Manpreet Gill, Shrihari Sridhar, & Rajdeep Grewal

Return on Engagement Initiatives: A Study of a Business-to-Business Mobile App

Firms are increasingly offering engagement initiatives to facilitate firm–customer interactions or interactions among customers, with the primary goal of fostering emotional and psychological bonds between customers and the firm. Unlike traditional marketing interventions, which are designed to prompt sales, assessing returns on engagement initiatives (RoEI) is more complex because sales are not the primary goal and, often, direct sales are not associated with such initiatives. To assess RoEI across varying institutional contexts, the authors propose and empirically implement a methodological framework to investigate a business-to-business mobile app that a tool manufacturer provides for free to engage its buyers. The data include sales by buyer firms that adopted the app over 15 months, as well as a control group of buyers that did not adopt. The results from a difference-in-differences specification, together with selection on observables and unobservables, show that the app increased the manufacturer's annual sales revenues by 19.11%–22.79%; even after accounting for development costs, it resulted in positive RoEI. This RoEI was higher when buyers created more projects using the app, so customer participation intensity appears to underlie RoEI. This article contributes to engagement literature by providing a methodological framework and empirical evidence on how the benefits of engagement initiatives materialize.

Keywords: causal effects, engagement initiatives, mobile applications, returns

Firms continually find new ways to interact with customers through myriad touch points in multiple channels and media (Lemon and Verhoef 2016). Often, the primary goal of these interactions is not to close a sale. For example, "the Sherwin-Williams' ColorSnap app allows users to capture desired colors on their mobile devices and then matches the colors to specific paint colors they can purchase at the paint store" (Urban and Sultan 2015, p. 33). In another setting, the milling, turning, tapping, and drilling calculators embedded in Sandvik Coromant's mobile apps[1] help engineers and machinists in buying firms perform machining and cost calculations and compare solutions for various parameters. During an annual fitness exposition, the protein supplement seller Nutrabolt helps runners and cross-training enthusiasts gauge their fitness levels through a CrossFit competition while also providing health and wellness tips.[2]

Two common themes underlie these organizational initiatives. First, they create customer value by providing product/service information, knowledge, and relevant assistance. For example, the Coromant app reduces the time that buyers spend selecting the right tools and machining processes. Second, firms use these initiatives primarily to foster interactions with customers, not to make sales. Sherwin-Williams does not obligate ColorSnap app users to buy its paint. Because these initiatives create value for customers but are not meant to prompt sales, we refer to them as "engagement initiatives." Consistent with a definition of engagement by Kumar and Pansari (2016), we define engagement initiatives as organizational initiatives that facilitate firm–customer interactions or interactions among customers, with the primary goal of fostering an emotional and psychological bond between customers and the firm.

Such initiatives have proliferated with the growth of the Internet and mobile devices, which intensify interactions between firms and customers (Manchanda, Packard, and Pattabhiramaiah 2015). Furthermore, with their complex products and services, long sales cycles, and varying interpurchase times, business-to-business (B2B) manufacturers frequently introduce mobile apps to interact with and engage buyers: one survey indicates that 80% of U.S. manufacturing companies have developed B2B apps (International Data Group

[1]Available at http://www.sandvik.coromant.com/en-gb/knowledge/calculators_and_software/apps_for_download (accessed October 2016).

[2]Available at http://www.fit360bcs.com/ (accessed November 2016).

Manpreet Gill is Assistant Professor of Marketing, Darla Moore School of Business, University of South Carolina (e-mail: msgill.usc@gmail.com). Shrihari Sridhar is Center for Executive Development Professor and Associate Professor of Marketing, Mays Business School, Texas A&M University (e-mail: ssridhar@mays.tamu.edu). Rajdeep Grewal is Townsend Family Distinguished Professor of Marketing, Kenan-Flagler Business School, University of North Carolina (e-mail: grewalr@unc.edu). The authors thank the Institute for the Study of Business Markets, Customer Analytics Program for providing access to the data. They thank Pranav Jindal, Gary Lilien, Vidya Mani, Chris Parker, Kapil Tuli, Lisa Scheer, and participants at the ISBM Customer Analytics Workshop 2015 for their helpful comments on previous drafts. Werner Reinartz served as area editor for this article.

ISSN: 0022-2429 (print)
1547-7185 (electronic)

Journal of Marketing
Vol. 81 (July 2017), 45–66
DOI: 10.1509/jm.16.0149

Market Report 2013).[3] These mobile apps can be effective engagement tools because they create multiple nonpurchase customer touchpoints (e.g., ColorSnap, Coromant), distinct from apps that function primarily as a sales channel (e.g., Amazon, Chipotle). Because B2B engagement apps are costly to develop and offered for free, and because B2B buying processes are not instantaneous or individual-specific, assessing returns to these apps is complex and difficult (e.g., Beebee 2013). Recent work has suggested that engagement initiatives might invoke nonsales outcomes such as trust, commitment, or loyalty (Brodie et al. 2011; Shiri, Beatty, and Morgan 2012), but scholars have called for explicit links to economic outcomes (Lemon and Verhoef 2016) to affirm the viability of such initiatives.

We accordingly propose a methodological framework for assessing economic returns on engagement initiatives (RoEI) and assess RoEI for a B2B mobile engagement app that a manufacturer provides for free to its buyers. We utilize the framework with novel data from a leading U.S. manufacturer (pseudonym: XYZ) that sells tools and industrial materials. XYZ devoted significant resources to launching its free manufacturing app, which can be downloaded on mobile devices and provides both product recommendations and process-design assistance (e.g., customized machining suggestions). This context is pertinent and typical of engagement initiatives for several reasons. First, the app is offered for free, creating direct economic (development and maintenance) costs, with no direct economic benefit. However, it might produce indirect economic benefits (e.g., increased revenues outside the app), which need to be quantified. Second, even if adopters generate indirect revenue benefits, to determine their value, we must safeguard against self-selection biases (i.e., revenue-generating buyers that adopt the app strategically) to avoid confounding the causal assessments of RoEI. Third, we aim to establish the RoEI mechanism to understand the source of the returns as well as develop managerial insights about which in-app interactions generate RoEI.

We use difference-in-differences specification and matching estimators to address these objectives. Specifically, we use objective sales data from a sample of buyer firms that downloaded the app (treatment group) and compared sales of these firms in the 15 months after the app's launch with sales in the 15 months before its launch. In turn, we utilize data from a random sample of buyers that did not download the app (control group), over the same time intervals. To avoid a self-selection bias related to buyers that adopt the app strategically, we estimated the treatment effects using different methods, reflecting distinct perspectives on how to obtain the focal difference-in-differences comparison: (1) selection on observables through regression and matching estimators and (2) selection on unobservables through a formal selection equation with appropriate instrumental variables.

We therefore offer two key contributions. First, we make theoretical and empirical contributions to engagement literature. Unlike most extant research that has focused on the definition or scope of customer engagement, we address the firm's economic benefits that result from an engagement initiative. We offer a methodological framework to address identification issues and provide a causal estimate of RoEI. Furthermore, in support of RoEI, we find that buyers that adopted the free app generated additional annual sales of 19.11%–22.79% for XYZ (relative to the preadoption period and benchmarked against nonadopters), even in the presence of alternate estimators, matching strategies, and data transformations. Because XYZ's RoEI is higher for buyers that create more projects using the app, our findings also indicate the importance of participation intensity in an engagement initiative as an RoEI-generating mechanism. Second, our empirical findings contribute to emerging literature on apps, which largely overlooks B2B apps and focuses primarily on intermediate outcomes, such as customer visits to a firm's mobile website or attitudes toward the firm (e.g., Urban and Sultan 2015; Xu et al. 2014). Our findings should prove useful to B2B firms trying to develop profitable ways to engage with their buyers through mobile apps.

We next discuss some relevant literature and present our conceptual arguments pertaining to B2B sellers' payoffs from free mobile apps. Then, we describe the institutional setting and data, model setup, and identification strategies. Finally, we present the results and discuss their implications.

Conceptual Background

Conceptualizing Engagement Initiatives

Engagement initiatives have two salient differences from traditional marketing-mix interventions. First, unlike traditional marketing interventions, engagement initiatives do not intend to induce a sale but primarily aim to build strong, long-term relationships with customers. Second, unlike conventional forms of one-way communication from the firm to the customer, engagement initiatives tend to be interactive and elicit participative experiences. The growth of engagement initiatives might stem from the growth of customer relationship management and its underlying philosophy that customers may interact valuably with the firm without necessarily making a purchase. These interactions need to be measured and managed to build stronger relationships, which then can lead to profitable value extraction. Engagement initiatives also grant firms their own touch points, which they can use to monitor and improve firm–customer and customer–customer interactions.

Components of an Engagement Initiative

A typical engagement initiative begins with customers' interactive participation with the firm. For example, customer participation in the Sherwin-Williams ColorSnap app is inherently interactive because customers search for their desired paint colors and interact with the app to choose a specific color. Buyers in the tooling industry first provide their machining parameters to the Sandvik machining calculator to obtain the desired machining processes and tolerance levels. Customers with greater participation intensity generally develop stronger emotional bonds and higher perceived interconnectedness with the firm, even when they do not engage in explicit purchasing activities (Van Doorn et al. 2010).

[3]App downloads increased from 4.5 billion in 2010 to 138.8 billion in 2014 (a 2,984.4% increase), with consumers spending an average of 2 hours and 19 minutes a day using mobile apps in 2014.

Heightened customer participation ideally leads to value creation for the customer, which constitutes the second component of the engagement initiative. Depending on their nature, engagement initiatives could provide value to customers at different purchase stages, and this value might be intrinsic, justifying the end unto itself, or extrinsic, by enabling a customer to perform a task related to the product or service (e.g., customization, designing a service solution) (Shiri, Beatty, and Morgan 2012). In a B2B context, increased interactions with the seller enable buyers to (1) articulate their business/product needs, (2) specify how they want the process customized, and (3) learn how to use the seller's products and solutions to fit their evolving needs (Sawhney 2006). The ubiquity and ease of access of mobile apps may be particularly valuable, in that they can alleviate the difficulty of acquiring static product information from a physical catalog as well as provide a platform that helps buyers create their own, customized solutions (Xu et al. 2014).

The last component of the engagement initiative is the firm's appropriation of value created for the customer. Engagement initiatives help firms strengthen their bonds with customers, and these bonds could lead to economic benefits in the future. A customer that learns about the firm through an engagement initiative may develop more favorable attitudes toward the firm, which should produce favorable economic outcomes. Increased perceived value fosters trust and loyalty, which also may increase purchase volumes (Reinartz and Ulaga 2008). Moreover, when buyers derive more value from participating in an engagement initiative, they might start to rely on seller-provided knowledge that otherwise would be costly or impossible to obtain, and the seller likely becomes a preferred supplier. Value extraction thus can follow from interactive participation through several routes, including increased purchasing behaviors, referrals, and influences on other customers (Kumar et al. 2013). Each behavioral outcome would signal the appropriation of value from an engagement initiative by providing clear pathways to incremental customer demand.

Assessing RoEI

Two challenges impede our ability to obtain causal estimates of RoEI. First, engagement initiatives lead to no direct economic benefits, and thus RoEI stems from indirect economic benefits, or net revenue increase generated from the value created by the engagement initiatives. Quantifying these indirect economic benefits is challenging; it requires a causal assessment of the impact of the initiative in the presence of multiple confounds, such as other environmental trends or marketing efforts, which co-occur with the engagement initiative.

Second, a firm's decision to offer the engagement initiative is strategic, as is the customer's decision to participate. That is, firms likely offer engagement initiatives to customers they believe will produce positive economic returns, and customers likely self-select into engagement initiatives according to their strategic evaluation of expected benefits. Buyers might adopt a tooling calculator app because they anticipate economies of scale and improved buying processes, for example, which would create a self-selection confound in assessing RoEI. Because information on all the reasons that firms use decide to launch an initiative or criteria that customers use to decide to participate is not observable to researchers, the omitted variables could lead to endogeneity in the RoEI estimates.

In Table 1, we summarize four potential approaches to assessing RoEI: event studies, seller-level observational inference designs, customer-level randomized experiments, and customer-level observational inference designs. We discuss each approach in turn next.

