
Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# The Effect of Customers' Social Media Participation on Customer Visit Frequency and Profitability: An Empirical Investigation

Rishika Rishika, Ashish Kumar, Ramkumar Janakiraman, Ram Bezawada,

*Information Systems Research*
Vol. 24, No. 1, March 2013, pp. 108–127
ISSN 1047-7047 (print) | ISSN 1526-5536 (online)

http://dx.doi.org/10.1287/isre.1120.0460




# The Effect of Customers' Social Media Participation on Customer Visit Frequency and Profitability: An Empirical Investigation

Rishika Rishika
Mays Business School, Texas A&M University, College Station, Texas 77843,
rrishika@mays.tamu.edu

Ashish Kumar
Department of Marketing, Aalto University School of Business, FI-00076 Aalto, Finland,
ashish.kumar@aalto.fi

Ramkumar Janakiraman
Mays Business School, Texas A&M University, College Station, Texas 77843,
janakiraman.ramkumar@gmail.com

Ram Bezawada
School of Management, State University of New York, Buffalo, New York 14260,
bezawada@buffalo.edu

In this study we examine the effect of customers' participation in a firm's social media efforts on the intensity of the relationship between the firm and its customers as captured by customers' visit frequency. We further hypothesize and test for the moderating roles of social media activity and customer characteristics on the link between social media participation and the intensity of customer-firm relationship. Importantly, we also quantify the impact of social media participation on customer profitability. We assemble a novel data set that combines customers' social media participation data with individual customer level transaction data. To account for endogeneity that could arise because of customer self-selection, we utilize the propensity score matching technique in combination with difference in differences analysis. Our results suggest that customer participation in a firm's social media efforts leads to an increase in the frequency of customer visits. We find that this participation effect is greater when there are high levels of activity in the social media site and for customers who exhibit a strong patronage with the firm, buy premium products, and exhibit lower levels of buying focus and deal sensitivity. We find that the above set of results holds for customer profitability as well. We discuss theoretical implications of our results and offer prescriptions for managers on how to engage customers via social media. Our study emphasizes the need for managers to integrate knowledge from customers' transactional relationship with their social media participation to better serve customers and create sustainable business value.

*Key words*: social media marketing; social media participation; customer-firm relationship; shopping visit frequency; customer profitability; propensity score matching; quasi-experiment; difference-in-differences
*History*: Chris Dellarocas, Senior Editor; Pei-Yu Chen, Associate Editor. This paper was received on January 12, 2012, and was with the authors 4 months for 2 revisions. Published online in *Articles in Advance* December 20, 2012.

## 1. Introduction

Social media has changed the way individuals interact with each other and is redefining the way firms connect with their customers. With 75% of Internet users participating in some form of social media (Forrester Research 2008), companies are increasingly investing in the creation of social media platforms and strategies to augment customer-firm interactions. A recent article in ZDNet (2011) reports that 69% of small business owners use some form of social media (e.g., Facebook, Twitter) and about 78% of them plan to increase their budgets devoted to this medium. Social media enables firms to converse about their offerings with customers interactively while also providing them the opportunity to share information about products (Godes and Mayzlin 2004, Agarwal et al. 2008). Thus, social media has provided firms—both large and small—with a new tool for customer engagement.

Despite the eagerness on the part of firms to embrace social media to connect with customers, there is also much skepticism about its efficacy. For example, a recent IBM report (2011, p. 3) mentions that "social media is no longer the adorable baby everyone wants to hold, but the angst-filled adolescent—still immature yet no longer cute—who inspires mixed feelings." The doubts about the effectiveness of social media arise because the link between firms' social media efforts and the return on their investment has not been established. In particular, no study, using individual level data to our knowledge, has connected customers' social media participation with their actual behavior. Recent studies in the broad area of social media have primarily focused on the effectiveness of user generated content (UGC) in stimulating product sales (Godes and Mayzlin 2004, Dellarocas et al. 2007, Forman et al. 2008, Zhu and Zhang 2010), encouraging content diffusion (Susarla et al. 2012), and fostering acceptance of recommendations (Ho et al. 2011). However, from the perspective of a firm that is contemplating social media investments, it is imperative to know if such expenditures help in furthering the intensity of customer-firm relationships and in creating sustainable firm value. In this regard, because most of the extant studies use aggregate data, they are unable to uncover critical individual level differences that are essential for obtaining insights at the customer level. Thus, the objective of this paper is to systematically examine the effect of customers' participation in firm initiated social media efforts on the business value generated thereof for the firm. For this purpose, we choose to work with customers' shopping visit frequency as the focal variable because it captures the intensity of relationship between a firm and a focal customer and serves as a key driver of customer lifetime value (Venkatesan and Kumar 2004). Furthermore, we also examine the impact of customer social media participation on customer profitability because that directly contributes to a firm's bottom line.

To study the central question concerning the effectiveness of firms' social media strategy, we first examine the main effect of customers' participation in social media on the intensity of relationship between customers and the firm (as measured by customers' frequency of shopping visits). We then hypothesize and test for the moderating role of overall activity in the social media site and a set of customer characteristics. Because contribution margin from customers is another key component of the customer lifetime value, we next complement our core findings related to visit frequency by examining the effect of social media participation on customer profits. In order to accomplish our objectives, we assemble a novel data set in which we observe customer participation decisions in a firm initiated social media site and connect this to individual purchase data for the *same* set of customers. We believe that our study is the first to establish the link between customers' social media participation and the transaction driven firm value generated thereof using actual behavioral data.

However, in our research context, it is important to account for endogeneity issues arising because of customer self-selection. For instance, customers who have an affinity toward the firm are also more likely to join its social media site and have an intensive relationship with the firm. To control for this self-selection issue, we employ the combination of propensity score matching (PSM) technique and difference-in-differences (DID) analysis. In particular, we utilize a quasi-experimental research design with two distinct groups of customers: (1) a "treatment" group that comprises customers who participate in the firm initiated social media site and whose behaviors we wish to analyze and (2) a "control" group consisting of customers who are not part of or never participate in the firm's social media site. We use PSM in combination with DID analysis because it helps mimic a random experimental design, thus producing an unbiased estimate of the "treatment effect," i.e., the impact of customer social media participation (Dehejia and Wahba 1999, Huang et al. 2012). We then extend the DID modeling framework to examine how the social media participation effect varies across customers depending on their level of relationship with the firm.

Our study makes three substantial contributions to the social media literature. First, we attempt to uncover the direct benefits that accrue to firms as a consequence of their social media efforts. Our approach is different from the extant research (e.g., Dholakia and Durham 2010, Algesheimer et al. 2010) in that we use actual behavioral data to establish the causal link between customers' social media participation and their buying behavior. Moreover, unlike the above studies, we explicitly account for endogeneity issues arising because of self-selection. Second, no study to our knowledge has integrated customers' social media participation behavior with their transaction relationship with the firm and studied how customer-firm relationships in online and offline domains may interact to create value for the firm. This study helps fill this gap in the literature. Third, our research also examines how the effect of social media participation behavior varies across different types of customers. By focusing on transaction based customer characteristics, we offer insights on when participation becomes more salient. Moreover, we quantify the economic benefits that result for firms due to customers' social media participation. Managers can use these results to understand how customer relationships cultivated in the online social media environment can be leveraged to optimize return on investment.

## 2. Hypotheses Development

The focus of the study is on examining how social media efforts by the firm help transform the intensity of customer-firm relationships. Accordingly, we measure this through a key characteristic representing customers' transactional relationship with the firm, frequency of visits. We believe that customer visit (or purchase) frequency captures the strength of customer-firm relationship because prior research has considered visit frequency to be an important indicator of the relationship that is also positively linked to customer lifetime value (e.g., Venkatesan and Kumar 2004).

In examining the impact of social media participation on the intensity of customer-firm relationship, we focus on two sets of factors that can influence this relationship: (a) overall social media activity and (b) customer characteristics. Social media has become a preferred platform for cultivating meaningful relationships between firms and their customers. However, from a managerial perspective, it is crucial to understand the moderating factors that make customers' social media participation more or less salient. In particular, we investigate how the level of activity on the social media site in terms of number of messages and comments posted by the firm and customers moderates the effect of social media participation. Furthermore, managers often segment customers based on their transaction characteristics for better targeting and optimal allocation of resources. For business executives, it is important to know on which customer segments they must spend greater resources in order to build and strengthen relationships (Rigby et al. 2002). Thus, an understanding of how the impact of social media participation differs across customer segments can enable firms to develop better customer relationship management (CRM) initiatives. We draw on CRM literature to identify key customer characteristics (i.e., purchase amount, buying focus, deal sensitivity, and premium product purchases) that can help in ascertaining the differential response among different customer segments and aid in managerial decision making.

### 2.1. Customers' Participation in Firm Hosted Social Media

By participating in the firm initiated social media site, customers can strengthen their relationship with the firm through many processes. First, the social media platform provides an opportunity for customers to voice their opinions and share experiences about the firm with other customers as well as the firm, thereby creating social interactions (Agarwal et al. 2009). In the words of Kietzman et al. (2011, p. 241), "social media employ mobile and web-based technologies to create highly interactive platforms via which individuals and communities share, co-create, discuss, and modify user-generated content." Thus, this ability of social media to facilitate greater interactions between consumers and the firm can help in establishing a deeper connection between them.

Second, participation in a firm's social media site provides an easy access to a social network consisting of other customers' opinions and experiences. An active network with message postings from other customers regarding their experiences with the firm and its products creates the feeling of a brand/firm community that will positively influence a focal customer's intensity of relationship with the firm. Research suggests that relationships among fellow customers, who are an integral part of a community, help in the fostering of firm loyalty, thereby intensifying the relationship with the firm (McAlexander et al. 2002).

Third, participating in a firm hosted social media site provides customers easy access to information on firms' offerings and other related messages. Through these postings, customers can infer the firm's level of involvement with its customer base and its commitment toward enhancing their experiences. Such increased interactions can lead to better customer identification with the firm and create greater trust and customer loyalty (Algesheimer et al. 2005). Thus, firm hosted social media opens a new channel of communication through which customers can connect and communicate directly with the firm and other customers that will positively affect the intensity of customer-firm relationship. Based on the above discussion, we hypothesize the following:

Hypothesis 1 (H1). *Participation in firm hosted social media by a focal customer will have a positive impact on the intensity of the customer-firm relationship.*

### 2.2. Interaction Effect with Social Media Activity

Customers who choose to engage with the firm through its social media site gain easier access to messages from both the firm and other customers. We believe that the impact of social media participation on customer-firm relationship will be influenced by the level of activity in the social media site in terms of the number of messages posted by the firm and other customers. First, because one of the main purposes of the firm hosted social media site is to create a platform for communicating with its customers and building relationships with them, a firm must maintain a high level of activity on the social media site. A more active and vibrant community with regular new message postings will create trust and allow customers to infer the level of a firm's relationship commitment and bolster customers' bond with the firm. Extant research suggests that commitment and trust in a relationship are positively associated with a party's intentions to continue the relationship with the other party (Morgan and Hunt 1994). Upon creating social media sites (e.g., Facebook pages, Twitter accounts), if firms fail to interact regularly with their clientele through this new medium, it could create customer skepticism and result in less favorable behavioral outcomes. Second, a higher number of postings by other customers reflects the level of customers' involvement with the firm. Individuals often wish to validate their choices in an environment that further strengthens their own affinity with the firm (Zhu et al. 2009). In the context of a social media site, by scrutinizing other customers' behavior and comments, focal customers can feel more confident about their own choices, thus impacting their intensity of relationship with the firm.[1]

Next, a mere exposure effect may also exist that may make the firm more accessible to the customer (Janiszewski 1993). As customers are exposed to a greater number of messages from the firm and other customers about its products, they become more favorably disposed toward the firm. Thus, an active social media page with regular new messages/postings can help customers form more positive attitudes toward the firm and strengthen the customer-firm relationship. Hence, we hypothesize that the impact of customers' social media participation on the intensity of their relationship with the firm will be greater for a higher level of activity (e.g., messages, comments) on the social media site.

