
    a. How does changing the number of iterations impact my results?
    b. How do I increase the number of iterations in the simulation?
    c. How do I determine the optimal number of iterations needed for my simulation to achieve convergence?
    d. How do I set the seed to ensure an identical starting point each time I employ the iterations?

7. Visualization
    a. Give me some ideas to visualize the simulation results.
    b. I would like to plot the distribution of betas from each iteration.
    c. Can you recommend techniques to visualize the convergence and stability of my simulation over iterations?

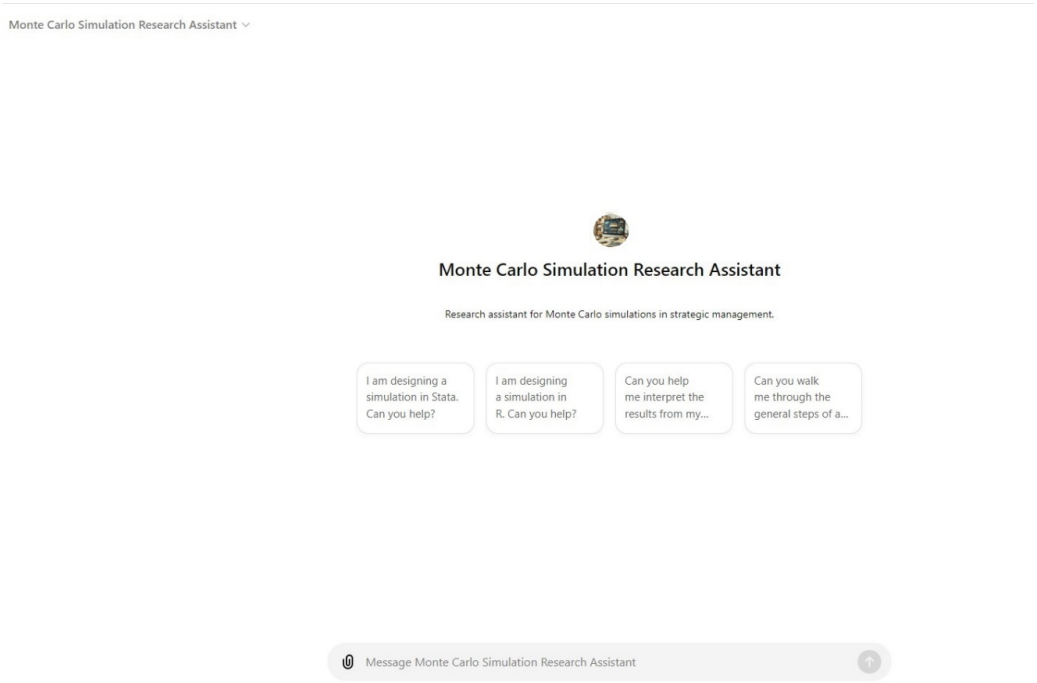


**Figure 1**. Screenshot of the custom GPT – Problem overview

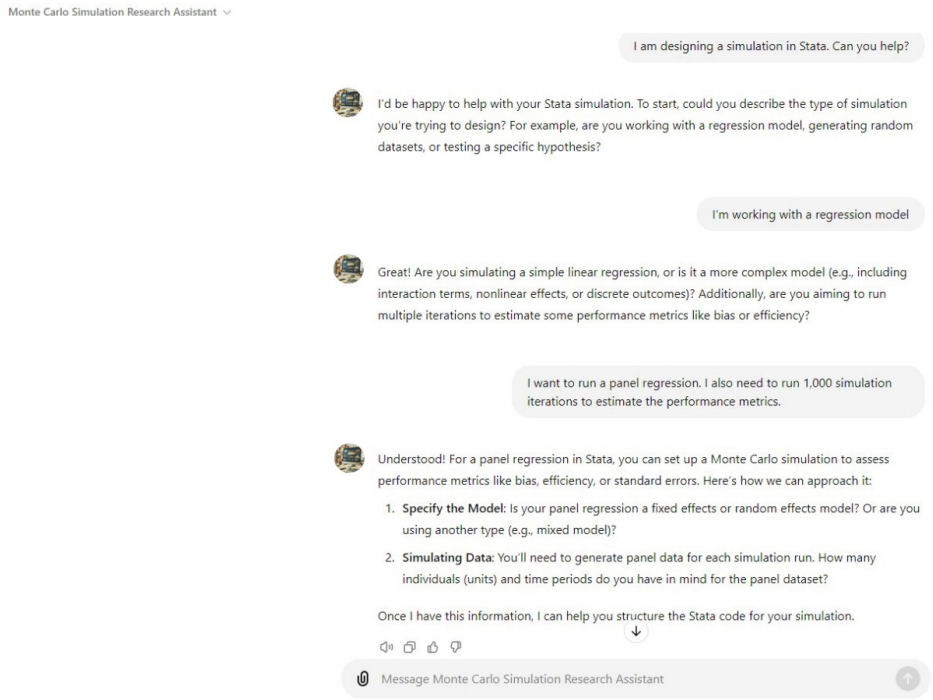