In an event study method, researchers could treat the announcement of a seller's engagement initiative as an economic event and estimate the impact of the event on the creation of the seller's shareholder wealth. Thus, data would be needed on the announcement dates of a sufficient sample of engagement initiatives across different sellers[4] over the study's time horizon (e.g., two years). Subsequently, researchers could assess the event's impact on the seller's shareholder value by obtaining a measure of the abnormal returns on the seller's stock price and testing the significance of the abnormal returns in an appropriate event window (for a discussion of the steps to define a market event and estimate abnormal returns, see Srinivasan and Bharadwaj [2004]). Estimates of RoEI using event studies constitute the market's belief about the potential economic value of an initiative. Although this method does not directly assess self-selection by sellers into such initiatives, the abnormal return attributable to the announcement would help adjust for the returns that stem from other variables causing price fluctuations in the market.

With seller-level observational inference designs, the primary objective is to establish the causal link between the presence of engagement initiatives and seller performance. Data are required on relevant aggregate seller-level outcomes (e.g., sales, firm value, profitability) from multiple sellers (ideally across several industries) during a time frame before and after each of the sellers launches an engagement initiative. Subsequently, researchers would estimate RoEI as the estimate of how much a seller's performance would change as a result of the introduction of the engagement initiative and would assess heterogeneity in RoEI estimates using both seller-type and industry-type moderators. However, the researchers would need to control for the notion that sellers likely possess private knowledge about whether, when, or how to launch a profitable engagement initiative. This private knowledge about the perceived efficacy of introducing an engagement initiative is unobserved to the researchers, is correlated to the eventual outcome of the engagement initiative, and could lead to endogeneity bias in RoEI estimates. To correct for this bias, we might rely on corrections such as instrumental variables, control functions, or precise knowledge about the institutional rules governing sellers' launch of engagement initiatives (for a similar discussion in the context of a firm's decision to have a chief marketing officer in the C-suite, see Germann, Ebbes, and Grewal [2015]).

The final two approaches, customer-level randomized experiments and customer-level observational inference design,

[4]In some cases, sellers could announce multiple engagement initiatives over the duration of the study, and these announcements would be included as separate events with appropriate statistical corrections.

TABLE 1
Comparison of Frameworks to Assess RoEI

Method	Aggregation	Data	Measure of RoEI	Solving Endogeneity from Selection
Event studies	Seller level	Data on the announcement dates of a large sample of engagement initiatives across different sellers	Abnormal market returns that a firm experiences on the day it announces an engagement initiative	Adjust for price fluctuations from the entire market on the same day
Seller-level observational inference designs	Seller level	Data on relevant aggregate seller-level outcomes (e.g., sales, firm value, profitability) from multiple sellers (ideally across several industries) during a time frame before and after each of the sellers launches an engagement initiative	Compare a relevant seller-level outcome (e.g., sales, firm value) across sellers, before and after the firm introduces engagement initiatives	To control for strategic selection by seller firms into engagement initiatives, use instrumental variables, control functions, and seller firms' decision rules
Customer-level randomized experiment	Single seller, customer level	Data from one seller on relevant aggregate economic outcomes (e.g., sales, margins, revenue) from multiple customers during a time frame before and after the seller's launch of the engagement initiative	Compare a relevant customer-level outcome (e.g., purchase order, quantity) across consumers, before and after the seller introduces an engagement initiative	Offer an engagement initiative to a random preselected treatment group but not to the control group
Customer-level observational inference design	Single seller, customer level	Data from one seller on relevant aggregate economic outcomes (e.g., sales, margins, revenue) from multiple customers during a time frame before and after the seller's launch of the engagement initiative	Compare a relevant consumer-level outcome (e.g., purchase order, quantity) across consumers, before and after the seller introduces an engagement initiative	Selection on observables, selection on unobservables

focus on one seller's engagement initiative and use across-customer variation to estimate RoEI. Thus, in each of these two cases, one would need data from one seller as well as relevant aggregate economic outcomes (e.g., sales, margins, revenue) from multiple customers during a time frame before and after the seller's launch of the engagement initiative.

In customer-level randomized experiment setup, researchers would infer the incremental economic benefit using customers' revealed purchase behaviors after the introduction of the engagement initiative. To prevent customer self-selection into the initiative, exposures to the engagement initiative would be randomized so that some preselected customers (treatment group) would have access, while other preselected customers (control group) would not. The random assignment implies that the difference in the average economic outcomes across treatment and control groups represents the treatment effect, or RoEI. To control for existing purchasing patterns in both groups, this method compares the change in economic outcomes (rather than levels) before and after the launch, across both groups. This robust version of RoEI would be the difference in the change in economic outcomes (difference-in-differences) across the treatment and control groups, after controlling for permanent differences across groups and time shocks common to both (for a discussion of the steps to infer causal economic effects of interventions using a randomized experimental design, see Athey and Imbens [2016]).

Finally, a customer-level observational inference designs is useful when, for pragmatic or fairness-related reasons, a seller cannot randomize the engagement initiative offering to its customers (as in our data). Thus, it might be possible that customers self-select into the engagement initiative, so RoEI estimates must modify the difference-in-differences estimate from the randomized design case, using empirical strategies that control for self-selection by customers (for a discussion in the context of online communities, see Manchanda, Packard, and Pattabhiramaiah [2015]). Because our data fall in this category, we subsequently describe three such strategies to overcome self-selection bias.

Engagement Initiatives in a Mobile App Context

Nascent but burgeoning literature on mobile apps (see Table 2) reflects the proliferation of mobile apps in the marketplace, with two main streams relevant for our research: demand for and

TABLE 2
Summary of Literature on Mobile Apps

Study	Focus Area	Data and Context	Key Finding
Bellman et al. (2011)	Mobile app effectiveness	Pre-/posttest experiment with general public from Australia (69) and United States (159); survey-based data.	Branded mobile phone apps increase attitude toward the brand and purchase intentions.
Xu et al. (2014)	Mobile app effectiveness	Repeated cross-sectional data (Q4 2009 and Q2 2010) from comScore MobiLens on 5,600 smartphone users; survey-based data.	Adoption of the news app significantly increases the probability of visiting the mobile website.
Einav et al. (2014)	Mobile app effectiveness	Mobile and nonmobile activity (including purchases) of users of eBay's shopping app and website.	Adoption of eBay's mobile application increased total platform purchases.
Urban and Sultan (2015)	Mobile app effectiveness	App to assist users who intend to move and "dream mover" app to help users purchase or rent new homes; survey data.	Benevolent apps increase app users' trust, brand consideration, and purchase likelihood.
Carare (2012)	Factors affecting mobile app demand	Daily download rankings for 166 days of top 100 paid and free apps in the United States available through Apple's app store.	Apps' past sales ranks affect their current sales.
Garg and Telang (2013)	Factors affecting mobile app demand	Daily app ranking and pricing data for two months, on 200 paid and 200 free apps for iPad and iPhone.	Method to estimate apps' demand using publicly available data on apps' ranks and prices.
Ghose and Han (2014)	Factors affecting mobile app demand	App characteristics and daily panel data on 4,706 iOS-based and 2,624 Android-based smartphone apps' sales for a period of four months.	App demand increases with the app version, app age, number of apps developed, positive user reviews, number of platforms on which app is released, and the presence of in-app purchase option.
Han, Park, and Oh (2016)	Factors affecting mobile app demand	Individual-level weekly data on usage of Android mobile apps.	Consumer utility from app usage varies by product category and consumers' demographic characteristics.
Kwon et al. (2016)	Factors affecting mobile app demand	Individual-level weekly panel data on Facebook and Anipang apps.	Consumers are rationally addicted to social and gaming apps.

effectiveness of mobile apps. In particular, research into app demand denotes the influences of customer characteristics, such as age, gender, and education (Han, Park, and Oh 2016). Younger users exhibit more affinity toward social networking, gaming, and photo apps relative to seniors; women express more preference for communication and entertainment apps than men. Younger customers also have low satiation for social networking and gaming apps, such that their usage appears to mimic uses of habit-forming substances such as alcohol (e.g., Kwon et al. 2016). Garg and Telang (2013) find that app demand also increases because of app characteristics such as its age, platform, version, and rank. According to Ghose and Han (2014), in-app advertisements negatively influence demand, whereas Carare (2012) shows that app demand increases with the valence and volume of customer reviews. An implicit assumption in this stream is that higher app demand is better for the app developer, but the return on investment remains unexplored.

The second research stream relates to the effectiveness of mobile apps. Firms use mobile apps to engage customers and obtain a competitive edge; for example, branded mobile apps (e.g., eBay, Amazon) allow for continuous interactions with customers and thereby strengthen customer attitudes and purchase intentions (Bellman et al. 2011). Urban and Sultan (2015) argue that engagement apps foster fondness over repeated customer usage, which could strengthen mindset metrics such as brand attitudes, brand consideration, and purchase intentions. According to Xu et al. (2014), customer adoption of news apps increases their probability of visiting the newspaper's mobile website. Sales apps also can increase overall sales for an omnichannel retailer (Einav et al. 2014). A consistent argument thus holds that firms may use apps to engage their customer base and thereby create several important nonsales outcomes, such as trust, commitment, and loyalty (Brodie et al. 2011; Shiri, Beatty, and Morgan 2012). However, we find no explicit approaches for estimating RoEI.

Method

Data

We obtained data from a leading manufacturer (XYZ) of tooling and industrial materials (e.g., automatic lathes, cutting tools) on its free mobile app[5] launched in September 2013. XYZ's annual sales are more than $1 billion, with buyers on six continents. Like other manufacturers in this industry, XYZ provides detailed print and online catalogs to buyers, specifying appropriate tools for performing simple machining jobs as well as machining sequences (and assembly layouts) required to execute more complex processes. Buyers often find it time consuming to review print catalogs, and generational turnover and a general lack of interest in manufacturing jobs among younger workers led XYZ to expect this knowledge gap to widen. Furthermore, XYZ believed that younger buying managers might be comfortable using Internet-enabled technologies in design environments. Accordingly, the firm chose to digitize its existing product information in a mobile app, which would reduce the time required for buyers to identify optimal tools and create machining sequences while also creating a touch point between the buyer and XYZ's offerings.

Buying firm managers typically use the app at their own machining plants for either product search or more complex product assembly designs. As a product search enabler, the app collects information from the buyer, such as a specification of the focal machining operation, then returns optimal tool recommendations. Buyers can select tools across a range of customizable attributes built into the app; they also can bookmark product recommendations and share their search results with other buying managers. As a product assembly platform, the app allows a buyer to draw an entire manufacturing process, comprising a series of tooling operations, together with the specific tools and tolerance levels associated with that process. The app reviews the overall manufacturing process and provides usage recommendations to make the process more efficient (e.g., better tolerance levels with the current tools) or product improvements for the same operation (e.g., to reduce manufacturing time).

Because XYZ provided the app for free and it was not designed to stimulate direct sales, XYZ employed virtually no targeted marketing efforts to increase adoption, except for an e-mail to all buyers around the time of app launch, followed by a short press release highlighting its features. The sales force also operated independently of the app, and the app remained solely under the purview of the new product development and marketing functions. However, XYZ believed that existing buyers might purchase more tools because of this engagement initiative. When a manager from the buying firm downloads the app, the user must provide the buying firm's name and a unique sales identifier that XYZ provides each buying firm following its first transaction. Multiple managers within the same buying firm can use the app, but the unique sales identifier consistently refers to the buying firm level. We obtained data from 550 unique buyer firms that downloaded XYZ's app; however, for confidentiality reasons, we cannot disclose the total number of adopters.