Hypothesis 2 (H2). *The impact of participation in firm hosted social media on the intensity of the customer-firm relationship will be greater for a higher level of message postings in the social media site.*

### 2.3. Interaction Effects with Customer Characteristics

*Purchase Amount.* Although participation in a firm hosted social media site will have a positive effect on customer-firm relationship for all customers (as argued for H1), we expect that customers who have a stronger transaction relationship with the firm (before participation in social media), as measured by their spending levels, will likely experience greater benefits from association with the social media site. This is likely to happen through two mechanisms. First, extant literature in CRM suggests that exchange characteristics such as average transaction amount are associated with greater satisfaction (Bolton 1998) and commitment that results in better behavioral outcomes, which strengthens the relationship between the firm and the consumer (Reinartz and Kumar 2003). Thus, customers who spend a higher amount with the firm are likely to be more committed and more relationship oriented than are customers who spend less with the firm (Crosby et al. 1990). Because they have a greater affinity toward the firm, customers with a larger pecuniary investment will exhibit a greater response to the firm's relationship building communications through social media than will customers with lower financial investments. Second, prior research also shows that customers with higher spending levels, due to their greater financial investments with the firm, feel that they are more important to the firm; therefore, they are more demanding and expect greater relationship investments on the part of the firm (Wangenheim and Bayón 2007). Thus, high value customers are likely to value firms' relationship investments in social media more than low value customers are. Hence, we expect response to social media participation to be greater for customers with higher spending levels as compared to customers with lower spending levels. Hence, we hypothesize the following:

Hypothesis 3 (H3). *The impact of participation in firm hosted social media on the intensity of the customer-firm relationship will be greater for customers with a larger purchase amount.*

*Focus of Buying.* Buying focus refers to the breadth of interaction between a customer and the focal firm, and extant research considers buying focus to be an important indicator of relationship status (Verhoef and Donkers 2005). High buying focus typically entails limited customer-firm interaction, wherein a focal customer tends to buy exclusively from a single product category and thereby does not broaden the scope of interactions with the firm. Research suggests that lower customer satisfaction with the offerings of the firm can manifest itself as highly focused purchasing behavior (Verhoef et al. 2001). Lower customer satisfaction has been linked to less favorable outcomes in several studies (e.g., Anderson and Sullivan 1993, Bolton 1998). On the other hand, less focused buying (purchasing across a wide range of products and categories) increases customers' familiarity with the firm and can lead to greater satisfaction with the firm. For these less focused buyers, resources on the social media site can prove to be particularly valuable. Because the social media site provides exposure to information on new and different products, less focused customers can reap greater benefits from participation in social media because they tend to purchase across multiple product categories. Thus, we expect that the impact of social media participation on the intensity of customer-firm relationship will be

[1] We note that these arguments are valid only if most customer comments are positive. We also wish to point out that in our context, all of the messages that originate from the firm are positive and more than 90% of customer comments are also positive in nature. This is likely because this is a firm-hosted social media site where most of the messages originate from the firm and customer comments are in response to these messages and generally take the form of positive comments or general queries. This is unlike a third party website where customers may post product reviews and recommendations that can be either positive or negative. We thank an anonymous reviewer for seeking clarification on this issue.

lower for customers having a greater buying focus. Therefore,

Hypothesis 4 (H4). *The impact of participation in firm hosted social media on the intensity of the customer-firm relationship will be lower for customers who have a greater focus in buying.*

*Deal Sensitivity.* An important customer characteristic in the context of customer-firm relationship is the extent to which a focal customer looks to buy items on promotion. Extant research suggests that customers who are more deal prone tend to search more for products from competing sellers and tend to make their purchases with the seller who offers the lowest price (Ailawadi et al. 2001). Thus, deal prone consumers are likely to be less committed to the firm and have reduced interest in developing a stronger relationship with the firm. These customers may visit the firm's social media site only to search for coupons or promotions. In such a case, we would expect firms' relationship building efforts in the form of engaging in interactive communications with their customers through social media to elicit a greater response from less deal prone consumers who may appreciate firms' relationship investments more.[2] Thus, we argue that online social media participation would strengthen the customer-firm relationship for customers who are less deal prone:

Hypothesis 5 (H5). *The impact of participation in firm hosted social media on the intensity of the customer-firm relationship will be greater for customers with lower deal sensitivity.*

*Share of Premium Products.* Customers who purchase the firm's most valuable offerings are also considered to be its most lucrative customers, and firms often invest considerable amount of resources in nurturing relationships with them (Reinartz and Kumar 2000). In addition, premium products (because of their higher price tags) are often high involvement purchases, which in turn are associated with greater levels of uncertainty and perceived risk (Pavlou et al. 2007).[3] Customers with a larger proportion of high ticket purchases are likely to have greater switching costs and thus may be more inclined to continue the relationship with their current firm and attempt to gain preferential status with it. Moreover, interactions with the firm through the online social media can help such customers achieve these objectives through obtaining a greater degree of personalization and information for premium products in real time. Moreover, they may also utilize social media to gather cues about premium products from other customers of the social media site to mitigate uncertainty (Pavlou et al. 2007). Thus, participation in social media is likely to strengthen the intensity of customer-firm relationship for customers who purchase a greater share of premium products:

Hypothesis 6 (H6). *The impact of participation in firm hosted social media on the intensity of the customer-firm relationship will be greater for customers who have a greater share of premium product purchases.*

## 3. Research Setting

The data set for this study comes from a large specialty firm that owns multiple stores and sells wine and like products in the northeastern United States.

### 3.1. The Firm's Customer Engagement Initiative via Social Media

In August 2009, the focal firm, for the first time, created a webpage on a major social networking website (e.g., Facebook) with the objective of connecting with its customers through social media.[4] Since then, it has made conscious efforts to continue to invest in social media by regularly updating the site through message postings and by encouraging customers to visit the site and become involved by posting, commenting, and other such activities. The messages posted by the firm on the above site differ both in their content and characteristics. In particular, the site is used by the firm to build store equity, foster better customer relationships, communicate to customers about various events/activities, and disseminate information about its products and promotions. Examples of equity and relationship building messages include "we are currently touring wineries in Argentina to get you the best wines" and "we are constantly striving to improve your shopping experiences," respectively. Messages that inform customers about various events usually relate to community based activities (e.g., 5K runs for charity). Other message postings can be solely product (e.g., "how to select a good

[2] We note that in case the firm's social media site is mostly devoted to offering coupons or promotions, we would expect the opposite effect where deal prone customers would shop with the firm more. However, our focus is on understanding the impact of firms' relationship building efforts through social media, where firms engage in a variety of communications with their consumers (both product and non-product related), which is the strategy used by the firm under study. We thank an anonymous reviewer for bringing our attention to this issue.

[3] We note that uncertainty can be due to either seller or product uncertainty. Our focus here is on product related uncertainty because consumers perceive a high risk associated with high-involvement purchases because of their higher uncertainty (Pavlou et al. 2007). We thank an anonymous reviewer for helping us clarify the issue.

[4] The confidentiality arrangements with the firm preclude us from disclosing the name of the firm. We also cannot furnish details that allow identification of the social networking website.

Chardonnay") or promotion (e.g., "buy one bottle at regular price and get another free") related.[5]

When the firm launched the social media site, it also undertook a marketing campaign to inform its clientele about these developments. As a result, existing as well as potential customers began participating by signing up and/or following the firm on its social media site. However, they differ with respect to their timing of participation, with some customers participating soon after the site was launched whereas others did so at a later date. No incentives (either monetary or promotional) were offered, and customers signed up of their own volition.

### 3.2. Social Media Participation and Customer Transaction Data

A key and notable contribution of our study is that we combine data related to customers' participation in the focal firm's social media site with their actual purchases. This data gathering process is not straightforward and presents a rather onerous task given the nature and idiosyncrasies associated with this kind of work. We adopted the following multi-step approach to identify customers who participate (e.g., become a "fan" of the firm on Facebook) in the firms' social media site. We first begin with a survey that was sent to the firm's customer base in early 2011. This survey that was sent to 5,000 customers featured a section devoted to social media wherein customers were specifically asked if they participate in the firm's social media site in question (e.g., the firm's Facebook page).[6]

From the survey, we identified 845 customers who actively participate in the firm's social media site (e.g., Facebook). Working in conjunction with the firm, we then matched the above customers with actual participants of the specific social media site in question. This was possible because the firm was tracking its participants on this site (e.g., Facebook) since the inception of its program. For this purpose, complete names were used and whenever there were multiple matches they were discarded and only unique exact names were retained. Additionally, these names were also matched with the transaction/purchase data of the retailer and duplicate matches were once again eliminated. To be further sure about this process, demographic and residence/location information obtained from the surveys for the respective customers was utilized and this was cross tabulated with the corresponding information available from the firm's transaction data. Specifically, information pertaining to customers' gender, age, and race along with their residence address was used. Only those customers whose demographics and residence location/address matched closely were retained. The process described above was done entirely by the retailer, and we were furnished with the data set containing unique codes assigned to customers upon masking/de-identifying all sensitive information (e.g., names and residence location). The customers' participation in social media can be linked to their transaction data using these unique customer codes. This process was adopted to ensure complete customer confidentiality.

We use data from January 2008 to March 2011 to conduct our empirical analysis. As will be evident in §4, we also need reasonable numbers of customer purchases and sufficient data period post customer participation to reliably estimate our model parameters. Consequently, we do not consider customers who joined the social media site during the last six months of the data time period.[7] The criteria described above left us with 394 customers whom we include for further analysis.

### 3.3. Variable Operationalization

The dependent variable for our analysis is customers' intensity of relationship with the firm. For our purposes, following precedence (e.g., Venkatesan and Kumar 2004) we use visit frequency as a measure for this variable. Accordingly, we analyze if customers' social media participation affects their frequency of visits and operationalize it as the number of times a customer visits the store in a given time period (denoted by *freq*). With respect to operationalization of the independent variables, customers' participation (*CParT*) is a binary variable that takes the value 1 if the customer is a participant in the firm's social media site (e.g., becomes a "fan" on the firm's Facebook page) at the given time period and 0 otherwise.[8] We note that different customers participate in the firm's social media site at different points of time during the estimation period. Social media activity relevant to a focal customer is measured by the number of messages posted by the firm and all other customers. Because recent postings may be more impactful than older ones, we formulate a stock variable of postings (*PostingsStock*) in the following manner (e.g., Tellis and Weiss 1995):

$$PostingsStock_t = \lambda Postings_t + (1-\lambda) PostingsStock_{t-1}. \quad (1)$$

[5] We note that no one particular type of posting dominates.