**Figure 2**. Screenshot of the custom GPT – Model specification

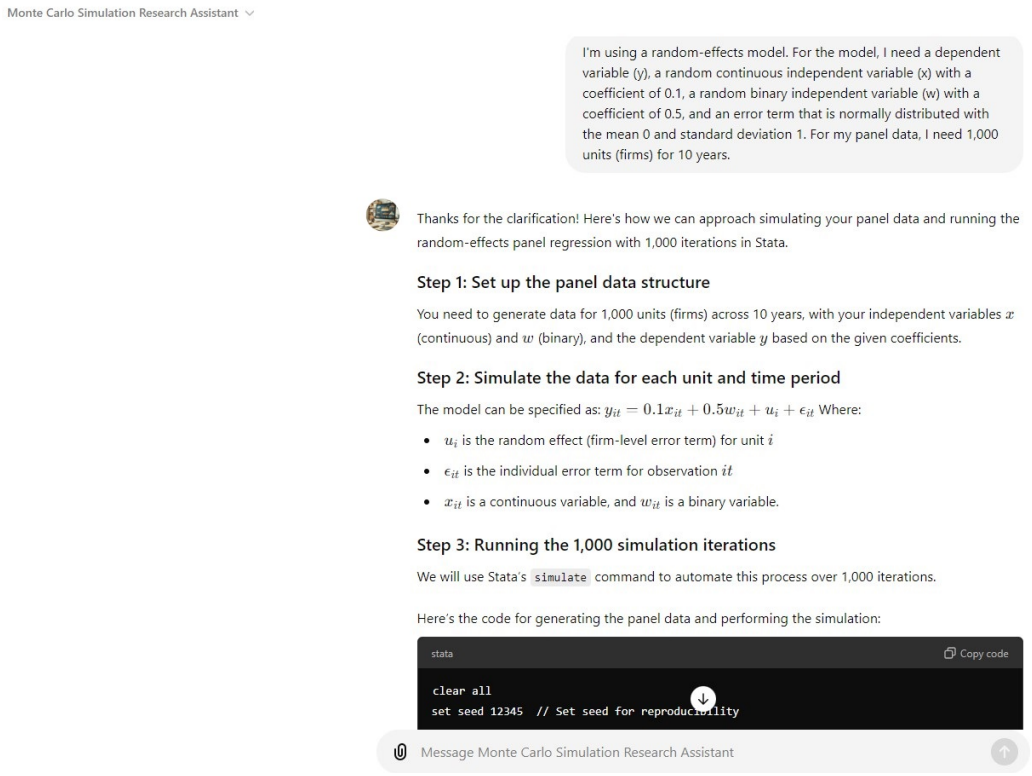


**Figure 3**. Screenshot of the custom GPT – Simulation steps

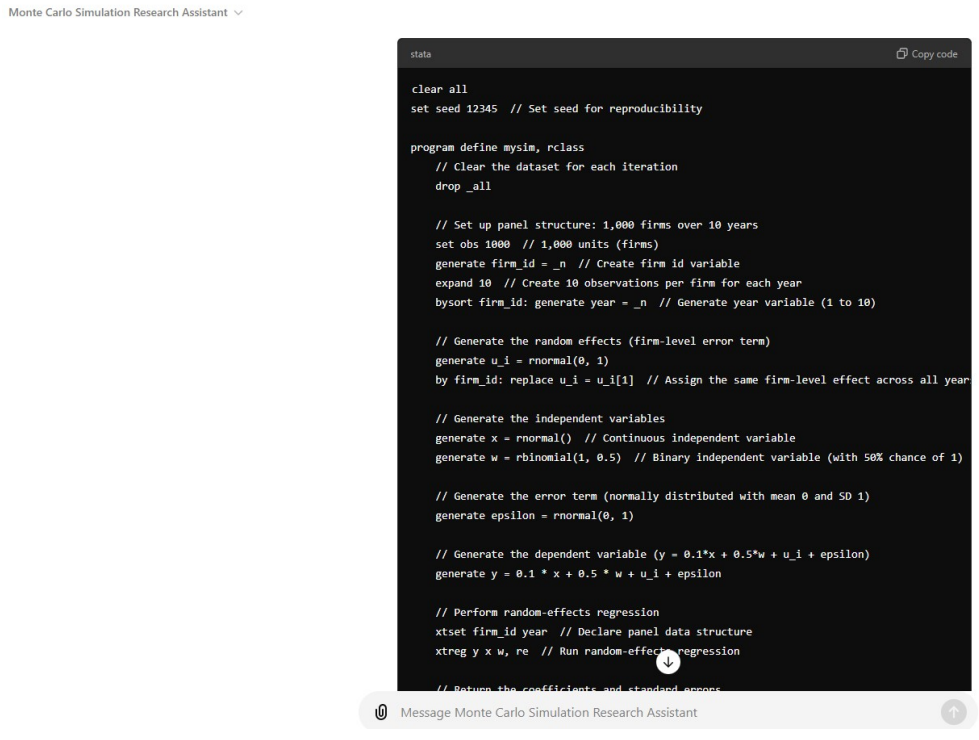


**Figure 4**. Screenshot of the custom GPT – Simulation code

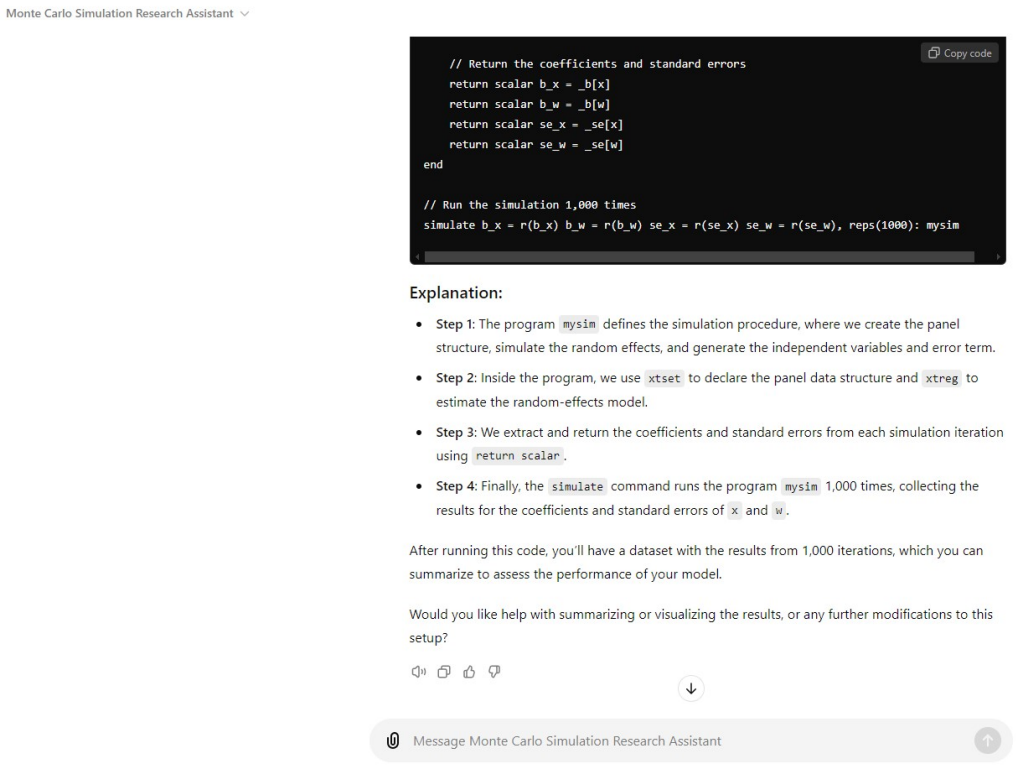


**Figure 5**. Screenshot of the custom GPT – Simulation code and explanation