We obtained all (offline) transaction sales data for these buyers for a 15-month period from September 2013 (launch month) to November 2014, as well for a 15-month period preceding the launch. Thus, we create a two-period customer-level observational inference design with a control and treatment group. We aggregate each set of 15 months of data for two reasons. First, interviews with the app director and sales managers at XYZ indicated that buying firms' purchase cycles are generally long (5 months on average) but also vary significantly (2–12 months). Disaggregation thus could lead to misrepresentations of sales changes due to organic differences in the purchase cycles across the buying firms. Second, we did not observe the exact date when buyers downloaded the app. Some might have done so early in the postlaunch period, whereas others did so later. By treating the entire 15-month period as the postlaunch period, we assume that buyers who downloaded the app did so right after its launch, which offers a conservative assessment because it limits the treatment effect for later adopters, which effectively must start at the moment of the app launch to indicate business benefits to XYZ. Manchanda, Packard, and Pattabhiramaiah (2015) use a similar conservative assumption; Bertrand, Duflo, and Mullainathan (2004) also recommend such an aggregation to help mitigate potential issues related to serial correlation and grouped error term effects.

Next, we obtained transactional sales data from a randomly drawn sample of 700 unique buyers that did not download the app. For the comparison, we consider a subsample of buyers that purchased at least once in both pre- and postlaunch periods, to mitigate potential endogenous entry or exit effects. We thus have data from 522 buyers that downloaded the app and 626 buyers that did not.

Identification Strategy

Our goal is to assess if XYZ's introduction of the free app increases sales revenue from buyers that adopted the app. In an experimental sense, XYZ exposes app-adopting buyers to a treatment, and we aim to infer the treatment effect, as represented by the incremental sales revenue from these buyers resulting from their adoption of the app. In an ideal setting, we could randomize the treatment, then observe sales from buyers that did not get the app (S_0) and sales from buyers that obtained it (S_1). With such a random assignment, the difference in these average sales, or $\overline{S_1} - \overline{S_0}$, represents the treatment effect—that is, the incremental economic benefit of introducing the app. However, for fairness, XYZ's app was available to all buyers. Thus, in our data (as in most observational data settings), buyers' app adoption is not random, and we need to account for buyers self-selecting into the treatment group. Not all their adoption reasons are observable; for example, we cannot observe improvements in the buying process that result from app adoption. Omitted variables that drive strategic app adoption could correlate with the sales XYZ earns from these buyers, which would involve an endogeneity bias. Therefore, we consider three potential solutions that vary in the extent to which they correct for selection bias to establish the causal

[5]We refer to smartphones and tablets when we use the use the term "mobile device."

link between app adoption and sales: (1) difference-in-differences, (2) difference-in-differences, augmented with selection on observables, and (3) difference-in-differences, augmented with selection on unobservables.

Difference-in-differences. The difference-in-differences approach compares the sales differential (posttreatment sales − pretreatment sales) of buyers in the treatment group with buyers in the control group. Thus,

$$S_{ijt} = \beta_0 + \beta_1 I_j + \beta_2 I_t + \beta_3 I_j \times I_t + \varepsilon_{ijt}, \tag{1}$$

where S_{ijt} is buyer i's sales from group j at time t, and ε_{ijt} is a random error term, clustered across buyers and the two periods. Our data set contains two groups j (treatment and control) and two time periods t (pre- and postlaunch periods). Then the indicator variable I_j picks up mean differences in the sales between the treatment group and the control group, referred to as group fixed effects and indicated by the coefficient β_1. The indicator variable I_t indicates the mean differences in post-relative to prelaunch period sales, similar to time fixed effects and indicated by the coefficient β_2. Finally, β_3 captures the difference in the change in sales outcomes (difference-in-differences) across the treatment and control groups, after controlling for permanent differences across groups and the time shocks common to both groups. Thus, β_3 is the estimate of the treatment effect, given as

$$\beta_3 = \left[E\left(S_{ijt} \middle| j = 1, t = 1\right) - E\left(S_{ijt} \middle| j = 1, t = 0\right)\right] - \left[E\left(S_{ijt} \middle| j = 0, t = 1\right) - E\left(S_{ijt} \middle| j = 0, t = 0\right)\right]. \tag{2}$$

From Equation 2, β_3 can also be viewed as the incremental economic benefit to XYZ of introducing the app, or RoEI. A key identifying assumption of the difference-in-differences approach is that the treatment and control groups are identical, so the time trends in sales for the treatment and the control group buyers are also identical (parallel trends assumption), apart from the treatment itself. Using this assumption, the deviation in the difference in sales for the treatment group from that of the control group provides a causal estimate of the treatment effect. Group fixed effects also eliminate time-invariant, buyer-specific unobservable variables—and, thus, self-selection—to the extent that this bias is driven by group-specific, time-invariant omitted variables.

However, the critical parallel trends assumption could be violated in our study context because buyer-specific unobservable variables (which influence both buyers' adoption decisions and sales) could vary across buyers, resulting in heterogeneous, dissimilar groups. The group fixed effects, meant to smooth out the permanent differences between groups, then would not eliminate buyer unobservable variables that are distinct from the group-specific, time-invariant unobservable variables. Failing to account for them in the difference-in-differences analysis could make our control group an inappropriate counterfactual for the treatment group because of the violation of the parallel time trends assumption. Thus, we augment the difference-in-differences analysis.

Difference-in-differences with selection on observables. Compositional differences between the control and treatment groups (thus violating the parallel trends assumption) arise because buyer firms self-select into the app due to unobservable variables that also correlate with their sales. For example, the app may offer more cost savings for some buyers, which could affect their unobserved preference for XYZ's offerings more in the treatment group than in the control group. Selection on observables corrects for this self-selection by assuming that the researcher observes all variables that buyers consider while deciding to adopt.

In our study setting, buyers' motivations to adopt the app may be due to cost-related advantages. A selection-on-observables strategy uses buyer-specific observables to proxy for cost advantages, such that the treatment and control groups look similar and the parallel trends assumption is preserved. Then, the outcome (i.e., sales) is independent of the treatment (i.e., app adoption); formally,

$$S_i \perp T_i | Z_i, \tag{3}$$

where S is sales, T is an indicator of app adoption, Z indicates the observables, and $\perp$ is an orthogonality operator.

We operationalize this approach by augmenting our difference-in-differences model from Equation 1 with all the observed buyer firm variables (e.g., Angrist and Pischke 2009) as follows:

$$S_{ijt} = \beta_0 + \beta_1 I_j + \beta_2 I_t + \beta_3 I_j \times I_t + \boldsymbol{\beta}_4 \mathbf{Z_{ij}} + \varepsilon_{ijt}, \tag{4}$$

where the added vector $\mathbf{Z_{ij}}$ captures the set of observables, the effects of which are estimated through the coefficient vector $\boldsymbol{\beta}_4$. The treatment effect thus is given as

$$\beta_3 = \left[E\left(S_{ijt} | j = 1, t = 1, \mathbf{Z_{ij}}\right) - E\left(S_{ijt} | j = 1, t = 0, \mathbf{Z_{ij}}\right)\right] - \left[E\left(S_{ijt} | j = 0, t = 1, \mathbf{Z_{ij}}\right) - E\left(S_{ijt} | j = 0, t = 0, \mathbf{Z_{ij}}\right)\right]. \tag{5}$$

Difference-in-differences with selection on unobservables. The assumption that we can observe all the important variables is a strong one, so we also need to account for unobservable variables. We combine the difference-in-differences analysis with a formal Heckman-style selection model, in which the errors in the selection equation (required to model the buyer's decision to adopt) and the errors in the outcome equation (i.e., difference-in-differences model) correlate and follow a bivariate normal distribution. In turn, we can derive the inverse Mills ratio (IMR) to account for unobservable variables in the outcome equation (Heckman 1979). Adding this ratio to the outcome equation accounts for omitted unobservable variables, so this strategy is called selection on unobservables (see Appendix A).

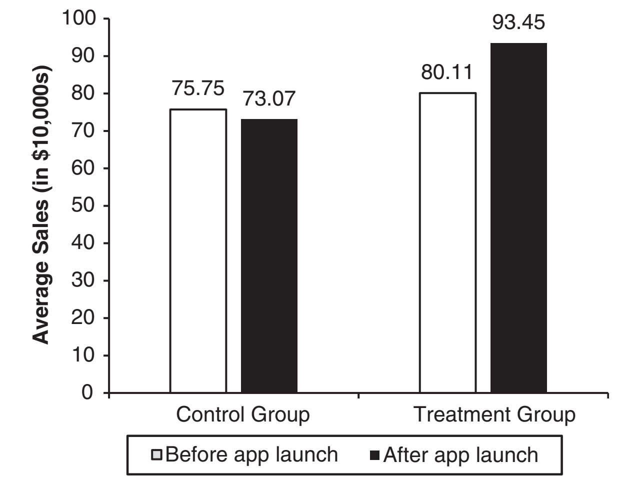
We first model buyers' decision to adopt the app as a function of all the observable variables with a probit model, which we use to calculate the IMR for the buyer firms in the treatment and control groups. Then, we augment our difference-in-differences model in Equation 4 as follows:

$$S_{ijt} = \beta_0 + \beta_1 I_j + \beta_2 I_t + \beta_3 I_j \times I_t + \boldsymbol{\beta}_4 \mathbf{Z_{ij}} + \beta_5 \mathrm{IMR}_{ij} + \varepsilon_{ijt}, \tag{6}$$

The treatment effect thus becomes

$$\begin{aligned} \beta_3 = &\left[E\left(S_{ijt} | j = 1, t = 1, \mathbf{Z_{ij}}, \mathbf{IMR_{ij}}\right) - E\left(S_{ijt} | j = 1, t = 0, \mathbf{Z_{ij}}, \mathbf{IMR_{ij}}\right)\right] \\ &- \left[E\left(S_{ijt} | j = 0, t = 1, \mathbf{Z_{ij}}, \mathbf{IMR_{ij}}\right) - E\left(S_{ijt} | j = 0, t = 0, \mathbf{Z_{ij}}, \mathbf{IMR_{ij}}\right)\right]. \end{aligned} \tag{7}$$

FIGURE 1
Model-Free Evidence



Results

Model-Free Evidence

As we show in Figure 1 and Table 3, raw mean total sales (scaled in $10,000s) in the control and treatment groups were not statistically different in the prelaunch period (treatment = 80.11, control = 75.75, n.s.). Treatment group sales were higher in the postlaunch than in the prelaunch period (post = 93.44, pre = 80.11, $p < .05$), whereas the sales in the control group stayed approximately the same in both periods (post = 73.07, pre = 75.75, n.s.). Sales also remained about the same for buyers in the control group, but they increased for buyers in the treatment group, indicating the need for a more formal comparison.

Selection of Covariates

From detailed interviews with the app program director, the app marketing team, and the app developer at XYZ, we learned that buyers' strategic motivation to adopt the app was cost savings. Specifically, app adoption can reduce the time buyers expend on product searches and help streamline their entire assembly design process because the app provides a common platform for all buying units to create machine assemblies. Thus, we include a set of buyer-specific observable variables that proxy for buyers' strategic motives to reduce costs by adopting the app, described in the following subsections.

Buyer power. We measured buyer power as the ratio of the buyer's total sales in the prelaunch period T_1 to the sum of total sales by XYZ to all buyers in the same industry division. Buyers that transact often with XYZ likely would enjoy cost advantages by adopting the app because of the efficiency gains of using a single app to design all offerings. Moreover, buyers that transact more often with a seller tend to value relationship-specific investments by the seller because they observe these investments during every transaction.

Buyer's industry competitiveness. We measured competitiveness in the buyer's industry by obtaining the buyer's industry concentration ratio from U.S. Census reports, which reflects the ratio of the sales of the top 20 firms in an industry to total sales in the industry.[6] We subtracted the concentration ratio from 1 to measure competitiveness in the buyer's industry (Lee et al. 2015). Buyer industry competitiveness ranges from 0 to 1, where 0 refers to highly monopolistic industries and 1 implies highly competitive industries. Prior B2B technology adoption literature has shown that greater competitive intensity induces significant heterogeneity in firms' new technology adoption speed (Lee and Grewal 2004). Firms anticipate cost-related gains from adopting early (and being first movers) because they can limit the losses that might accrue from unsubstantiated early adoptions. Lee and Grewal (2004) show that as competition increases, heterogeneity in adoption (vs. nonadoption) increases; firms decide quickly whether to move early or not adopt at all (and wait for the benefits to trickle down), so it becomes crucial to control for competitive intensity.

Buyer firm size. We measured buyer size using the number of employees in the buying firm. Larger buyer firms might have more cost savings from app adoption than smaller firms because of scale advantages of reduced product search time and product assembly guidance.