[6] The survey had 1,249 valid questionnaires returned, representing a response rate of 24.98%.

[7] Because the customer participation period spans the time period August 2009–March 2011, we consider only those customers who joined during the time period August 2009–September 2010.

[8] In our case, customer participation takes the form of becoming a fan of the firm, liking the firm, etc. Our sample customers do not "unparticipate" once they decide to participate/take part in the firm's social media site.

*PostingsStock* is the number of messages posted excluding that of the focal customer at time $t$ and $\lambda \in (0, 1)$ is the decay coefficient.[9] Purchase amount (*PurAmt*) is the spending incurred by the customer across all products in a given time period. Buying focus (*BuyFocus*) represents how broad or narrow the buying pattern is for a particular customer. In our case, because we consider wine purchases, following current literature in this area (Huerta et al. 1998), we use wine color type to capture this variable. The color is a predominant characteristic for wines and serves as an important element for their classification because it depends on the grape variety (Jensen et al. 2008). Wine is classified broadly into three color types: white, red, and rosé. For each customer, during the calibration period,[10] we calculate her spending for each of these wine color types and designate that color as the principal one on which her spending is the most. The buying focus (calculated during the estimation period) then is defined as a dummy variable that takes the value 1 if the customer buys the principal color type at a given time and 0 otherwise.

To calculate the share of premium products (*PremiumShare*), we first determine which products are premium. For this purpose, we calculate the mean price per unit ($/ml) for all the products and plot the distribution. On finding that this is distributed normal, we designate those products as being premium that are one standard deviation above the mean. The share of premium products bought by a focal customer is then obtained as the ratio of the number of premium to the total number of products purchased by the customer. Deal sensitivity (*Deal*) designates how sensitive a customer is toward deals and is defined as the proportion of products the customer buys on promotion, i.e., the number of products bought on promotion divided by the total number of products purchased in a given time period.

We also include in our model several control variables such as customer loyalty and demographics. We follow prior literature and measure loyalty (*Loyalty*) as the share of wallet or the percentage of total customer spending at the focal retailer in a given time period (e.g., Ailawadi et al. 2008). To operationalize this variable, we need to know the total customer spending on wine products across all stores (and not just the focal retailer). In order to compute this, we rely on the survey information wherein customers were specifically asked to report how much they spend on wine products in a typical year. Loyalty is then defined as the ratio of the expenditures at the focal retailer to the (self-reported) total spending incurred on such categories in each time period.[11] In case of demographics, we include the actual age in years (*Age*), the logarithm of customers' income in dollars (*Income*), and dummy variables for *Gender* and *Race* which take the value 1 if the customer is male and white, respectively, and 0 otherwise.

## 4. Self-Selection Issue and Endogeneity

It is possible that customer specific unobserved factors may jointly determine customers' decision to participate in the firm initiated social media site and their intensity of relationship with the firm. For example, customers' positive predisposition toward the firm and its products (beyond customer loyalty that we control for) may simultaneously influence both their willingness to participate in the social media site and their intensity of relationship with the firm. Thus, there could be endogeneity issues arising because of self-selection that need to be considered for accurate identification of the social media participation effect.

To resolve the above issues and establish a causal link between customers' social media participation and their intensity of relationship with the firm, as recommended by current literature (e.g., Huang et al. 2012, Girma and Gorg 2007, Blundell and Dias 2000), we utilize propensity score matching in combination with difference-in-differences analysis. A DID estimator represents the difference in pre and post participation differences between two groups of customers, the treatment group (i.e., customers who participate) and the control group (i.e., customers who do not participate). Note that the difference in the pre and post behavior of participating customers alone may not produce accurate results because of the presence of extraneous factors that might affect behavior (e.g., frequency of visits) in one period but not the other. Thus, a better research design entails using as reference or benchmark customers who do not participate in the firm initiated social media (i.e., the control group). Comparing the behavior of the treatment and control groups in the pre and post participation periods helps to control for the influence of extraneous factors. The challenge then is to compose the control group that comprises customers who are very similar

[9] The decay coefficient $\lambda$ accounts for the possibility that recent messages would matter more than the earlier postings. To avoid computational burden, we do not estimate $\lambda$; instead, we use a grid search procedure and find that 0.83 provides the best fit.

[10] We use 12 months of data prior to January 2008 as the calibration period to calculate/initialize some of the model variables in this case.

[11] In our case, we need to allocate the yearly (self-reported) total expenditures to specific time periods/shopping trips. We use historical spending patterns covering two full years (January 2006–December 2007) prior to our estimation time period to compute average weights, which are then used to assign expenditures, because spending may not be even or consistent across months.

to the treatment group customers (participating customers) in all respects but for the fact that they have not participated in the firm hosted social media, in contrast to the treatment customers who have participated. The difference in behavior between the two groups (i.e., treatment and control) can then be attributed to social media participation.

In order to reliably construct the aforementioned control group, we use PSM. This is because matching through propensity scores substantially improves the similarity between the treatment and the control group because customers who are like treatment customers are given a greater weight for inclusion in the control group, thereby improving the inference related to DID analysis (Stewart and Swaffield 2004, Abadie et al. 2010). In other words, the control group (obtained through PSM) that is used as a reference or comparison in the DID estimation will be as similar as possible to the treatment group on the observed characteristics, which helps to eliminate temporally invariant estimation bias (Smith and Todd 2005). Moreover, PSM helps us mimic a randomized experimental setup (Rubin 2006).

To summarize, we use the PSM procedure to construct a matched pair set of customers, one from the "treatment" group—comprising customers who participate in the firm initiated social media site and whose behaviors we wish to analyze—and another from the "control" group—consisting of customers who are not part of or never participate in the firm's social media site. Subsequently, we employ the DID estimation on the resulting matched sample to estimate the impact of social media participation. We note that using PSM in combination with DID analysis, as documented by prior literature, considerably reduces the bias resulting from both observable and unobservable variables and helps improve the consistency of the estimates (Stewart and Swaffield 2004, Cao et al. 2011). In addition, it improves the reliability of the estimation process (Cao et al. 2011) and is also more efficient as compared to other approaches (e.g., instrument variables) because it obviates the need for strong exclusion restrictions (e.g., Girma and Gorg 2007, Blundell and Dias 2000).

We note that although instrument variable technique is commonly used to correct for selection bias, it is not well suited for our purpose because finding good instruments for social media participation is difficult and using weak instruments may only exacerbate the problem (Bound et al. 1995). Furthermore, even if valid instruments were to be found, the estimation process is still fraught with perils because the social media participation enters as a discrete variable in the analysis, which can result in poorly identified (Wooldridge 2002) and/or inconsistent estimates (Angrist and Pischke 2009).

### 4.1. Propensity Score Matching

The objective of PSM is to select treatment and control group customers who resemble each other in all relevant characteristics before social media participation (the treatment), thereby creating a statistical equivalence between the two groups by balancing them on observables (Rosenbaum and Rubin 1983, Brynjolfsson et al. 2010). For PSM to be feasible, data should be available for both the treatment and control customers *before* the intervention (i.e., social media participation). We utilize 82 weeks of data pertaining to customer characteristics for both groups *prior* to the launch of the social media site. We follow recommendations of recent literature and use the same data sources for both groups (Heckman et al. 1997). We model customer participation decision as a function of customer specific transaction and demographic characteristics.

In the interest of exposition, we discuss the salient points here and provide the detailed description, including the steps we follow for conducting PSM—which are based on current literature (e.g., Caliendo and Kopeinig 2008, Guo and Frasier 2010)—in the (online) appendix (available at http://dx.doi.org/10.1287/isre.1120.0460). We begin by checking if the variables used in the model are balanced between the treatment and control customers using social media participation as the grouping variable through appropriate bivariate tests and include them for further analysis if they are significantly different across the two groups.

We then conduct the propensity score analysis using a logistic model formulation because the program intervention (social media participation) is binary.[12] These results along with the fit statistics are reported in Table 1. We then perform the process of matching the treatment and control customers using the estimated propensity scores. Although many matching algorithms are available, we use the optimal pair matching technique, wherein each customer who participates is matched to another customer who does not participate. This procedure substantially improves the power of the analysis and reduces bias that can occur when multiple (potentially) dissimilar treatment and control customers are matched (Guo and Frasier 2010, Huang et al. 2012). Additionally, optimal pair matching also enables us to treat each distinct matched pair comprising treatment and control customers separately, thus allowing us to control for any residual variability emanating from social media participation effects (Huang et al. 2012). As a result of the above, we are able to satisfactorily match 287 treatment customers to equivalent control customers, all of whom are included for model estimation.

[12] We also fit a probit model for social media participation and the results were similar.

**Table 1** Logit Regression Model of Customers' Social Media Participation

| Variable | Parameter | Std. error |
|---|---|---|
| *PurAmt* | 0.0075*** | 0.0028 |
| *BuyFocus* | −3.0751*** | 0.3975 |
| *Deal* | 0.6113** | 0.2941 |
| *PremiumShare* | 0.0981** | 0.0409 |
| *Loyalty* | 3.4004*** | 1.2452 |
| *Age* | −0.0685*** | 0.0112 |
| *Gender* | 0.2053* | 0.1117 |
| *Income* | 0.1510 | 0.2322 |
| *Race* | 0.4478* | 0.2625 |
| Constant | 3.3035 | 2.7626 |
| Log likelihood | −301.5350 | |

*$p \leq 0.10$, **$p \leq 0.05$, ***$p \leq 0.01$.