Buyer T_0 period patterns. We controlled for buyers' intrinsic preference in transacting with XYZ by including buyers' past sales (buyer T_0 sales) and purchase frequency (buyer T_0 frequency) in the period five months before the preperiod T_1, which we denote as T_0. The T_0 period provides the baseline reference period for the study.

Other cultural and industry factors. We controlled for the location (continent) and the industry classification code (using the manufacturer's internal industry classification code) of the buyer firm, which might induce heterogeneity in buyers' perceptions of the cost savings achieved from using the app. In Table 4, we show that the composition of buyers in the control and treatment groups is similar. Nearly half the buyers are located in developing economies (Asia, South America, Africa), and the z-statistic shows that the composition between groups is statistically indistinguishable. The composition of buyers across transportation and aerospace, heavy equipment, and general engineering industries across treatment and control

TABLE 3
Mean Differences (Total Sales in $10,000) Between Control and Treatment Groups

	Control Group	Treatment Group	Difference
T_1: 15 months pretreatment	75.75	80.11	4.36
T_2: 15 months posttreatment	73.07	93.44	20.37**
Number of observations	626	522	

**$p < .05$.

[6]The U.S. Census reports are available at https://www.census.gov/econ/concentration.html.

TABLE 4
Mean Differences (Covariates) Between Control and Treatment Groups at T_1

Percentage of Observations from Emerging and Emerged Economies			
Economy	**Control Group**	**Treatment Group**	**z-Stat**
Developing economies	.13	.16	1.58
Developed economies	.87	.84	1.58
Percentage of Observations from Different Industry Divisions			
Industry	**Control Group**	**Treatment Group**	**z-Stat**
Transportation and Aerospace	.30	.28	.81
Heavy Equipment	.10	.12	1.27
General Engineering	.60	.60	.05
Firm Variables	**Control Group**	**Treatment Group**	**t-Stat**
Buyer firm size	4.51	5.48	.70
Buyer industry competitiveness	55.57	53.80	1.59
Buyer power	60.05	81.23	1.18
Buyer T_0 sales	18.59	20.11	–.40
Buyer T_0 purchase frequency	50.95	97.20	–5.91***
Number of observations	626	522	

***$p < .01$.
Notes: Buyer firm size was scaled down by a factor of 100, buyer industry competitiveness and buyer power (originally measured in %) were scaled up by a factor of 100. Buyer T_0 sales (scaled down by a factor of 10,000) represents the total sales of the firms in the five months before the start of the data window. Buyer T_0 purchase frequency represents the total purchase frequency of the firms in the five months before the start of the data window. We use this scaling throughout the article, except as explicitly noted.

groups also is statistically similar. In Table 4, we also report the mean values of the buyer power, buyer firm size, buyer industry competitiveness, and buyers' intrinsic preference to transact with the manufacturer (captured as T_0 sales and purchase frequency) across treatment and control groups. Buyers' average transaction share, firm size, industry competitiveness, and T_0 sales do not differ statistically across groups. However, buyers' T_0 purchase frequency in the treatment group is significantly higher than that of firms in the control group.[7]

Model-Based Results

Difference-in-differences. We begin by presenting the estimates for Model 1 (Table 5), without the control group. For buyers in the treatment group, sales in the 15-month postlaunch period were higher than sales in the 15-month prelaunch period by \$133,300, or an annual increase of \$106,640. The average annual sales of a buyer in the treatment group (\$640,880) thus reveals a sales increase of 16.64% as a result of adoption of the app.

[7]The significant difference in the T_0 purchase frequency between the treatment and control groups suggests a covariate imbalance, which violates the parallel trends assumption in difference-in-differences analysis. Accordingly, the use of a selection-on-observables strategy is warranted. We implement this strategy by (1) including all the covariates (including T_0 purchase frequency) in the difference-in-differences model and (2) matching firms in the treated group to firms in the control group as a function of the covariates. After matching (i.e., nearest neighbor), the difference in T_0 purchase frequency between the treated and control firms becomes statistically insignificant. We discuss the treatment effect estimates obtained by using nearest neighbor matching in Appendix B.

Next, we added the control group and estimated the treatment effect from the difference-in-differences specification (β_3) without any buyer-specific characteristics (Table 5, Model 2). The treatment effect was significant ($\beta_3 = 16.01$, $p < .05$), indicating a statistically significant economic impact for XYZ when the buyer adopts the app. The average annual sales of a buyer in the treatment group in the prelaunch period (\$640,880) enabled us to calculate a percentage sales increase of 19.99% in the postlaunch period.

Difference-in-differences with selection on observables. We augment the simple difference-in-differences model with buyer-specific characteristics in Model 3 in Table 5. Again, the treatment effect was significant ($\beta_3 = 16.01$, $p < .05$), indicating a statistically significant economic impact of buyers' adoption. According to the average annual sales of a buyer in the treatment group (\$640,880), the annual sales increase was 19.99%.

Difference-in-differences with selection on unobservables. To correct for buyers' potential self-selection into adoption, we used a two-stage Heckman (1979) correction. In the first stage, we model the app adoption choice using key drivers and a probit specification. Buyer power, industry competitiveness, firm size, T_0 period patterns (i.e., T_0 sales and purchase frequency), and cultural and industry factors can all proxy for buyers' strategic motives to reduce costs by adopting, so we included these covariates as predictors of app adoption in the first-stage model.

For identification, the covariate set driving the app adoption choice should contain at least one variable that provides an exclusion restriction, such that it affects app adoption but does not directly influence buyer sales. We used the number of buyer firm buying units; as the number of buying units increases, the chances that a buyer has its

TABLE 5
Treatment Effect Estimation Results

Variables	Model 1 Only Treatment	Model 2 Without Covariates	Model 3 With Covariates	Model 4 Heckman Model (No Covariates)	Model 5a Heckman Model (with Covariates)	Model 5b Heckman Model	Model 4a Selection Model	Model 4b Selection Model
Treatment effect		16.01**	16.01**	16.01**	16.01**	16.01**		
		(6.498)	(6.523)	(6.500)	(6.524)	(6.524)		
Time dummy	13.33***	−2.682	−2.682	−2.682	−2.682	−2.682		
	(4.058)	(5.103)	(5.122)	(5.104)	(5.123)	(5.123)		
Treatment group dummy		4.362	−20.33***	209.1***	−16.05	−8.841		
		(9.864)	(6.542)	(37.98)	(33.85)	(15.85)		
Buyer firm size	.0210		.597***		.597***	.596***	.769	2.578
	(.143)		(.217)		(.217)	(.215)	(1.822)	(1.898)
Buyer power	.0821***		.224***		.224***	.225***	−.246	−.249
	(.0226)		(.0667)		(.0673)	(.0670)	(.251)	(.344)
Buyer industry competitiveness	.184		−.290		−.280	−.265	−.411*	−.243
	(.157)		(.184)		(.189)	(.180)	(.233)	(.241)
Buyer T_0 sales	.310***		.0533		.535	.538	−.135**	−1.144**
	(.0352)		(.0383)		(.384)	(.384)	(.0615)	(.504)
Buyer T_0 purchase frequency	.209***		.435***		.431***	.425***	.00257***	.00292***
	(.0481)		(.0489)		(.0595)	(.0517)	(.000454)	(.000468)
IMR				−140.6***	−2.794	−8.407		
				(24.98)	(22.04)	(11.79)		
Number of buying units							−.389***	−.386***
							(.0538)	(.0536)
Firms that adopted the app								2.482***
								(.229)
Constant	32.48*	75.75***	76.93***	−16.42	74.45**	70.29***	.206	−1.015**
	(19.20)	(7.537)	(25.05)	(17.79)	(31.39)	(25.77)	(.386)	(.400)
Observations	1,044	2,296	2,296	2,296	2,296	2,296	1,148	1,148
Adjusted R-square	.689	.001	.461	.044	.461	.461		
Pseudo R-square							.08	.18
Division and continent fixed effects	Yes	No	Yes	No	Yes	Yes	Yes	Yes

*$p < .10$.
**$p < .05$.
***$p < .01$.

Notes: Robust standard errors are in parentheses. Dependent variable (sales) is in $10,000. In the selection model, buyer firm size was scaled down by a factor of 100,000; number of buying units was scaled down by a factor of 100; buyer T_0 sales was scaled down by a factor of 10,00,000; buyer power (originally measured in %) was scaled down by a factor of 10. Unlike Models 2–5, which consist of 2,296 observations (1,148 buyer firms over two time periods), the selection Model 4a has 1,148 observations, consisting of 522 treated buyer firms and 626 control buyer firms. “Firms that adopted the app” is the second instrument that captures the number of buyer firms, other than the focal firm, that adopted the app in the industry.

TABLE 6
Robustness Assessment

Variables	Outliers: In (Total Sales)	Competition Intensity Ratio 50	Competition Intensity Ratio 8	Competition Intensity Ratio 4
Treatment effect	.151**	16.01**	16.01**	16.01**
	(.0741)	(6.523)	(6.523)	(6.523)
Time dummy	.000761	−2.682	−2.682	−2.682
	(.0572)	(5.122)	(5.122)	(5.122)
Treatment group dummy	.699***	−20.87***	−19.77***	−19.59***
	(.125)	(6.580)	(6.502)	(6.485)
Buyer firm size	.826***	.594***	.612***	.623***
	(.205)	(.215)	(.220)	(.221)
Buyer power	.200***	.224***	.224***	.224***
	(.0678)	(.0667)	(.0667)	(.0667)
Buyer industry competitiveness	−.779**			
	(.361)			
Buyer T_0 sales	.000297	.0532	.0534	.0534
	(.000324)	(.0382)	(.0383)	(.0383)
Buyer T_0 purchase frequency	.00682***	.437***	.432***	.431***
	(.000719)	(.0490)	(.0488)	(.0488)
Competition intensity_50		−.303*		
		(.169)		
Competition intensity_8			−.199	
			(.195)	
Competition intensity_4				−.129
				(.220)
Constant	12.62***	74.31***	75.35***	72.85***
	(.539)	(24.48)	(25.97)	(27.80)
Observations	2,296	2,296	2,296	2,296
Adjusted R-square	.293	.461	.460	.460
Division fixed effects	Yes	Yes	Yes	Yes
Continent fixed effects	Yes	Yes	Yes	Yes

*$p < .10$.
**$p < .05$.
***$p < .01$.
Notes: Robust standard errors are in parentheses. In the model used to assess robustness against outliers, firm size was scaled down by 10,000, and power is in %.

own centralized product search or assembly unit should increase too, so its reliance on a manufacturer to provide this service is lessened, and more buying units should decrease the probability of app adoption. However, there is no reason to expect a priori that the number of buying units exerts any effect on the change in total buyer sales. The results of the first-stage (probit) model in Table 5 (Model 4a) that predicts app adoption according to buyer characteristics confirms that the number of buying units decreases the probability of adoption ($b = -.389$, $p < .05$). Then, we added the IMR as a selection correction term in the second-stage sales equation.

The main results in Table 5, Model 4 (no covariates), reveal that the selection correction term is significant, and the treatment effect remained statistically significant ($\beta_3 = 16.01$, $p < .05$). Model 5a (all covariates) confirms these results ($\beta_3 = 16.01$, $p < .05$), except that the selection correction term is statistically nonsignificant. Thus, the selection-on-unobservables strategy indicates a positive economic benefit to XYZ: the treatment effect increase translated into a 19.99% annual sales increase.

For robustness, we considered another instrument: the number of buyer firms in the focal firm's industry that have adopted the app. This instrument passes the validity criterion, because the number of buyer firms that have adopted the app should correlate positively with the decision of a focal firm to adopt. However, there is no reason that peer firms' adoption decisions should correlate with the focal firm's sales, conditional on industry competitiveness and time fixed effects. We modeled a buyer firm's app adoption as a function of various covariates and two instruments (i.e., number of buying units and the number of buyer firms in the focal firm's industry that have adopted the app), using a probit model. Accordingly, we obtained unobserved factors capable of influencing the buyer firms to adopt and their sales in the IMR, which we included in the outcome model with the other covariates. Our results further bolstered our claim regarding the treatment effect, which again turned out to be statistically significant ($\beta_3 = 16.01$, $p < .05$)[8] and implied a 19.99% sales increase for the buyer firms that adopted the app (see Models 5b and 4b, Table 5).