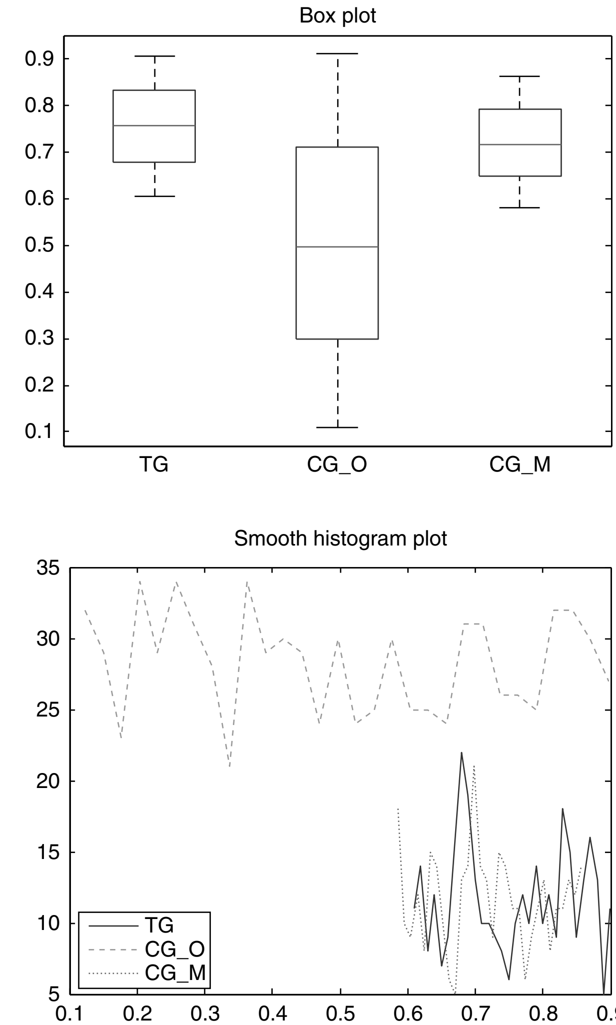
Next, we check to see if the underlying assumptions of the PSM process hold. We check if there is substantial overlap in the characteristics of the customers who participate in the social media and those who do not (i.e., common support condition). For this purpose, as recommended by Lechner (2002), we perform a visual analysis of the propensity score distributions through box-plots and histograms (Figure 1) and find evidence for the existence of common support.

To further ensure the quality of the PSM procedure, we check if the covariates are balanced between the treatment and control groups in the pre and post matching condition, the results of which are reported in Table 2. As can be seen from the table, the variables that are used for PSM are not significantly different across the two groups of customers post matching, implying statistical balance.

Finally, we conduct sensitivity analysis to check for "hidden" bias—the presence of unobserved factors that may affect the assignment into treatment (social media participation) and outcome (intensity of customer-firm relationship) simultaneously (Rosenbaum 2002). Essentially, this involves investigating whether some unobservable factors overly influence the treatment effect of customers' social media participation. The analysis based on the methods expounded in the literature (e.g., Rosenbaum 2002, Becker and Caliendo 2007) shows that "hidden" bias is not a serious concern in our case.

We report the overall summary statistics, which are obtained by averaging the corresponding values across the treatment and the matched control groups. With respect to number of postings, we find that both the firm and its customers actively post messages on the social media site. The average of the stock of the postings variable (see Equation (1)) is 58.23. Although we do not explicitly code the valence of the postings, as mentioned earlier, we find that all of the messages initiated by the retailer are positive and that more than 90% of postings made by the customers are positive in nature. Customers on an average spend \$61.05 per time period. The large amount is a reflection of the nature of category (wine) and also that customers buy multiple items on a trip. Premium selling products constitute 6.69% of the total purchases across the customers. Customers seem to be buying substantial amounts on deal, which could be due to the frequency of promotions that are offered. Moreover, they also exhibit considerable buying focus. Likewise, they are also loyal to the firm (as measured by share of wallet) plausibly because the focal firm is one of the dominant players in the region. As for demographics, we

**Figure 1** Distribution of Propensity Score Before and After Matching




*Notes.* The figure shows the box plots and smooth versions of histograms of propensity score distributions for following groups: TG—Treatment Group, CG_O—Control Group *before* Matching, CG_M—Control Group *after* Matching. Before matching, the distributions for treatment and control groups are quite different; however, after matching they are almost identical, providing evidence of common support. Furthermore, we also conduct a Kolmogorov-Smirnov (KS) test to compare the two distributions. The *p*-value of the KS-test between TG and CG_O is 0.0000 whereas *p*-value of the KS-test between TG and CG_M is 0.9377, which provide evidence of similarity of propensity score distributions between TG and CG_M and hence the existence of common support between the two distributions.

**Table 2 Summary Statistics and Covariate Comparison Before and After Matching**

| | Treatment group | Control group | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Before matching | | | | After matching | | | |
| | Mean | Mean | Mean diff. | *t* stat. | Var. ratio | Mean | Mean diff | *t* stat. | Var. ratio |
| *PurAmt* | 59.5052 | 72.8310 | −13.3259 | −3.0462 | 0.6454 | 62.6005 | −3.0953 | −1.0772 | 0.9442 |
| *BuyFocus* | 0.5504 | 0.7690 | −0.2186 | −10.2615 | 0.8730 | 0.6257 | −0.0753 | −1.2859 | 0.9354 |
| *Deal* | 0.6405 | 0.6744 | −0.0339 | −1.7948 | 1.1000 | 0.6615 | −0.0210 | −0.9731 | 0.9828 |
| *PremiumShare* | 0.0635 | 0.0769 | −0.0134 | −0.5611 | 1.1302 | 0.0701 | −0.0066 | −0.3045 | 1.0298 |
| *Loyalty* | 0.8256 | 0.8353 | −0.0097 | −4.2823 | 0.8025 | 0.8296 | −0.0040 | −1.0153 | 0.9915 |
| *Age* | 50.4550 | 56.1339 | −5.6789 | −7.1329 | 1.0922 | 51.4079 | −0.9529 | −1.6229 | 1.0084 |
| *Gender* | 0.4408 | 0.4017 | 0.0391 | 1.8698 | 1.0306 | 0.4342 | 0.0066 | 0.1152 | 1.0033 |
| *Income* | 10.9990 | 10.9675 | 0.0315 | 1.7515 | 1.0545 | 10.9828 | 0.0162 | 0.2455 | 1.0100 |
| *Race* | 0.8737 | 0.8605 | 0.0132 | 1.6724 | 0.8974 | 0.8715 | 0.0022 | 0.2014 | 0.9857 |

*Notes.* Mean difference (*Mean Diff.*) for each covariate is calculated by subtracting the mean of control group from the mean of treatment group. The *t* statistics (*t stat.*) for these differences in mean are also reported. The variance ratio (Var. *Ratio*) for each covariate is calculated as a ratio of variance of treatment group to variance of control group. The covariates are imbalanced and hence matching would be compulsory if mean difference is away from zero and variance ratio is away from one. The mean differences and variance ratios are close to 0 and 1, respectively, after matching, indicating the improvement in covariate balancing due to matching.

find the mean age to be about 51 years, reflecting the nature of the category (because 21 years is the legal minimum for alcoholic purchases), and the mean income is \$59,334; 43.75% of the sample is male and 87.26% of the customers in the sample are white.

## 5. Econometric Specification

### 5.1. DID Model: Main Effect of Social Media Participation

For each matched pair comprising treatment and control customers (see §4.1), the logarithm of frequency of visits is modeled as[13]

$$\log(freq_{ijt}) = \delta_{0j} + \delta_1 TreatD_{ij} + \delta_2 CParT_{ijt} + \delta_3 TreatD_{ij} \times CParT_{ijt} + \Theta X_{ij} + \xi_{it}. \quad (2)$$

In Equation (2), $i$ denotes a matched pair of customers, $j$ denotes a treatment group or a control group customer, and $t$ denotes the time period. $TreatD_{ij}$ is the treatment dummy variable (that equals 1 if the customer is in the treatment group and 0 if the customer is in the control group), and $CParT_{ijt}$ is a dummy variable denoting social media participation, which takes the value 0 and 1 for periods prior to and post participation, respectively, for customers belonging to the matched pair $i$. $X_{ij}$ represents a vector of demographic and behavioral control variables with $\Theta$ being their corresponding estimated coefficients. $\delta_{0j}$ are the customer-specific fixed effects that capture the differences in baseline relationship intensity and enable us to control for unobserved heterogeneity across customers. Note that in the above formulation consisting of matched treatment and control group customers, we use monthly data that spans both pre and post launch time periods of the social media site (January 2008–March 2011), which results in time-series data that are then stacked for estimation. The main parameter of interest is $\delta_3$, which captures the change in the frequency of visits for treatment customers post participation relative to control customers who do not participate in firm initiated social media.

[13] In the very few cases where the frequency is zero, we add a small value to it (order of 0.001) so as to enable the logarithmic transformation.

### 5.2. DDD Approach: Moderating Effects of Hypothesized Variables

We now describe an alternative version of the model to the one presented earlier (in Equation (2)) that enables us to test our hypotheses discussed in §2. Specifically, we segment customers into high and low levels for each of the moderating variables (i.e., social media activity, purchase amount, buying focus, deal sensitivity, and share of premium products). To accomplish this, consistent with the methods expounded in the current literature (e.g., Tucker et al. 2012), we use a median split based on the values of the above variables and divide them into high and low levels/types. Thus, for a certain variable (e.g., purchase amount) customers will be classified as belonging to the high level if they score greater than the median on that variable and low level otherwise. To make sure that the above process does not confound the estimation time period, we use the calibration period for this purpose.

The above approach allows us to test when the treatment effect (i.e., the social media participation effect) is more effective. Conducting such an analysis may be especially pertinent for managerial decision making because it would help in understanding when and for which customer segments social media efforts yield greater returns.[14] Therefore, to investigate the effect of social media participation at different levels of the moderating variables based on prior literature (e.g., Gruber 1994), we use the following formulation:

$$\begin{aligned}\log(freq_{ijt}) &= \gamma_{0j} + \gamma_1 TreatD_{ij} + \gamma_2 CParT_{ijt} + \gamma_3 TreatD_{ij} \times CParT_{ijt}\\ &\quad + \gamma_4 TreatD_{ij} \times CParT_{ijt} \times PostingsStock_{ij}\\ &\quad + \gamma_5 TreatD_{ij} \times CParT_{ijt} \times PurAmt_{ij} + \gamma_6 TreatD_{ij}\\ &\quad \times CParT_{ijt} \times BuyFocus_{ij} + \gamma_7 TreatD_{ij} \times CParT_{ijt}\\ &\quad \times Deal_{ij} + \gamma_8 TreatD_{ij} \times CParT_{ijt} \times PremiumShare_{ij}\\ &\quad + \Omega X_{ij} + \varepsilon_{it}. \end{aligned} \tag{3}$$

In Equation (3), the variables have the same meaning as in Equation (2).[15] As in the previous case, we include customer specific fixed effects and demographic/behavioral controls. Additionally, the equation includes three way interactions between treatment, social media participation, and each of the moderating variables (they take on the value of 1 for high levels of the variable and 0 otherwise). For instance, $TreatD_{ij} \times CParT_{ijt} \times PurAmt_{ij}$ is the three way interaction between the dummies of treatment, participation, and the high level of purchase amount. Because of the presence of third level interactions in Equation (3), the above framework results in a difference-in-difference-in-differences (DDD) specification. The key parameters of interest are the coefficients associated with the three way interaction terms ($\gamma_4 - \gamma_8$). They capture the effect of social media participation on visit frequency for customers with high level of the corresponding variable (e.g., purchase amount) relative to those customers with a low level of the same variable.

# 6. Results

## 6.1. DID Results for Visit Frequency

We display the raw difference-in-differences values for frequency of visits for the treatment and control groups before and after the social media site was launched by the firm (Table 3). As can be seen from the table, the mean visit frequency increase is not significant for the control group (2.486 post launch versus 2.417 pre-launch), but this increase is significant for the treatment group (2.762 post launch versus 2.540 pre-launch). Moreover, there is no significant difference in the frequency of visits between the treatment and control group customers before the social media site is launched by the firm, whereas the difference is significant post launch. This lends prima facie credence to the fact that participation in the firm initiated social media impacts the behavior of participating consumers. We investigate the above more formally using the DID analysis next.

**Table 3 Difference in Differences: Frequency of Visits of Treatment vs. Control Groups**

| | Group monthly mean | | | |
|---|---|---|---|---|
| | Treatment | Control | Difference | *t*-test |
| After social media launch | 2.7618 | 2.4855 | 0.2763 | 3.4652 |
| Before social media launch | 2.5403 | 2.4169 | 0.1234 | 0.9886 |
| Difference | 0.2215 | 0.0686 | 0.1529 | |
| *t*-test | 3.1764 | 1.5892 | | |

*Notes.* There is no significant difference in frequency of visits between treatment and control groups before the launch of social media; however, the difference becomes significant after the social media launch. Because these differences are taken for the matched sample, the significance of the difference in post launch period can be attributed to customers' social media participation.

To benchmark the results and statistical fit of our DID model, we estimate a series of alternative models (Table 4). Model (1) is the basic DID model without any control variables. As part of Model (2), we include demographic variables as controls, and Model (3) contains behavioral variables in addition to demographics as controls. Model (4) is an augmented version of Model (3) that adds customer specific fixed effects. As is evident from Table 4, the results are substantively similar across the models. The fit statistic ($R^2$) increases from Model (1) to Model (4), which lends face validity to the results.

We find that the parameter corresponding to the treatment effect of social media participation ($TreatD \times CParT$) is consistently positive and significant across the different models (for example, in Model (4) this estimate is 0.0509). This suggests that participation in the firm initiated social media positively influences customers' intensity of relationship with the firm in terms of their frequency of visits. Thus, we find support for H1—the central hypothesis of our study. In order to assess the magnitude of the effect, we compute the elasticity measure using the corresponding parameter estimate from Model (4), the best fitting DID model. We find that elasticity of participation on visit frequency to be 5.204,[16] which implies that customers who participate in the firm's social media visit the firm about 5.2% more frequently compared to customers who do not participate.

[14] We thank the AE for the suggestion.

[15] Although we include the various relevant two way interaction variables in the model, for the sake of exposition, we do not present and discuss the results related to these two way interaction variables. More information is available from the authors upon request.

[16] Because $TreatD \times CParT$ is a dummy variable and the regression equation is semi-logarithmic, the elasticity is given by $100 \times [\exp\{\hat{\delta}_3 - (0.5\mathrm{Var}(\hat{\delta}_3))\} - 1]$, where $\hat{\delta}_3$ is the estimate of $\delta_3$ and $\mathrm{Var}(\hat{\delta}_3)$ is its estimated variance (Kennedy 1981, Halvorsen and Palmquist 1980).

**Table 4 Impact of Social Media Participation on Customer Frequency of Visits**

| | Model (1) No controls | | Model (2) Control for demographic variables | | Model (3) Control for demographic and behavioral variables | | Model (4) Control for demographic and behavioral variables and customer fixed effects | | Model (5) Three-way difference model with customer fixed effects (DDD) | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variables | Estimate | SD | Estimate | SD | Estimate | SD | Estimate | SD | Estimate | SD |
| *TreatD* | 0.0508 | 0.0307 | 0.0383 | 0.0409 | 0.0327 | 0.0243 | 0.0502 | 0.2151 | 0.0342 | 0.2185 |
| *CParT* | 0.0279* | 0.0161 | 0.0853* | 0.0464 | 0.1260* | 0.0714 | 0.1542* | 0.0906 | 0.1541* | 0.0865 |
| *TreatD* × *CParT* | 0.0619*** | 0.0201 | 0.0583*** | 0.0202 | 0.0566*** | 0.0218 | 0.0509*** | 0.0180 | 0.0500*** | 0.0129 |
| *TreatD* × *CParT* × *PostingsStock* | | | | | | | | | 0.0002** | 0.0001 |
| *TreatD* × *CParT* × *PurAmt* | | | | | | | | | 0.0005*** | 0.0002 |
| *TreatD* × *CParT* × *BuyFocus* | | | | | | | | | −0.0405*** | 0.0071 |
| *TreatD* × *CParT* × *Deal* | | | | | | | | | −0.0436*** | 0.0162 |
| *TreatD* × *CParT* × *PremiumShare* | | | | | | | | | 0.0837** | 0.0345 |
| *PostingsStock* | | | | | 0.0016** | 0.0008 | 0.0020** | 0.0008 | 0.0019** | 0.0009 |
| *PurAmt* | | | | | 0.0002*** | 0.0001 | 0.0003*** | 0.0001 | 0.0002*** | 0.0001 |
| *BuyFocus* | | | | | −0.2177 | 0.1490 | −0.0552 | 0.0520 | −0.0533 | 0.0546 |
| *Deal* | | | | | −0.0986 | 0.0676 | −0.0301 | 0.0428 | −0.0376 | 0.0472 |
| *PremiumShare* | | | | | 0.1144** | 0.0537 | 0.0861*** | 0.0286 | 0.0663** | 0.0298 |
| *Loyalty* | | | | | 0.1141** | 0.0478 | 0.0055*** | 0.0019 | 0.0100*** | 0.0024 |
| *Age* | | | 0.0018 | 0.0020 | 0.0019 | 0.0018 | 0.0018 | 0.0115 | 0.0024 | 0.0115 |
| *Gender* | | | −0.0155 | 0.0154 | −0.0103 | 0.0153 | −0.0170 | 0.3743 | −0.0280 | 0.3751 |
| *Income* | | | 0.0021 | 0.0034 | 0.0019 | 0.0034 | 0.0282 | 0.2436 | 0.0049 | 0.3254 |
| *Race* | | | −0.0190 | 0.0425 | −0.0087 | 0.0420 | −0.0404 | 0.3228 | −0.0319 | 0.2444 |
| *Intercept* | 0.8795** | 0.1130 | 0.6022*** | 0.0583 | 0.0192*** | 0.0065 | | | | |
| Customer fixed effects | No | | No | | No | | Yes | | Yes | |
| $R$-squared | 0.1821 | | 0.2359 | | 0.3375 | | 0.4527 | | 0.5545 | |

*$p \le 0.10$, **$p \le 0.05$, ***$p \le 0.01$.

### 6.2. DDD Analysis of High vs. Low Type Customers

We now focus on the results of the DDD model, which is concerned with the effect of social media participation for high and low levels of the moderating variables. We report these results as Model (5) in Table 4. Because the DDD model also includes the two-way interaction of the treatment effect, the results can be interpreted as the *relative* impact of the corresponding level of the moderating variable on visit frequency. H2 suggests that the effect of customer participation on intensity of relationship will be greater for higher levels of activity or message postings in the social media site. We find evidence to support this ($\gamma_4 = 0.0002$; $p < 0.01$). As per hypotheses H3, the impact of customer participation in firms' social media is amplified for customers with high purchase amount levels. H4 suggests that the effect of social media participation is lower for customers with a greater buying focus. We find support for both these hypothesized interaction effects ($\gamma_5 = 0.0005$; $p < 0.01$ for purchase amount and $\gamma_6 = -0.0405$; $p < 0.01$ for buying focus).

We find that the effect of social media participation is greater for customers with lower deal sensitivity ($\gamma_7 = -0.0436$; $p < 0.01$). Thus, we find support for H5. Finally, we find that the effect of social media participation is higher for customers whose share of premium product purchases is greater ($\gamma_8 = 0.0837$; $p < 0.01$), thereby supporting H6. With respect to the control variables, we note that to be consistent with the DDD analysis, we report the results when the variables were used in levels (similar results were obtained when actual values of the variables were used). We find that higher level of activity in the social media site is associated with greater intensity of relationship between the firm and its customers. Moreover, customers with higher levels of spending and greater shares of premium products visit the firm more frequently, as do more loyal customers. We find no significant effects for buying focus, deal sensitivity, and the demographic variables (i.e., age, gender and race) plausibly because of better covariate balance that was achieved for these variables.

## 7. Economic Impact of Customer Social Media Participation

### 7.1. Profitability Analysis

In this section, we analyze the benefits that can be realized by firms as a result of their foray into social media. Specifically, we begin by documenting the incremental profits of customers who participate in the social media (i.e., treatment group customers) vis-à-vis those who do not participate (i.e., control group customers). For this purpose, we compute the

**Table 5 DID: Profitability of Treatment vs. Control Group**

| | Group monthly mean ($) | | | |
|---|---|---|---|---|
| | Treatment | Control | Difference | *t*-test |
| After social media launch | 25.2607 | 23.6989 | 1.5618 | 3.3972 |
| Before social media launch | 22.2549 | 22.0354 | 0.2195 | 1.2184 |
| Difference | 3.0058 | 1.6635 | 1.3423 | |
| *t*-test | 2.9672 | 1.5924 | | |

*Notes.* There is no significant difference in total customer profits (in dollars per month) between treatment and control group before the launch of social media site; however the difference becomes significant post launch. Because these differences are taken for the matched sample, the significance of the difference in post period can be attributed to the customers' social media participation.

aggregated per period net total profits separately for customers in both groups before and after the launch of the social media site by the firm, the results of which are reported in Table 5. Note that this is possible in our case because we have data pertaining to both retailer prices and costs at the individual product level. The values reported in the table are the mean total profits across all the products purchased by the respective group of customers. On comparing the table values, we once again find that there is a substantial increase in total profits for the treatment group after the launch of the social media site ($25.261 pre launch versus $22.255 post launch). In contrast, there is no significant difference in total profits before ($22.035) and after ($23.698) the launch of the social media site for the control group. This suggests that customer participation in social media improves firm profitability.