[8]Although the treatment effects estimates are similar across Models 2–5, their standard errors (reported in Table 5) are different. We conjecture that the similarity of the treatment effects could stem from the stability of the identification strategies.

TABLE 7
Novelty and Falsification Tests

Variables	Novelty	Falsification
Treatment effect	14.42***	–3.332
	(5.402)	(4.181)
Time dummy	–3.974	3.516
	(4.170)	(3.772)
Treatment group dummy	–16.89***	–8.071**
	(5.897)	(3.547)
Buyer firm size	.539***	.332***
	(.163)	(.101)
Buyer power	.177***	.122***
	(.0507)	(.0389)
Buyer industry competitiveness	–.226	–.143
	(.165)	(.0968)
Buyer T_0 sales	.0248	.0270
	(.0333)	(.0180)
Buyer T_0 purchase frequency	.363***	.198***
	(.0427)	(.0235)
Constant	66.86***	24.12**
	(22.41)	(9.630)
Observations	2,255	2,178
Adjusted R-square	.396	.437
Division fixed effects	Yes	Yes
Continent fixed effects	Yes	Yes

$^{**}p < .05$.
$^{***}p < .01$.
Notes: Robust standard errors are in parentheses.

Robustness Analyses

Potential outcomes framework. For our selection-on-observables strategy, we used a classic regression estimator to estimate the treatment effect. Specifically, after adding the set of observables $\mathbf{Z_{ij}}$, the treatment effect $[E(S_{ijt}|\, j = 1,\ t = 1, \mathbf{Z_{ij}}) - E(S_{ijt}|\, j = 1,\ t = 0, \mathbf{Z_{ij}})] - [E(S_{ijt}|\, j = 0,\ t = 1, \mathbf{Z_{ij}}) - E(S_{ijt}|\, j = 0, t = 0, \mathbf{Z_{ij}})]$ can be estimated directly from observations for which j and t are 0 and 1, respectively, using a conditional independence assumption. Thus, the regression-based approach conditions on the covariates to compare all members of the control and treatment groups. As an alternative selection strategy, we use a potential outcomes framework (Guo and Fraser 2010). The change in the buyer's outcomes when it adopts versus does not adopt the app (denoted as ∇S_{i0} and ∇S_{i1}, respectively) are potential outcomes. Their difference represents the firm-level treatment effect, averaged over the sample of firms to yield the average treatment effect (ATE). However, it is not possible to estimate ATE; we observe only one potential outcome for each buyer. Instead, we use various methods to impute missing potential outcomes and then calculate the ATE as an average of individual treatment effects in the sample. Specifically, we relied on nearest-neighbor matching and its variants (i.e., regression adjustment, inverse probability weighting, and inverse probability weighting with regression adjustment). We provide the estimates from these methods and their details in Appendix B.

TABLE 8
Disaggregate Analysis

Variables	Quarterly Aggregation
Treatment effect	3.19**
	(1.55)
Observations	10,148
Adjusted R-square	.50
Firm fixed effects	Yes
Quarter fixed effects	Yes

$^{**}p < .05$.
Notes: Robust standard errors are in parentheses. A similar analysis is available in Jin and Leslie (2009).

Role of outliers. We estimated a significant treatment effect ($\beta_3 = .151,\ p < .05$) with log-transformed sales (see Table 6) because the log transformation mitigates the threat of outliers. We ideally sought to demonstrate the effect with untransformed data and thus only used a log-transformed model to confirm outlier-related robustness.

Definition of competitive intensity. Buyer firm size and buyer transaction share are objective measures not prone to design choice variations; we also considered alternative measures of competitive intensity. Rather than subtracting 1 from the industry concentration ratio of the sales of the top 20 firms to total sales in the industry, we used the top 4, top 8, and top 50 firms' sales. As Table 6 reveals, the significance of the treatment effect remained unchanged when we used these alternative measures of competitiveness in the buyer's industry.

Novelty effects and falsification. We checked whether the positive economic impact for XYZ might stem from the buyer's early (potentially fleeting) interest in using the app. To rule out novelty effects, we considered a 12-month postlaunch period, starting in the 4th month after the app launch to the end of the 15th month. The prelaunch period then started nine months before the launch date and ran until three months after its launch. Thus, we moved the three-month period after the launch to the prelaunch period, which should be adequate time for the novelty perceptions to wear off. In Table 7, the treatment effect's significance ($\beta_3 = 14.42,\ \ p < .05$) did not change much as a result of this adjustment. Thus, the economic impact of a buying firm's adoption seems persistent and not necessarily subject to novelty effects.

We also designed a falsification test to check whether the increase in sales to buyers in the treatment groups was due to the launch of the app. Because there was no app before September 2013, the treatment effect in the prelaunch period should be zero. Accordingly, we performed a difference-in-differences analysis of the prelaunch period data, treating months 2–8 as the prelaunch period and months 9–15 as the postlaunch period. The results in Table 7 confirm our intuition that no treatment effect existed prior to the launch; the effect is not significant ($\beta_3 = -3.33,\ p > .05$).

Quarterly aggregation. We used a two-period model to minimize heterogeneity in sales cycles across buyers in the sample, but in this robustness check, we reestimated the model using quarterly data. Buyer fixed effects account for buyer-specific, time-invariant factors that contribute to differences in sales. We also use quarterly fixed effects to account for time period–specific factors (e.g., seasonality)

TABLE 9
Range of Estimates

Estimation Method	Treatment Effect Estimate	Implied Annual Sales (%)
Selection on observables	16.01	19.98
Selection on unobservables	16.01	19.98
Nearest-neighbor matching (Mahalnobis distance)	16.27	20.31
Nearest-neighbor matching (Ivariance distance[a])	15.31	19.11
Nearest-neighbor matching (Euclidean distance)	18.26	22.79
Nearest-neighbor matching (two neighbors)	16.98	21.20
Nearest-neighbor matching (three neighbors)	16.93	21.13
Regression adjustment	16.20	20.22
Inverse probability weighting	16.25	20.28
Inverse probability weighting: regression adjustment	16.65	20.78

[a]Inverse diagonal sample covariate covariance

that might induce changes to sales patterns. The interaction term (postlaunch period × treatment group) captures the treatment effect, identified as within-buyer and over-time variation in sales, after controlling for stable firm and time-specific factors that contribute to changes. As Table 8 reveals, we retrieve a significant treatment effect ($\beta_3 = 3.19$, $p < .05$) even with data disaggregated to quarterly time units (see also Jin and Leslie 2009). The average annual economic benefit to XYZ in this model was $127,600, or an increase of 19.91% resulting from the introduction of the app.

In summary, across all identification strategies (selection on observables, selection on unobservables, estimator from potential outcomes framework) and temporal aggregations (two-period, quarterly), the annual economic benefit to XYZ due to the app featured sales increases in the range of 19.11%–22.79%, or an annual RoEI of $122,480–$146,080. We summarize the models and results in Table 9.

Sources of RoEI

Having established the presence of a RoEI, through increased sales, we examine whether the sales increases indicate more frequent purchases, larger quantities, or a broader variety of product purchases. Thus, we use purchase frequency, purchase volume, and purchase breadth as dependent variables. Sales increases and RoEI mainly resulted from purchase frequency and purchase breadth (see Appendix C). Using median splits of the sample, based on buyers' T_0 sales, we also identify a low– and a high–T_0 sales group. Purchase frequencies increased marginally for the low–T_0 sales group, but it did not show any increases in purchase breadth. Instead, we observe significant increases in purchase frequency, purchase breadth, and sales for the high–T_0 sales group (Appendix C). Next, we split the sample according to (1) firm size (i.e., small and large firms), (2) the industries to which buyers belong, and (3) the economic region (developing vs. developed) in which buyers are situated. We find significant sales increases for small (relative to large) firms, firms that belong to the general engineering industry category, and those in developing economies (see Appendix C). These analyses suggest heterogeneity in RoEI across buyers, depending on their size, industry, and location.

Participation Intensity

In keeping with our previous arguments, buyers' participation intensity is manifest through their repeated activity with the app, and it likely creates more value over time, such that it might expand the value appropriation opportunities for XYZ (Brodie et al. 2011; Kumar and Pansari 2016). We operationalize participation intensity as the number of machining assembly projects created by buyers through the app. Of the 522 buyers that adopted, 63 used it solely as a product search provider, but the remaining 459 firms used the process platform. We plot the histogram of the resulting projects in Figure 2, which reveals varying levels of buyer participation intensity in the treatment group. This variation is unique to the treatment group in the postlaunch period, so we use it to identify engagement mechanisms that likely drive the economic impact for XYZ.

A new variable, participation intensity, is a continuous variable that captures the total number of projects created

FIGURE 2
Distribution of the Number of Projects Created by Buyer Firms

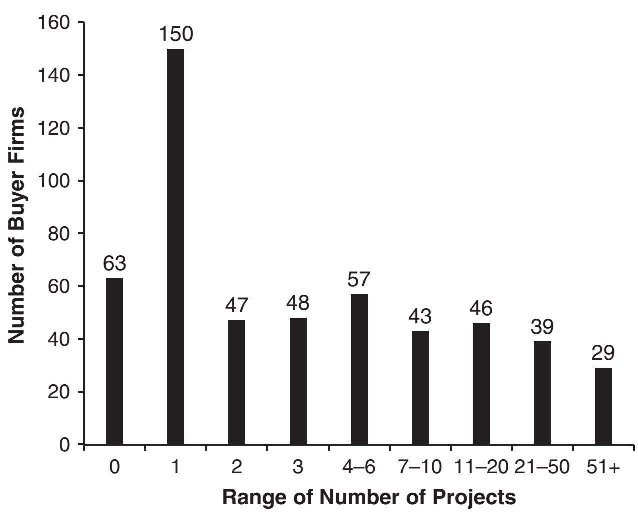


TABLE 10
Participation Intensity as Mechanism

Variables	Continuous	Nonparametric	Linear + Quadratic	Log Form	Square Root
Time dummy	–4.290	–2.682	–4.515	–10.86**	–12.44**
	(3.492)	(5.123)	(3.835)	(5.049)	(4.852)
Treatment dummy	–20.63***	–20.33***	–20.84***	–28.39***	–29.64***
	(5.735)	(6.543)	(5.681)	(6.105)	(5.925)
DD × Participation intensity	1.709***		1.781**		
	(.241)		(.701)		
DD × Low participation intensity		21.83			
		(17.35)			
DD × High participation intensity		15.21**			
		(6.678)			
DD × Participation intensity^2			–.00033		
			(.00219)		
DD × Log (Participation intensity)				21.92***	
				(5.771)	
DD × Participation intensity$^{.5}$					15.94***
					(3.463)
Buyer firm size	.597***	.599***	.597***	.580***	.582***
	(.215)	(.217)	(.215)	(.216)	(.215)
Buyer power	.224***	.224***	.224***	.224***	.224***
	(.0660)	(.0666)	(.0660)	(.0662)	(.0661)
Buyer industry competitiveness	–.286	–.290	–.285	–.291	–.287
	(.183)	(.184)	(.184)	(.184)	(.184)
Buyer T_0 sales	.0526	.0533	.0526	.0535	.0533
	(.0368)	(.0383)	(.0368)	(.0378)	(.0374)
Buyer T_0 purchase frequency	.403***	.435***	.402***	.412***	.402***
	(.0462)	(.0489)	(.0470)	(.0476)	(.0469)
Constant	80.20***	76.83***	80.27***	80.67***	82.48***
	(25.23)	(24.98)	(25.24)	(25.37)	(25.47)
Observations	2,296	2,296	2,296	2,296	2,296
R-square	.486	.466	.486	.473	.479
Division fixed effects	Yes	Yes	Yes	Yes	Yes
Continent fixed effects	Yes	Yes	Yes	Yes	Yes

*$p < .10$.
**$p < .05$.
***$p < .01$.
Notes: DD = product of treatment dummy and time dummy. Robust standard errors are in parentheses. For buyer firms that adopted the app, the participation intensity variable is nonzero in the postadoption period, but it is zero in the preadoption period. Participation intensity is also zero for the buyer firms that did not adopt.

by buyers. We incorporate this measure to reflect economic impacts, effectively scaling the difference-in-difference coefficient to capture the economic impact of app adoption as follows[9]:

$$S_{ijt} = \beta_1 I_j + \beta_2 I_t + \beta_3^{mech} I_j \times I_t \times \text{Participation Intensity}_i + \boldsymbol{\beta_4 Z_{ij}} + \varepsilon_{ijt}. \quad (8)$$

The interpretation of the scaled difference-in-difference coefficient (β_3^{mech}) thus changes. It still measures the change in sales in the treatment group (pre- vs. postlaunch period) with respect to the control group, but with our definition of participation intensity, we anticipate an increase in treatment effect size as participation intensity increases.