To investigate the above issue more formally, we utilize the DID framework of Equation (2) with customer profits as the dependent variable and report the results in Table 6. Like earlier, we estimate a series of models ranging from the basic DID model sans any control variables (Model 1) to the one that includes both behavioral and demographic controls along with customer fixed effects (Model 4). We once again find that the parameter related to the treatment effect of social media participation (*TreatD* × *CParT*) is positive and significant across all the different model specifications. For instance, in Model (4) the parameter estimate is 0.0550, which corresponds to profit elasticity of 0.0563. This implies that customers who participate in the firm's social media contribute more to the firm's bottom line. We also estimate the DDD model that includes the three way interaction effects between the treatment variable, customer social media participation, and the indicator variable for the high level of the moderating variables (Equation (3)). We find that the results have the same directionality and similar significance as reported for visit frequency

**Table 6 Impact of Social Media Participation on Customer Profitability**

| | Model (1) No controls | | Model (2) Control for demographic variables | | Model (3) Control for demographic and behavioral variables | | Model (4) Control for demographic and behavioral variables and customer fixed effects | | Model (5) Three-way difference model with customer fixed effects (DDD) | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variables | Estimate | SD | Estimate | SD | Estimate | SD | Estimate | SD | Estimate | SD |
| *TreatD* | 0.0098 | 0.0482 | 0.0108 | 0.0546 | 0.0609 | 0.0479 | 0.0247 | 0.2572 | 0.0927 | 0.2811 |
| *CParT* | 0.0726* | 0.0394 | 0.0878* | 0.0462 | 0.0868* | 0.0235 | 0.0225* | 0.0125 | 0.0334* | 0.0187 |
| *TreatD* × *CParT* | 0.0578*** | 0.0318 | 0.0569*** | 0.0183 | 0.0561*** | 0.0187 | 0.0550*** | 0.0213 | 0.0505*** | 0.0186 |
| *TreatD* × *CParT* × *PostingsStock* | | | | | | | | | 0.0007** | 0.0003 |
| *TreatD* × *CParT* × *PurAmt* | | | | | | | | | 0.0003*** | 0.0001 |
| *TreatD* × *CParT* × *BuyFocus* | | | | | | | | | −0.0170*** | 0.0060 |
| *TreatD* × *CParT* × *Deal* | | | | | | | | | −0.0110** | 0.0052 |
| *TreatD* × *CParT* × *PremiumShare* | | | | | | | | | 0.0315*** | 0.0109 |
| *PostingsStock* | | | | | 0.0003** | 0.0001 | 0.0004** | 0.0002 | 0.0009* | 0.0005 |
| *PurAmt* | | | | | 0.0027*** | 0.0003 | 0.0028*** | 0.0003 | 0.0027*** | 0.0003 |
| *BuyFocus* | | | | | −0.0975 | 0.0632 | −0.1625 | 0.1221 | −0.1278 | 0.0793 |
| *Deal* | | | | | −0.1005 | 0.0931 | −0.1491 | 0.1012 | −0.1653 | 0.1096 |
| *PremiumShare* | | | | | 0.0538*** | 0.0107 | 0.0389** | 0.0179 | 0.0880*** | 0.0129 |
| *Loyalty* | | | | | 2.2030*** | 0.0548 | 2.1367*** | 0.0640 | 2.1304*** | 0.0641 |
| *Age* | | | 0.0003 | 0.0017 | 0.0001 | 0.0009 | 0.0221 | 0.0137 | 0.0213 | 0.0137 |
| *Gender* | | | 0.1026 | 0.0947 | 0.0087 | 0.0175 | 0.9551 | 0.6475 | 0.9593 | 0.6481 |
| *Income* | | | 0.0077 | 0.0356 | 0.0247 | 0.0179 | 0.0585 | 0.3859 | 0.088 | 0.3887 |
| *Race* | | | 0.1769 | 0.1258 | 0.0187 | 0.0482 | 0.4572 | 0.2912 | 0.5013 | 0.4920 |
| *Intercept* | 3.0928*** | 0.9891 | 2.8112*** | 0.8107 | 1.9887*** | 0.2094 | | | | |
| Customer fixed effects | No | | No | | No | | Yes | | Yes | |
| *R*-squared | 0.2168 | | 0.2864 | | 0.4526 | | 0.5407 | | 0.6246 | |

$^{*}p \le 0.10$, $^{**}p \le 0.05$, $^{***}p \le 0.01$.

(i.e., Model 5 in Table 4). Thus, the above analysis establishes that social media participation has a positive impact on profitability (a key metric for firms) in a manner akin to intensity of relationship as measured by visit frequency.

### 7.2. Segment Level Analysis

The results from the DDD analysis establish that the relative impact of social media participation differs across different customer segments (i.e., high versus low type). However, from a managerial perspective, it might be more interesting to understand the actual magnitude of the treatment effect for the two segments. In other words, managers may be interested in understanding not only whether the impact of social media is greater for high type versus low type customers but also the absolute magnitude of the impact for each of these segments. This kind of segment level analysis can help in understanding which segments are more lucrative and should be allocated greater marketing dollars. To investigate this issue, we divide the sample data set into two subsamples, with one consisting of customers with high levels and the other with low levels of the moderating variable, respectively (we do this for each of the moderating variables). We utilize the same process for dividing customers into high and low levels/types as was used for the DDD analysis (i.e., median split analysis as described earlier in the paper). We then conduct the DID analysis for each customer segment to distinguish the impact of social media participation at the segment level for both visit frequency as well as profits (Model 4 formulation as presented in Equation (2)).

We provide the results of the above analysis in Table 7. In the interest of exposition, we present the main results concerning the treatment effect only. We find that the effect of participation for the high segment is positive and significant in case of frequency of visits for all the moderating variables. Interestingly, we also find that the impact of social media participation is still positive and significant for low segment customers. In other words, social media participation influences the behavior of not only high level/segment customers but also those belonging to the low level/segment with respect to their intensity of relationship with the firm. We obtain similar results for the customer profitability model as well. We also represent these results pictorially in Figure 2, wherein the four panels succinctly establish that although social media participation matters for both high and low type customers, the differences between the two are more pronounced for certain moderating variables. Such an exercise can be useful in quantifying the benefits that can accrue to firms as a result of their social media initiatives at the individual customer segment level and can also help discern the relative impact for each variable.

## 8. Robustness Checks

### 8.1. Inclusion of Additional Variables for Matching

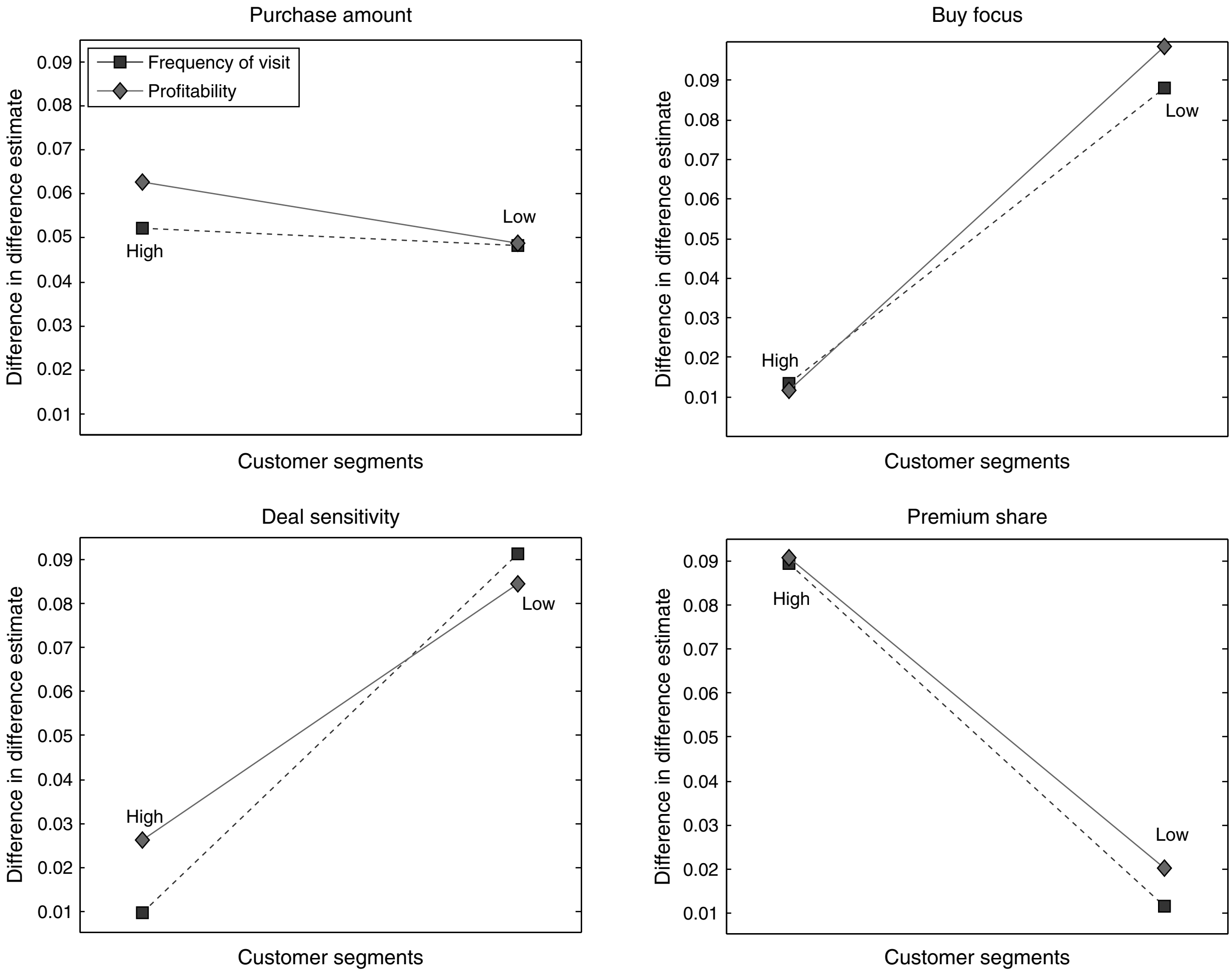
In addition to the original nine variables that are used for the main analysis, we check the validity of our results by using an augmented set comprising three additional variables: Internet risk aversion, technology comfortness, and social networking attitude. We choose to include these variables, which are obtained from the survey for both participants and nonparticipants, because they may be related to customers' propensity to participate in firm initiated social media activities. For example, customers who score high on social networking attitude may be more likely to join the firm's social media. The pertinent details including a description of the relevant variables and sources used to construct them are provided in the online appendix. Note that we do not include the above mentioned variables in the main analysis because it is recommended that they be collected prior to customers' social media participation (Caliendo and Kopeinig 2008), whereas the survey (that is the source of these variables) was administered at a later date. We find that the results pertaining to the effect of social media participation using the

**Table 7 DID Estimates for Segment (High vs. Low) Level Analysis**

| | Postings stock | | Purchase amount | | Buying focus | | Deal sensitivity | | Premium share | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | High | Low | High | Low | High | Low | High | Low | High | Low |
| | | | | | Frequency of visits | | | | | |
| *TreatD* | 0.0718 | 0.0323 | 0.0612 | 0.0403 | 0.0104 | 0.0896 | 0.0078 | 0.0921 | 0.0912 | 0.0098 |
| *CParT* | 0.1884 | 0.1186 | 0.2216* | 0.0834* | 0.1219* | 0.1867* | 0.1383* | 0.1843* | 0.2154* | 0.0932* |
| *TreatD* × *CParT* | 0.0642*** | 0.0463** | 0.0523*** | 0.0484*** | 0.0135*** | 0.0881*** | 0.0099*** | 0.0911*** | 0.0894*** | 0.0116*** |
| | | | | | Profitability | | | | | |
| *TreatD* | 0.0307 | 0.0162 | 0.0462 | 0.0008 | 0.0026 | 0.0446 | 0.0099 | 0.0325 | 0.0492 | 0.0014 |
| *CParT* | 0.0318 | 0.0133 | 0.0403* | 0.0052* | 0.0028* | 0.0427* | 0.0121* | 0.0328* | 0.0428* | 0.0021* |
| *TreatD* × *CParT* | 0.0787*** | 0.0294*** | 0.0628*** | 0.0487*** | 0.0116*** | 0.0987*** | 0.0262*** | 0.0844*** | 0.0908*** | 0.0203*** |

*$p \leq 0.10$, **$p \leq 0.05$, ***$p \leq 0.01$.

**Figure 2** **DID Estimate Plots for Segment Analysis (High vs. Low Groups)**




augmented set are substantively similar to the results that we presented earlier. We summarize these results in the appendix.

### 8.2. Matching with Alternative Techniques

As mentioned previously in §4.2, we use the optimal pair matching for our analysis. However, we also verify the robustness of our results by employing alternative matching methods. Specifically, we use greedy matching (nearest neighborhood without calipers) and Mahalanobis distance (Dehejia and Wahba 2002) and find that the results with these alternative matching techniques are very similar to those obtained from the optimal pair matching method. We provide some pertinent results of the above alternative matching techniques in the online appendix.

### 8.3. Robustness of the DID Analysis

We conduct a series of robustness checks to confirm the results of the DID analysis. For our case, we use data consisting of 18 months pre and post launch of the social media site by the firm. To make sure that the results are not because of some unobservable factors present during certain periods and are robust to the specification of time windows, we conduct the following analysis. We run the DID models (for both visit frequency and profitability) wherein we restrict the post treatment period (i.e., data post social media launch) to a time frame consisting of 3, 6, 9, 12, and 15 months, the results of which are reported in Table 8. We find that the DID estimates are positive and significant in each of the above cases, indicating that customer participation in social media positively impacts visit frequency and profitability. Moreover, the parameter estimate of the treatment effect gradually converges with the estimate at 15 months being close to that obtained overall (i.e., using 18 months). For illustration purposes, we also plot the DID estimates obtained from the different time windows in the above analysis along with the 95% confidence interval in Figure 3. As can be seen from the graph, as expected, these estimates seem to be stabilizing with time.

To further check that the results of our analysis are not simply a consequence of spurious correlation

**Table 8 Results of DID Analysis with Different Time Windows for Treatment Periods**

| | Treatment window (post launch period) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 15 months | | 12 months | | 9 months | | 6 months | | 3 months | |
| Variable | Visits | Profitability | Visits | Profitability | Visits | Profitability | Visits | Profitability | Visits | Profitability |
| *TreatD* | 0.0297 | 0.0484 | 0.0298* | 0.0485* | 0.0302* | 0.0488* | 0.0305** | 0.0492** | 0.0317*** | 0.0502** |
| | (0.0223) | (0.0320) | (0.0173) | (0.0289) | (0.0163) | (0.0261) | (0.0123) | (0.0221) | (0.0023) | (0.0221) |
| *CParT* | 0.0572* | 0.0281* | 0.0702* | 0.0440* | 0.0768* | 0.0523* | 0.0832* | 0.0594* | 0.0877 | 0.0655 |
| | (0.0322) | (0.0170) | (0.0401) | (0.0229) | (0.0421) | (0.0299) | (0.0421) | (0.0319) | (0.0621) | (0.0519) |
| *TreatD* × *CParT* | 0.0501*** | 0.0534*** | 0.0478*** | 0.0498*** | 0.0439*** | 0.0457*** | 0.0371*** | 0.0391*** | 0.0267*** | 0.0283*** |
| | (0.0032) | (0.0028) | (0.0033) | (0.0030) | (0.0035) | (0.0032) | (0.0039) | (0.0035) | (0.0048) | (0.0044) |

$^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$.

and/or due to model misspecification, we run a series of "phantom" regressions (Chetty et al. 2009). For these regressions, we consider only the data period (18 months) prior to the actual launch of the social media site by the firm (i.e., from January 2008 to August 2009) and designate a random month from this period as the time of the launch of the site when customers could participate. For example, if we designate the end of the ninth month as the time when the social media site went into effect, then months 10–18 constitute the "treatment" period and months 1–9 make up the "nontreatment" period. We run a series of DID regressions with different time periods (third, sixth, and ninth month) randomly designated as the month when the social media site was launched, the results of which are reported in Table 9. We find that parameter estimate corresponding to the treatment effects is statistically indistinguishable from zero in all the cases. The above analysis with "placebo" estimates suggests that our results can be attributed to customers' social media participation.

Because our data consist observations over time, a potential concern is the presence of serial correlation. To make sure that the above does not pose a large concern for the model results, we adopt the following method based on Bertrand et al. (2004). We aggregate the data set, by ignoring the time series aspect of it, so that we end up with two distinct data points—one each for pre and post participation periods—for each customer in the matched pair. On reestimating the DID models using this aggregated data set, we find that our results are substantially similar (Table 10). We additionally replicate our regressions using Newey-West corrected standard errors and also perform Dickey-Fuller tests to make sure that our dependent and independent variables are stationary. Finally, we also reestimate our DID models with random effects in lieu of customer specific fixed effects and find that the results are very similar.

### 8.4. Cohort Effect

It is plausible that early participants affect the behavior of later participants. In our data set, because we observe when the customers become participants of the firm initiated social media site, we classify them into their respective cohorts (early or late participants)

**Figure 3 95% Confidence Interval for DID Estimates with Different Treatment Windows**

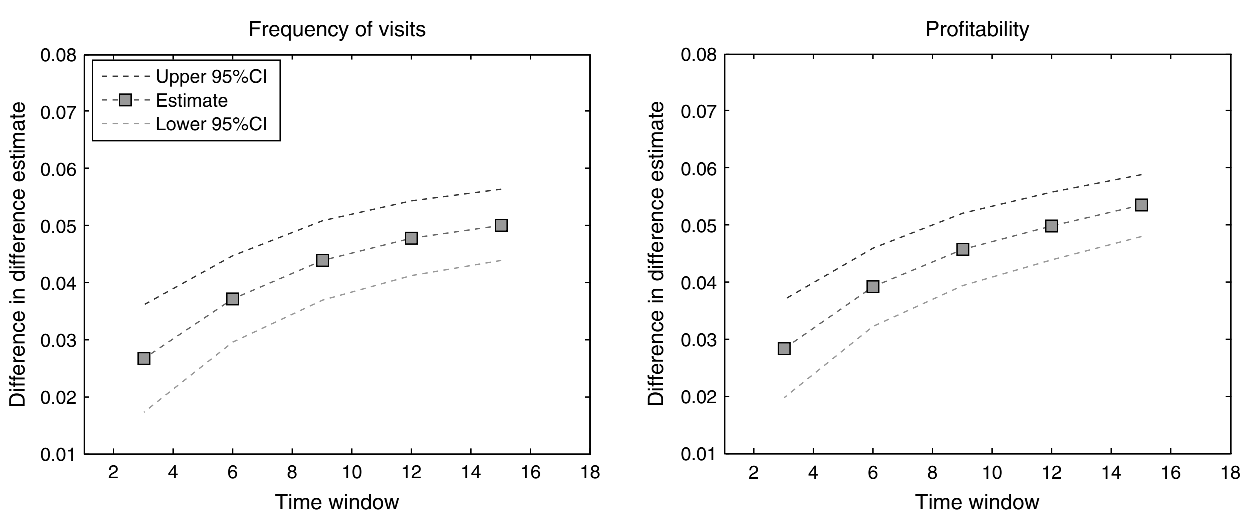


**Table 9 DID Analysis for "Phantom" Regressions**

| | "Phantom" treatment window | | | | | |
|---|---|---|---|---|---|---|
| | 3 months | | 6 months | | 9 months | |
| Variable | Visits | Profitability | Visits | Profitability | Visits | Profitability |
| *TreatD* | −0.0049 | 0.0050 | −0.0023 | 0.0053 | −0.0014 | 0.0023 |
| | (0.0039) | (0.0036) | (0.0032) | (0.0050) | (0.0030) | (0.0027) |
| *CParT* | 0.0489 | 0.0189 | 0.0486 | 0.0195 | 0.0481** | 0.0184*** |
| | (0.0323) | (0.0130) | (0.0324) | (0.0138) | (0.0226) | (0.0024) |
| *TreatD* × *CParT* | 0.0029 | 0.0078 | 0.0029 | 0.0068 | 0.0033 | 0.0020 |
| | (0.0055) | (0.0051) | (0.0046) | (0.0042) | (0.0042) | (0.0039) |

$^{*}p \le 0.10$, $^{**}p \le 0.05$, $^{***}p \le 0.01$.

**Table 10 Aggregate DID Results**

| | Visit frequency | | Profitability | |
|---|---|---|---|---|
| Variable | Estimate | SD | Estimate | SD |
| *TreatD* | 0.0002 | 0.0020 | 0.0001 | 0.0019 |
| *CParT* | 0.0498* | 0.0280 | 0.0178* | 0.0099 |
| *TreatD* × *CParT* | 0.0049*** | 0.0018 | 0.0069*** | 0.0026 |
| *PostingsStock* | 0.0012** | 0.0006 | 0.0002** | 0.0001 |
| *PurAmt* | 0.0002*** | 3.33E–05 | 0.0001*** | 2.13E–05 |
| *BuyFocus* | −0.0044 | 0.0054 | −0.0128** | 0.0050 |
| *Deal* | −0.0092* | 0.0050 | −0.0016 | 0.0047 |
| *PremiumShare* | 0.0753*** | 0.0264 | 0.1049** | 0.0518 |
| *Loyalty* | 0.0009*** | 0.0003 | 0.0304*** | 0.0116 |
| *Age* | 0.0001 | 0.0001 | 0.0009 | 0.0008 |
| *Gender* | −0.0030 | 0.0035 | 0.0042 | 0.0034 |
| *Income* | 0.0002 | 0.0002 | 0.0004 | 0.0002 |
| *Race* | −0.0021 | 0.0022 | 0.0024 | 0.0020 |
| Customer fixed effects | Yes | | Yes | |
| *R*-squared | 0.3557 | | 0.4694 | |

$^{*}p \le 0.10$, $^{**}p \le 0.05$, $^{***}p \le 0.01$.

using a time based median split. We then reestimate our models separately for each cohort and find the substantive results to be similar.

### 8.5. Alternative Operationalization of Variables

We check the sensitivity of the model results to different operationalization of the variables. Specifically, we use the total quantity purchased by customers for purchase amount. Similarly, we considered two standard deviations above the mean as the cut-off for customers' purchase of premium share products. We also considered transformation of the pertinent variables such as logarithm of purchase amount, age, and postings. We find that the results are very similar in all the above cases.

## 9. Discussion, Conclusion, and Directions for Future Research

In this paper, we study the impact of customer participation in a firm's social media site on an important indicator of the intensity of customer-firm relationship, customer shopping visit, or purchase frequency. We further examine the moderating roles of social media activity and customer characteristics. In addition, we quantify the effect of customer social media participation on a key metric that contributes to firms' bottom line, i.e., customer profitability. We draw on CRM and social media literatures to formulate our hypotheses and test them using a novel individual level transaction data set that is complemented with data on social media participation for the *same* set of customers. We rely on a quasi-experimental setup that uses PSM in combination with DID analysis to rule out customer self-selection issues concerning social media participation and the intensity of customer-firm relationship as measured by frequency of visits. Our results show that customer engagement through social media increases customers' shopping visits. Our results also suggest that this effect is greater for a greater level of activity in the firm hosted social media site. Furthermore, we find that the effect is greater for customers who have higher levels of spending and share of premium products and lower levels of buying focus and deal sensitivity. We find that these results hold for customer profitability as well.