As the results in the first column of Table 10 show, we find a statistically significant coefficient ($\beta_3^{mech} = 1.709$, $p < .05$); the economic impact is increasingly positive for XYZ as participation intensity increases.[10] To verify the robustness of the results, we estimated separate difference-in-difference coefficients for lower– and higher–participation intensity buyers. Low participation intensity refers to buyers who downloaded the app but did not create any projects (n = 63). The high–participation intensity buyers instead downloaded the app and created at least one project (n = 459). In the second column of Table 10, we find a significant focal coefficient for high–participation intensity buyers ($\beta_3^{mech} = 15.210$, $p < .05$) but a statistically nonsignificant effect for low–participation intensity buyers ($\beta_3^{mech} = 21.830$, n.s.).

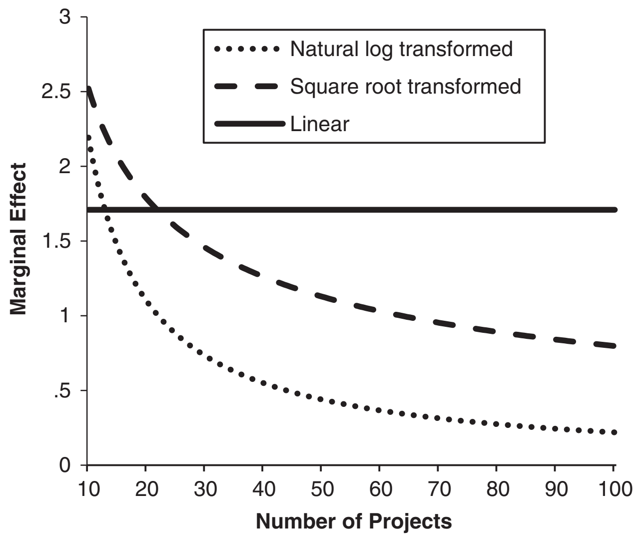
9Equation 8 includes only the three-way interaction term. We deliberately excluded lower-order interaction terms because their inclusion would induce perfect collinearity and fail to identify the impact of a one-unit increase in participation intensity on the buyer firm's purchases from the manufacturer. Note that this perfect collinearity arises because customers that do not adopt the app cannot have participation intensity.

10Participation intensity, operationalized as the number of machining assemblies, could be highly correlated with sales, so it might not qualify for a true mechanism. However, we find that participation intensity is not highly correlated with sales (.40). A high correlation likely would arise only if buyer firms had to purchase the products they used to create the project assemblies in the app.

Next, to understand nonlinearity in the participation intensity mechanism, we reestimated Equation 8 using linear quadratic (Column 3), logarithmic (Column 4), and square root (Column 5) functional forms instead of the linear functional form (Table 10). The results suggest a significant focal coefficient for the linear term of participation intensity (β_3^{mech}(linear) = 1.781, $p < .05$) but a statistically nonsignificant effect for its quadratic term (β_3^{mech}(quadratic) = −.0003, n.s.). The coefficient of the log of participation intensity ($\beta_3^{mech} = 21.92$, $p < .05$) is significant, indicating diminishing returns as participation intensity increases. This result is substantiated by the results showing a significant focal coefficient of the square root of participation intensity ($\beta_3^{mech} = 15.94$, $p < .05$), in support of the diminishing returns that occur as participation intensity increases. In Figure 3, we plot the marginal impact of the number of projects on sales with linear, logarithmic, and square root functional forms (we omitted the linear quadratic model because the quadratic term was not statistically significant). Each of these plots demonstrates visual evidence that RoEI increases with increasing participating intensity but also exhibits diminishing returns, as is common with marketing-mix interventions.

We thus uncover that the true economic impact on XYZ of offering the app stems from its ability to induce buyers to create projects. This source of continuous interaction between the buyer and seller seemingly enables indirect economic benefits to XYZ. Moreover, our results suggest a nonlinear impact of the number of projects on economic benefits to XYZ; an increasing treatment effect emerges as the number of projects increase, but with diminishing returns.

FIGURE 3
Marginal Effects from Various Transformations of Number of Projects



Notes: The marginal effect when the number of projects is log transformed is d(Sales)/d(Projects) = 21.92/# of projects, the marginal effect when the number of projects is square root transformed is d(Sales)/d(Projects) = 15.94/(2 × $\sqrt{\text{\# of projects}}$), and the marginal effect when number of projects remains the same is d(Sales)/d(Projects) = 1.709.

Discussion

Business-to-business firms use various touch points to interact with customers and maintain deep, continuous relationships. With firm-offered engagement initiatives, these firms aim to increase their interconnectedness with customers, even if the initiatives do not provoke any immediate sales outcomes. Engagement initiatives have been lauded for their ability to connect firms to their customers, but they also invoke direct economic costs, without direct economic benefits. We argue that despite these direct economic costs, engagement initiatives can create indirect economic benefits (e.g., increased revenues) by providing customer value through customer participation intensity, which can be appropriated in the form of RoEI. Our methodological framework, which provides suitable self-selection corrections, applied to novel, observational data from a manufacturer that launched a B2B app, confirms RoEI presence for the manufacturer. We leverage buyer-level variation in app adoption to establish buyer participation intensity as a driving mechanism. Our findings in turn have implications for both theory and practice.

Theoretical Implications

Foremost, our theoretical arguments have implications for customer engagement literature, which thus far has focused primarily on customer engagement (Kumar and Pansari 2016) as opposed to engagement initiatives, their definition, their role in enhancing customers' experiences (Lemon and Verhoef 2016), and their impact on psychological outcomes such as trust and referrals (Shiri, Beatty, and Morgan 2012). We instead take a return-on-marketing view, questioning whether and how engagement initiatives pay off for firms. Unlike traditional marketing interventions that work to stimulate sales outcomes, RoEI lacks a straightforward return-on-marketing-outcomes path. With our proposed methodological framework, we show that gauging RoEI requires researchers to test for indirect economic returns and overcome several econometric challenges, including controlling for self-selection by firms and customers. By doing so, we also add to the nascent but burgeoning research on the use of observational inference to document causal effects of strategic marketing decisions (e.g., Germann, Ebbes, and Grewal 2015; Shi et al. 2016). We find empirical evidence that RoEI increases nonlinearly with increased customer participation intensity, suggesting a rich link between immediate manifest outcomes of engagement initiatives and their economic returns

Furthermore, our empirical findings have implications for app research. From a substantive standpoint, the empirical evidence about the efficacy of engagement apps complements extant research that has focused mainly on sales apps (e.g., Einav et al. 2014; Ghose and Han 2014). Engagement apps can yield positive economic returns too, through indirect effects. This proposition has been discussed conceptually (Urban and Sultan 2015), but we add empirical evidence, obtained using objective sales data that pertain to the causal economic returns to engagement apps. Research designed to assess the efficacy of mobile apps (e.g., Einav et al. 2014; Urban and Sultan 2015; Xu et al. 2014) thus should account for both direct (e.g., nonzero app prices) and indirect effects when calculating overall marketing effectiveness.

Managerial Implications

Given the exponential rise of engagement initiatives, there is an urgent need to assess their returns to justify effective marketing spending. For managers, we provide an implementable methodological framework to test the hypotheses of positive RoEI, thus eliminating conjecture surrounding whether engagement initiatives generate indirect benefits. Moreover, we propose an approach that uses data on sales transactions and firmographic variables, both of which are ubiquitous in marketing organizations. Firms offering free mobile apps could use our approach to estimate the causal impact of free apps with a sample of buyers. Subsequently, they could extend the average sales increase from a sample to all buyers to arrive at the estimated gross economic benefit of the app, which can then be compared with the development cost of the app. Thus, our analysis had direct, tangible, managerial benefits for firm's product development efforts.[11]

Furthermore, in the context of mobile apps, our research provides evidence that RoEI can be positive, with participation intensity as the underlying mechanism. As increase in buyer participation intensity increases economic returns, app designers should incorporate features in the app that are specific to the institutional context under consideration to increase user participation intensity. For example, for business-to-customer (B2C) contexts, app designers might consider including features such as social sharing, product reviews, and instructional videos provided that these features would enhance participation intensity in their specific B2C context. Moreover, we find that RoEI is heterogeneous across customers, such that it varies substantially by buyer size, industry type, and region. Our results suggest that managers could consider customizing RoEI across customer segments to maximize the overall benefit from engagement initiatives.

Limitations

We close by noting the primary limitations of our study, which present avenues for further research. First, we used only one context to assess RoEI; thus, assessments of RoEI in other contexts (e.g., health care, B2C apps) are likely to be beneficial; eventually, a meta-analysis on the magnitude of RoEI would be useful. Second, we provided initial evidence for heterogeneity in RoEI across buyers, depending on their size, industry, and location. Conceptual understanding of engagement would likely improve with the development of formal hypotheses about whether and when to expect differences in RoEI. Third, although we focused on establishing the effectiveness of RoEI from the introduction of a free app, we did not try to explain substitution patterns across channels that resulted from the introduction of the app; future studies could examine this issue. Fourth, a fruitful avenue for further research would be in identifying the profit-maximizing level of engagement initiative spending, which could be possible using data on the cost structure of engagement initiatives. Fifth, whereas we use observational inference methods to establish the existence of RoEI, further research could leverage event studies or randomize field experiments to validate and augment our findings. Sixth, we focused on accounting performance (revenue) in our study, but future researchers could document the impact of engagement initiatives on customer mindset (e.g., satisfaction), product-market performance (e.g., market share), and financial market performance metrics (e.g., investor returns) (Katsikeas et al. 2016). Finally, we did not focus on employee engagement in our research. It would be worthwhile to determine how RoEI affects employee engagement and subsequent performance (Kumar and Pansari 2016).

[11]Our results, provided two years after the launch of the app (i.e., 6 months after the data period) convinced the program director that XYZ's app produced a positive RoEI within 15 months of its launch. After the chief marketing officer reviewed these data, XYZ approved further app development and marketing efforts. This prompted increased internal marketing efforts by XYZ to educate its salespeople about the app's functionalities, so marketing and sales force efforts could jointly encourage adoption.

Appendix A: Selection on Unobservables Strategy

We account for potential unobserved factors affecting buyer firms' decision to adopt the mobile app and sales by the manufacturer by including unobserved factors, obtained in the form of IMRs in the difference-in-differences model. We calculated the IMR for the firms in the treatment group and the control group using the expressions provided in Equations A1 and A2.

$$\text{(A1)} \qquad \text{Inverse Mills Ratio} = \frac{\Phi(P_i\theta)}{\Phi(P_i\theta)} \text{ if } \text{Treatment}_i = 1, \text{ and}$$

$$\text{(A2)} \qquad \text{Inverse Mills Ratio} = \frac{-\Phi(P_i\theta)}{1 - \Phi(P_i\theta)} \text{ if } \text{Treatment}_i = 0.$$

Thus,

$$z_i^* = P_i\theta + \zeta_i \text{ (Selection Equation)},$$

$$\text{Treatment}_i = 1 \text{ if } z_i^* > 0$$

$$y_{1i} = a_1 + X_i\beta + \varepsilon_{1i} \text{ (Outcome Equation)},$$

where y_1 is the outcome when firm i adopts the mobile app. Similarly,

$$\text{Treatment}_i = 0 \text{ if } z_i^* < 0$$

$$y_{0i} = a_0 + X_i\beta + \varepsilon_{0i} \text{ (Outcome Equation)},$$

where y_0 is the outcome when firm i does not adopt the mobile app. The treatment effect then is

$$\begin{aligned} & E(y_{1i}|z_i^* > 0) - E(y_{0i}|z_i^* < 0) \\ & = a_1 + X_i\beta + E(\varepsilon_{1i}|z_i^* > 0) - a_0 + X_i\beta + E(\varepsilon_{0i}|z_i^* < 0) \\ & = [a_1 + X_i\beta + E(\varepsilon_{1i}|P_i\theta + \zeta_i > 0)] - [a_0 + X_i\beta \\ & \quad + E(\varepsilon_{0i}|P_i\theta + \zeta_i < 0)] \\ & = [a_1 + X_i\beta + E(\varepsilon_{1i}|\zeta_i > -P_i\theta)] \\ & \quad - [a_0 + X_i\beta + E(\varepsilon_{0i}|\zeta_i < -P_i\theta)] \\ & = (a_1 - a_0) + [E(\varepsilon_{1i}|\zeta_i > -P_i\theta) - E(\varepsilon_{0i}|\zeta_i < -P_i\theta)] \\ & = \text{Treatment Effect} + \text{Unobserved Component}. \end{aligned}$$

The unobserved component will not equal zero if the errors in the selection equation and the errors in the outcome equation are correlated. However, if the unobserved component is not equal to zero (i.e., if $E(\varepsilon_{1i}|\zeta_i > -P_i\theta) - E(\varepsilon_{0i}|\zeta_i < -P_i\theta) \neq 0$), our estimate of the treatment effect would be biased. One way to overcome biased estimations of the treatment effect is to use parametric assumptions to model the unobserved component and include them along with the other covariates; conditional on the observed covariates and the unobserved component (i.e., selection on unobservables), the treatment effect should be unbiased. Thus, using the Heckman (1979) model, we assess unobserved component by assuming that the errors in the selection model and those in the outcome model are bivariate normally distributed, such that the unobserved component can be obtained as follows:

$$\text{Inverse Mills Ratio}_i = \text{Treatment}_i \frac{\Phi(P_i\theta)}{\Phi(P_i\theta)} + (1 - \text{Treatment}_i)\frac{-\Phi(P_i\theta)}{1 - \Phi(P_i\theta)}.$$

Appendix B: Potential Outcomes Framework

The treatment effect of app adoption represents the difference in the change in sales of a buyer due to app adoption (treatment) from the change in its sales without app adoption; when averaged across the population of firms, it represents the ATE, or mathematically:

$$\text{ATE} = E(y_1 - y_0). \quad \text{(B1)}$$

The sample equivalent of ATE (or ATE is estimated from a sample of size N) is

$$\widehat{\text{ATE}} = \frac{1}{N}\sum_{i=1}^{N}(y_{i1} - y_{i0}). \quad \text{(B2)}$$

Collectively, y_{i1} and y_{i0} are potential outcomes for a firm, and we must observe both outcomes to estimate the firm's treatment effect. However, it is not possible to estimate a firm-level treatment effect, because each firm either receives a treatment or does not. We only observe the outcome of a firm when it receives a treatment (denoted as y_{i1}) or does not receive a treatment (denoted as y_{i0}). The potential outcomes framework argues that a firm-level treatment effect could be estimated by treating the nonavailability of one of the potential outcomes as a missing data problem, then imputing the missing data using the methods available in treatment effects literature. We discuss some of these methods here.

Nearest-Neighbor Matching

In the nearest-neighbor matching method, the imputed value of the missing potential outcome of a buyer in the treatment or control condition is the outcome of the buyer that is most similar to the focal buyer, but present in a condition different from that of the focal buyer. A buyer could be similar to the focal buyer if its distance from the focal buyer—calculated using the observed covariates and employing Euclidean, Ivariance, or Mahalanobis distance metrics—is smaller than that of the other buyer s. Formally, let $x_i = \{x_{i1}, x_{i2}, x_{i3}, ..., x_{ip}\}$ be a vector of observed covariates of firm i. Then the distance between firm i and firm j is given as

$$||x_i - x_j|| = \left\{(x_i - x_j)'S^{-1}(x_i - x_j)\right\}, \quad \text{(B3)}$$

where S is a symmetric positive definite matrix, determined by the type of distance metric used for the nearest-neighbor matching. That is, S is an identity matrix when Euclidean distance serves to calculate the distance between two buyers; S can be a diagonal matrix consisting of the variance of all the covariates to account for the variation in covariates while calculating the distance between two firms; and S could be a variance-covariance matrix when using the Mahalanobis distance to account for the variance in each covariate and the correlation between the covariates. Thus, according to the distance metric used, the set of neighbor m_i firms, formally represented in Equation B4, could be considered similar to firm i. Both j and l refer to firms, but firms other than i.

$$\text{(B4)} \quad J_m^x(i) = \{j_1, j_2, j_3, \ldots, j_{mi}|t_j = 1 - t_i, ||x_i - x_j|| < ||x_i - x_l||, t_l = 1 - t_i\}, l \neq j\},$$

where t indicates the treatment. The potential outcomes then could be imputed as

$$\text{(B5)} \quad \widehat{y_i(1)} = \begin{cases} y_i \text{ if } t_i = 1 \\ \frac{1}{\#J_m^x}\sum_{j \in J_m^x} y_j \text{ if } t_i = 0 \end{cases}, \text{ and}$$

$$\text{(B6)} \quad \widehat{y_i(0)} = \begin{cases} y_i \text{ if } t_i = 0 \\ \frac{1}{\#J_m^x}\sum_{j \in J_m^x} y_j \text{ if } t_i = 1 \end{cases}.$$

The ATE is

$$\text{(B7)} \quad \text{ATE} = \frac{1}{N}\sum_{i=1}^{N}\left\{\widehat{y_i(1)} - \widehat{y_i(0)}\right\}.$$

In Table B1, we provide the results of the treatment effect estimated from a variety of nearest neighbor models.

Model-Based Imputation of Potential Outcomes

Unlike nearest-neighbor matching, model-based imputation relies on regression methods to impute potential outcomes by modeling the outcome (outcome model), the treatment (treatment model), or both. In the outcome model, separate regressions are first estimated (i.e., sales are regressed on the observed covariates) for buyers in the treatment group and buyers in the control group. Then the model-based estimates from the treatment group impute potential outcomes for a buyer in the control group. Estimates from the control group similarly function to impute potential outcomes for the treatment group. After obtaining all the potential outcomes, the ATE is calculated as $\text{ATE} = (1/N)\sum_{i=1}^{N}\{\widehat{y_i(1)} - \widehat{y_i(0)}\}$.

The inverse probability weighting method is a treatment model that accounts for the missing potential outcome

TABLE B1
Nearest-Neighbor Matching Model Results

	Nearest-Neighbor Matching				
	Euclidian Distance	Ivariance Distance[a]	Mahalanobis Distance	Two-Neighbors	Three-Neighbors
Treatment effect	18.26**	15.31**	16.27***	16.98**	16.93**
	(7.64)	(6.27)	(6.41)	(6.82)	(6.91)
Observations	1,148	1,148	1,148	1,148	1,148

***p* < .05.
****p* < .01.
[a]Inverse diagonal sample covariate covariance.
Notes: Robust standard errors are in parentheses.

problem by weighting the observations in the treatment and control groups by the inverse of the probability of receiving the treatment and not receiving the treatment, respectively, which can be estimated using various covariates in either probit or logit models.

Finally, in inverse probability weighting with regression adjustment, we would use a combination of regression adjustment and inverse probability weighting methods, with both outcome and treatment models used to obtain the treatment effect. The inverse probability weights come from estimating the treatment model using logit or probit. These estimated weights then reveal the weighted regression coefficients required to impute the potential outcomes, as in the regression adjustment method. After all the potential outcomes are imputed, the treatment effect can be estimated using $\mathrm{ATE} = (1/\mathrm{N})\sum_{\mathrm{i}=1}^{\mathrm{N}}\{\widehat{\mathrm{y_i}(1)} - \widehat{\mathrm{y_i}(0)}\}$.

In Table B2, we provide the results of the treatment effect estimated from the regression adjustment, inverse probability weighting method, and inverse probability weighting method with the regression adjustment method.

Table B2
Model-Based Imputation of Potential Outcomes

	(1)	(2)	(3)
	RA	IPW	IPWRA
Treatment effect	16.20***	16.25***	16.65**
	(6.28)	(6.20)	(6.48)
Observations	1,148	1,148	1,148

***p* < .05.
****p* < .01.
Notes: Robust standard errors are in parentheses. RA = regression adjustment; IPW = inverse probability weighting; IPWRA = inverse probability weighting with regression adjustment.

Appendix C: Source of RoEI

TABLE C1
Source of RoEI

Variables	Sales	Purchase Frequency	Purchase Volume	Volume Per Purchase	Purchase Breadth
Treatment effect	16.01**	22.90**	4,046	–45.68	.407**
	(6.523)	(11.24)	(3,982)	(54.59)	(.162)
Time dummy	–2.682	–2.337	1,151	51.47	–.0895
	(5.122)	(5.920)	(1,343)	(53.55)	(.102)
Treatment group dummy	–20.33***	18.74	–2,669	–89.68***	3.050***
	(6.542)	(14.56)	(4,702)	(28.65)	(.266)
Buyer firm size	.597***	.393**	410.8	.127	.0117***
	(.217)	(.199)	(279.0)	(.843)	(.00349)
Buyer power	.224***	.128**	67.05***	.274**	.00311***
	(.0667)	(.0630)	(21.53)	(.135)	(.00107)
Buyer industry competitiveness	–.290	.258	212.5**	–.746	–.000328
	(.184)	(.402)	(101.4)	(1.144)	(.00760)
Buyer T_0 sales	.0533	–1.433***	158.2	1.002	–.00151
	(.0383)	(.206)	(174.3)	(.825)	(.00224)
Buyer T_0 purchase frequency	.435***	5.086***	225.5***	–.409***	.0194***
	(.0489)	(.109)	(24.08)	(.141)	(.00182)

TABLE C1
Continued

Variables	Sales	Purchase Frequency	Purchase Volume	Volume Per Purchase	Purchase Breadth
Constant	76.93*** (25.05)	25.02 (53.59)	26,182** (12,603)	399.0** (187.5)	11.60*** (1.217)
Observations	2,296	2,296	2,296	2,296	2,296
Adjusted R-square	.461	.843	.255	.021	.317
Division fixed effects	Yes	Yes	Yes	Yes	Yes
Continent fixed effects	Yes	Yes	Yes	Yes	Yes

**$p < .05$.
***$p < .01$.
Notes: Robust standard errors are in parentheses. Purchase frequency is operationalized as the number of invoices by the buyer firm with the manufacturer. Purchase volume is the total number of units purchased. Purchase breadth is the number of distinct products purchased by buyer firms.

TABLE C2
Sales, Purchase Frequency, and Purchase Breadth Based on Prior Buyer Sales

Variables	Sales	Purchase Frequency	Purchase Volume	Volume Per Purchase	Purchase Breadth
Low T_0 Sales					
Treatment effect	5.835 (6.821)	9.023* (5.442)	7,780 (7,617)	14.82 (28.38)	.0457 (.0987)
Time dummy	3.669 (2.926)	−.0997 (3.152)	535.9 (630.4)	6.058 (23.41)	.0484 (.0625)
Treatment group dummy	.0258 (1.529)	13.34*** (4.591)	1,554 (1,342)	−48.90*** (14.77)	.717*** (.104)
Buyer firm size	−.186 (.182)	−.463 (.420)	−272.2 (202.2)	−1.330 (1.505)	−4.94e-05 (.00701)
Buyer power	.281** (.115)	.0809 (.0711)	48.16 (31.49)	.747*** (.193)	.00105 (.000851)
Buyer industry competitiveness	−.0205 (.144)	.339** (.143)	112.1 (128.7)	−.415 (.497)	−.00273 (.00254)
Buyer T_0 sales	2.077 (1.347)	1.944 (1.986)	730.9 (1,159)	16.42 (10.45)	.224*** (.0346)
Buyer T_0 purchase frequency	.0116 (.0756)	3.985*** (.239)	102.9 (94.06)	−1.163*** (.449)	.0137*** (.00295)
Constant	23.79*** (7.941)	70.13*** (24.92)	12,206*** (3,929)	123.3*** (35.34)	3.993*** (.529)
High T_0 Sales					
Treatment effect	26.97** (12.32)	34.44* (20.89)	935.4 (4,466)	−114.9 (119.5)	.238** (.102)
Time dummy	−10.79 (11.09)	−5.193 (12.93)	1,937 (2,964)	109.4 (118.7)	.0291 (.0696)
Treatment group dummy	−40.39*** (12.84)	38.38 (27.35)	−5,644 (9,199)	−123.1** (49.94)	.856*** (.140)
Buyer firm size	.554*** (.205)	.555*** (.207)	417.0 (280.7)	−.234 (1.064)	.00153 (.00148)
Buyer power	.249*** (.0727)	.165** (.0754)	68.54*** (21.58)	.223* (.115)	.000874** (.000339)

TABLE C2
Continued

			High T_0 Sales		
Buyer industry competitiveness	–.347 (.346)	.147 (.843)	413.4** (184.1)	–.404 (2.261)	–.00450 (.00434)
Buyer T_0 sales	.458 (.328)	–1.442*** (.226)	125.5 (147.3)	.787 (.654)	–.000506 (.000396)
Buyer T_0 purchase frequency	.380*** (.0434)	5.212*** (.120)	202.7*** (24.19)	–.650*** (.191)	.00432*** (.000493)
Constant	112.3*** (40.79)	3.630 (101.2)	30,970 (20,289)	536.5 (334.3)	5.276*** (.570)
Observations	1,148	1,148	1,148	1,148	1,148
Division and continent fixed effects	Yes	Yes	Yes	Yes	Yes

*$p < .10$.
**$p < .05$.
***$p < .01$.
Notes: Robust standard errors are in parentheses. T_0 is prior sales, and Low T_0 and High T_0 sales are sales above/below median, respectively.

TABLE C3
Treatment Effect by Firm Size

Variables	Small Firms	Big Firms
Treatment effect	18.84*** (6.983)	13.65 (11.27)
Time dummy	1.431 (2.650)	–6.956 (10.12)
Treatment group dummy	–1.175 (5.003)	–37.83*** (10.68)
Buyer firm size	–12.71 (11.69)	.358 (.220)
Buyer power	.0818*** (.0220)	.378*** (.0525)
Buyer industry competitiveness	.195 (.121)	–.220 (.309)
Buyer T_0 sales	3.048*** (.449)	.315 (.233)
Buyer T_0 purchase frequency	.198*** (.0517)	.445*** (.0566)
Constant	17.14 (41.24)	85.62*** (29.80)
Observations	1,148	1,148
Adjusted R-square	.669	.495
Division fixed effects	Yes	Yes
Continent fixed effects	Yes	Yes

***$p < .01$.
Notes: Robust standard errors are in parentheses. We operationalized firm size as the number of employees in a firm.

TABLE C4
Treatment Effect by Industry Type

Variables	Transportation and Aerospace	Heavy Equipment	General Engineering
Treatment effect	19.02 (14.39)	36.79 (36.21)	11.30* (6.030)
Time dummy	1.178 (8.005)	–23.97 (35.62)	–1.115 (4.873)
Treatment group dummy	–7.125 (6.937)	–50.22* (27.66)	–4.120 (7.111)
Buyer firm size	.189** (.0822)	.392 (.335)	.164 (.206)
Buyer power	.534*** (.141)	.00434 (.105)	.173*** (.0585)
Buyer industry competitiveness	–.251 (.209)	.0165 (.443)	–.352 (.266)
Buyer T_0 sales	1.807*** (.483)	3.835*** (.599)	.317 (.203)
Buyer T_0 purchase frequency	–.109 (.140)	.382*** (.134)	.460*** (.0390)
Constant	39.14*** (14.43)	212.2* (112.0)	71.64** (27.75)
Observations	674	252	1,370
Adjusted R-square	.500	.665	.520
Division fixed effects	Yes	Yes	Yes
Continent fixed effects	Yes	Yes	Yes

*$p < .10$.
**$p < .05$.
***$p < .01$.
Notes: Robust standard errors are in parentheses.

TABLE C5
Treatment Effect by Economy Type

Variables	Developing Economies	Developed Economies
Treatment effect	70.28***	7.166
	(25.42)	(6.171)
Time dummy	–34.32**	1.888
	(15.16)	(5.430)
Treatment group dummy	–20.13	–18.45***
	(13.56)	(6.785)
Buyer firm size	2.987	.650***
	(1.826)	(.222)
Buyer power	.459***	.240***
	(.130)	(.0583)
Buyer industry competitiveness	1.198**	–.386**
	(.597)	(.185)
Buyer T_0 sales	2.427***	.495
	(.571)	(.357)
Buyer T_0 purchase frequency	–.0685	.433***
	(.156)	(.0461)
Constant	–1.104	44.16***
	(24.80)	(13.15)
Observations	324	1,972
Adjusted R-square	.473	.491
Division fixed effects	Yes	Yes
Continent fixed effects	Yes	Yes

$^{**}p < .05$.
$^{***}p < .01$.
Notes: Robust standard errors are in parentheses.

Table C6
Average Firm Size by Industry Type and Economy Type

Industry	Avg. Employees
Transportation and Aerospace Industry	7.38
Heavy Equipment	10.21
General Engineering	2.79
Economy Type	**Avg. Employees**
Developing	4.16
Developed	5.08

Notes: Firm size is operationalized as the number of employees.

REFERENCES

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

Athey, Susan, and Guido Imbens (2016), "The Econometrics of Randomized Experiments," https://arxiv.org/abs/1607.00698.

Beebee, Cheryl (2013), "The Increasing Adoption of Apps in the Manufacturing Workplace," Fishman Company Technical Report (accessed March 10, 2016), http://www.fishmancorp.com/manufacturing-apps.

Bellman, Steven, Robert F. Potter, Shiree Treleaven-Hassard, Jennifer A. Robinson, and Duane Varan (2011), "The Effectiveness of Branded Mobile Phone Apps," *Journal of Interactive Marketing*, 25 (4), 191–200.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), "How Much Should We Trust Difference-in-Difference Estimates?" *Quarterly Journal of Economics*, 119 (1), 249–75.

Brodie, Roderick J., Linda D. Hollebeek, Biljana Juric, and Ana Ilic (2011), "Customer Engagement: Conceptual Domain, Fundamental Propositions, and Implications for Research," *Journal of Service Research*, 14 (3), 1–20.

Carare, Octavian (2012), "The Impact of Bestseller Rank on Demand: Evidence from the App Market," *International Economic Review*, 53 (3), 717–42.

Einav, Liran, Jonathan Levin, Igor Popov, and Neel Sundaresan (2014), "Growth, Adoption, and Use of Mobile e-Commerce," *American Economic Review*, 104 (5), 489–94.

Garg, Rajiv, and Rahul Telang (2013), "Inferring App Demand from Publicly Available Data," *Management Information Systems Quarterly*, 37 (4), 1253–64.

Germann, Frank, Peter Ebbes, and Rajdeep Grewal (2015), "The Chief Marketing Officer Matters!" *Journal of Marketing*, 79 (May), 1–22.

Ghose, Anindya, and Sang Pil Han (2014), "Estimating Demand for Mobile Applications in the New Economy," *Management Science*, 60 (6), 1470–88.

Guo, Shenyang, and Mark W. Fraser (2010), *Propensity Score Analysis: Statistical Methods and Applications: Statistical Methods and Applications*. Thousand Oaks, CA: Sage Publications.

Han, Sang Pil, Sungho Park, and Wonseok Oh (2016), "Mobile App Analytics: A Multiple Discrete-Continuous Choice Framework," *Management Information Systems Quarterly*, 983–1008.

Heckman, James J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47 (1), 153–61.

International Data Group Market Report (2013), "Nearly 80% of Manufacturers to Develop Mobile Application This Year, According to IDC Manufacturing Insig," (July 29), https://www.idg.com/news/nearly-80-of-manufacturers-to-develop-mobile-application-this-year-according-to-idc-manufacturing-insig/.

Jin, Ginger Zhe, and Phillip Leslie (2009), "Reputational Incentives for Restaurant Hygiene," *American Economic Journal. Microeconomics*, 1 (1), 237–67.

Katsikeas, Constantine S., Neil A. Morgan, Leonidas C. Leonidou, and G. Tomas M. Hult (2016), "Assessing Performance Outcomes in Marketing," *Journal of Marketing*, 80 (March), 1–20.

Kumar, V., Vikram Bhaskaran, Rohan Mirchandani, and Milap Shah (2013), "Creating a Measurable Social Media Marketing Strategy: Increasing the Value and ROI of Intangibles and Tangibles for Hokey Pokey," *Marketing Science*, 32 (2), 194–212.

Kumar, V. and Anita Pansari (2016), "Competitive Advantage Through Engagement," *Journal of Marketing Research*, 53 (August), 497–514.

Kwon, Hyeokkoo Eric, Hyunji So, Sang Pil Han, and Wonseok Oh (2016), "Excessive Dependence on Mobile Social Apps: A Rational Addiction Perspective," http://pubsonline.informs.org/doi/abs/10.1287/isre.2016.0658?journalCode=isre.

Lee, Ju-Yeon, Shrihari Sridhar, Conor M. Henderson, and Robert W. Palmatier (2015), "Effect of Customer-Centric Structure on Long-Term Financial Performance," *Marketing Science*, 34 (2), 250–68.

Lee, Ruby P., and Rajdeep Grewal (2004), “Strategic Responses to New Technologies and Their Impact on Firm Performance,” *Journal of Marketing*, 68 (October), 157–71.

Lemon, Katherine N., and Peter C. Verhoef (2016), “Understanding Customer Experience Throughout the Customer Journey,” *Journal of Marketing*, 80 (November), 69–96.

Manchanda, Puneet, Grant M. Packard, and Adithya Pattabhiramaiah (2015), “Social Dollars: The Economic Impact of Customer Participation in a Firm-Sponsored Online Customer Community,” *Marketing Science*, 34 (3), 367–87.

Reinartz, Werner, and Wolfgang Ulaga (2008), “How to Sell Services More Profitably,” *Harvard Business Review*, 86 (5), 90–96.

Sawhney, Mohanbir (2006), “Going Beyond the Product, Defining, Designing and Delivering Customer Solutions,” in *The Service-dominant Logic of Marketing: Dialog, Debate, and Directions*, Robert F. Lusch and Stephen L. Vargo, eds. New York: M.E. Sharpe, 356–80.

Shi, Huanhuan, Shrihari Sridhar, Rajdeep Grewal, and Gary Lilien (2016), “Sales Representative Departures and Customer Reassignment Strategies in Business-to-Business Markets,” Journal of Marketing, 81 (March), 25–44.

Shiri, D. Vivek, Sharon E. Beatty, and Robert M. Morgan (2012), “Customer Engagement: Exploring Customer Relationships Beyond Purchase,” *Journal of Marketing Theory and Practice*, 20 (2), 122–46.

Srinivasan, Raji, and Sundar Bharadwaj (2004), “Event Studies in Marketing Strategy Research,” in *Assessing Marketing Strategy Performance*, Christine Moorman and Donald R. Lehmann, eds. Cambridge, MA: Marketing Science Institute, 9–28.

Urban, Glen L., and Fareena Sultan (2015), “The Case for ‘Benevolent’ Mobile Apps,” *MIT Sloan Management Review*, 56 (2), 31.

Van Doorn, Jenny, Katherine N. Lemon, Vikas Mittal, Stephan Nass, Doreén Pick, Peter Pirner, et al. (2010), “Customer Engagement Behavior: Theoretical Foundations and Research Directions,” *Journal of Service Research*, 13 (3), 253–66.

Xu, Jiao, Chris Forman, Jun B. Kim, and Koert Van Ittersum (2014), “News Media Channels: Complements or Substitutes? Evidence from Mobile Phone Usage,” *Journal of Marketing*, 78 (July), 97–112.