### 9.1. Theoretical Implications

Our results extend findings from social media and CRM literatures while also contributing to them in several ways. First, although many studies have attempted to analyze the impact of online social media, they have primarily considered the influence of online reviews on product sales (Chevalier and Mayzlin 2006) or new product diffusion (Susarla et al. 2012). Our study adds to this burgeoning literature by analyzing the role of social media in driving business value for firms by linking customers' social media participation to the additional value created for the firm.

Second, we view a firm's social media efforts as an extension of its relationship building efforts with its customers. A significant stream of literature in CRM attempts to understand the drivers of customer equity and bring accountability to relationship management expenditures (Rust et al. 2004). This research has been primarily limited to studying the impact of firms' offline relationship building expenditures and customer characteristics. We add to this research and provide insights into how firms' online social media activities may influence customer visit (or purchase) frequency, a key driver of customer lifetime value (Venkatesan and Kumar 2004).

Third, our study takes a holistic look at social media participation and adds new insights to growing research that considers the impact of community participation on customer behavior. For example, Dholakia and Durham (2010) conduct surveys and find that customers who became fans of a firm on Facebook reported increased spending with the firm as compared to non-fans. Although these findings are interesting, the study relies mainly on self-reported data and does not establish a causal connection between Facebook participation and customer behavior. Another study, that of Algesheimer et al. (2010), reports that participation in the eBay community leads to a decline in the amount of money spent with the firm. Although this research establishes the effect of participation in an online community, these findings do not help understand how firms' social media efforts can transform a customer-firm relationship, which is the focus of this study. Furthermore, our study also helps tie customers' social media participation to their transaction relationship with the firm and thus provides managerial guidelines for better decision making. Finally, the few studies that examine the impact of social media tend to be aggregate in nature (e.g., Mudambi and Schuff 2010). Although such studies help provide an overview, analyzing the above phenomenon at the individual level and adopting a customer centric view enriches the extant social media literature while providing novel insights into behavior subsequent to customer social media participation.

## 9.2. Managerial Implications

The results of our study yield several important managerial implications. As customers spend more time on social media sites, managers are devoting more resources to them. However, in turbulent times when firms are under increased pressure to justify all their promotional budgets, social media spending needs to be more financially accountable. Thus, our research helps managers gain a better understanding of return on investment accruing to social media initiatives. Based on our findings, we offer the following recommendations for practice.

*Nurture Customer Relationships Through Social Media.* Our results show that firms' online social media activities help strengthen the bond between the customer and the firm and contribute to financial performance in the long run by increasing customer visit frequencies and profitability. Although our findings suggest a positive effect of customer social media participation on customer visit frequency and profitability, managers must be cautious and not interpret this result to imply that simply creating, say, a Facebook page and inviting customers to become fans will lead to positive customer outcomes. They must carefully manage and devise opportunities to create and nurture relationships with customers through the social medium. For example, maintaining a user-friendly social media site interface, providing regular updates about events, sending personalized messages to individual customers, and encouraging member contributions can create interactive communication that can enhance firm equity (Agarwal et al. 2008).

*Be Wary; Not All Customers are Created Equal.* Our results also provide actionable recommendations for managers in terms of segmenting and targeting customers based on customer purchase histories and interactions with the firm. Our results suggest that managers need to understand that not all customers will have a similar response to the firm's social media efforts. It is vital that managers integrate their knowledge about customers from both offline transactions and online social media sources in order to better serve them.

*Create Product Based Social Media Forums.* Besides finding a differential response of social media participation across different types of customers, we also find an interaction effect between social media participation and customers' shopping baskets. Thus, managers can increase the response to social media participation by creating specialized subcommunities or discussion forums for customers looking for premium and unique products.

In summary, we believe the results of the study provide unique theoretical and managerial insights into the relationship between a firm's social media efforts and firm performance. However, the study suffers from some limitations that can be addressed by future research. First, we are able to provide results from only one firm. It would be interesting to study whether the results are generalizable for other companies. Furthermore, checking for generalizability of results across different forms of social media would also be an interesting addition to this research. Second, although we studied the impact of social media postings, due to data limitations we were not able to examine how the valence of messages affects customers' purchase behavior. In particular, we note that most of the message postings in our data are positive, which limits us from studying the impact of valence of messages. Next, future research could also extend the current findings by studying whether different types of message postings by firm (e.g., product or price related) will elicit a different response from customers. Finally, the results of this study also suffer from the limitation that they are specific to the social media strategy employed by the firm. Future research can address this by examining the impact of different social media strategies across multiple, disparate firms.

## Electronic Companion

An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/isre.1120.0460.

## Acknowledgments

The authors thank Professor Arun Jain and the Research Group in Integrated Marketing (RIM) at the School of

Management, SUNY-Buffalo, for accesses to the data and research assistance. They thank the participants of the *ISR* special issue workshop at University of Maryland and the INFORMS 2012 Marketing Science Conference. They also thank the Senior Editor, the Associate Editor, and the two anonymous reviewers for valuable comments and suggestions. The authors contributed equally to the article and the author ordering is reverse alphabetical. All errors are the authors' own.

## References

Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco Control Program. *J. Amer. Statist. Assoc.* 105(490):493–505.

Agarwal R, Gupta AK, Kraut R (2008) The interplay between digital and social networks. *Inform. Systems Res.* 19(3):243–252.

Agarwal R, Animesh A, Prasad K (2009) Social interactions and the "digital divide": Explaining variations in Internet use. *Inform. Systems Res.* 20(2):277–294.

Ailawadi KL, Neslin SA, Gedenk K (2001) Pursuing the value-conscious consumer: Store brands versus national brand promotions. *J. Marketing* 65(1):71–89.

Ailawadi KL, Pauwels K, Steenkamp J-BEM (2008) Private-label use and store loyalty. *J. Marketing* 72(6):9–30.

Algesheimer R, Dholakia UM, Hermann A (2005) The social influence of brand community: Evidence from European car clubs. *J. Marketing* 69(3):19–34.

Algesheimer R, Borle S, Dholakia U, Singh SS (2010) Impact of customer community participation on customer behaviors: An empirical investigation. *Marketing Sci.* 29(4):756–769.

Anderson EW, Sullivan MW (1993) The antecedents and consequences of customer satisfaction for firms. *Marketing Sci.* 12(2):125–143.

Angrist JD, Pischke JS (2009) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ).

Becker SO, Caliendo M (2007) Sensitivity analysis for average treatment effect. *Stata J.* 7(1):71–83.

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *Quart. J. Econom.* 119(1):249–275.

Blundell R, Dias MC (2000) Evaluation methods for non-experimental data. *Fiscal Stud.* 21(4):427–468.

Bolton RN (1998) A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Sci.* 17(1):45–65.

Bound J, Jaeger D, Baker R (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J. Amer. Statist. Assoc.* 90(430):443–450.

Brynjolfsson E, Hu Y(J), Smith MD (2010) The longer tail: The changing shape of Amazon's sales distribution curve. http://dx.doi.org/10.2139/ssrn.1679991.

Caliendo M, Kopeinig S (2008) Some practical guidance for the implementation of propensity score matching. *J. Econom. Surveys* 22(1):31–72.

Cao Z, Song X, Hamilton L (2011) Difference-in-difference and instrumental variables approaches: An alternative and complement to propensity score matching in estimating treatment effects. Thomson Reuters; CER Issue Brief, http://healthcare.thomsonreuters.com/cer/assets/DifferenceCERIssueBrief.pdf.

Chetty R, Looney A, Kroft K (2009) Salience and taxation: Theory and evidence. *Amer. Econom. Rev.* 99(4):1145–1177

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* 43(August): 345–354.

Crosby LA, Evans KR, Cowles D (1990) Relationship quality in services selling: An interpersonal influence perspective. *J. Marketing* 54(3):68–81.

Dehejia RH, Wahba S (1999) Causal effects in non-experimental studies: Reevaluating the evaluation of training programs. *J. Amer. Statist. Assoc.* 94(448):1053–1062.

Dehejia RH, Wahba S (2002) Propensity score-matching methods for non-experimental causal studies. *Rev. Econom. Statist.* 84(1):151–161.

Dellarocas C, Zhang X, Awad NF (2007) Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interactive Marketing*. 21(4):23–45.

Dholakia UM, Durham E (2010) One café chain's Facebook experiment. *Harvard Bus. Rev.* 88(3):26.

DiPrete T, Gangl M (2004) Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociol. Methodology* 34(1):271–310.

Forman C, Ghose A, Wiesenfeld B (2008) Examining the relationship between reviews and sales: The role of reviewer identity discloser in electronic markets. *Inform. Systems Res.* 19(3): 291–313.

Forrester Research (2008) *The Growth of Social Technology Adoption* (Forrester Research, Cambridge, MA).

Girma S, Gorg H (2007) Evaluating the foreign ownership wage premium using a difference-in-differences matching approach. *J. Internat. Econom.* 72(1):97–112.

Godes D, Mayzlin D (2004) Using online conversations to study word-of-mouth communication. *Marketing Sci.* 23(4):545–560.

Gruber J (1994) The consumption smoothing benefits of unemployment insurance. *Amer. Econom. Rev.* 87(1):192–205

Guo S, Frasier MW (2010) *Propensity Score Analysis: Statistical Methods and Applications* (Sage, Menlo Park, CA).

Halvorsen R, Palmquist R (1980) The interpretation of dummy variables in semilogarithmic equations. *Amer. Econom. Rev.* 70(3):474–475.

Heckman J, Ichimura H, Todd P (1997) Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Rev. Econom. Stud.* 64(4):605–654.

Ho SY, Bodoff D, Tam KY (2011) Timing of adaptive Web personalization and its effects on online consumer behavior. *Inform. Systems Res.* 22(3):660–679.

Huang Q, Nijs VR, Hansen K, Anderson ET (2012) Wal-Mart's impact on supplier profit. *J. Marketing Res.* 49(2):131–143.

Huerta MRS, Masoud T, Alonso GL (1998) Wine differentiation according to color using conventional parameters and volatile components. *J. Food Composition Anal.* 11(4):363–374.

IBM Report (2011) 11 Key Marketing Trends for 2011. Unica/IBM, New York, NY.

Janiszewski C (1993) Pre-attentive mere exposure effects. *J. Consumer Res.* 20(3):376–392.

Jensen JS, Demiray S, Egego M, Meyer AS (2008) Prediction of wine color attributes from the phenolic profiles of red grapes (Vitis vinifera). *J. Agricultural Food Chemistry* 56(3):1105–1115.

Kennedy BW (1981) Variance component estimation and prediction of breeding values. *Canadian J. Genetics Cytology* 23(4):565–578.

Kietzmann JH, Hermkens K, McCarthy IP, Silvestre BS (2011) Social media? Get serious! Understanding the functional building blocks of social media. *Bus. Horizons* 54(3):241–251.

Lechner M (2002) Some practical issues in the evaluation of heterogeneous labor market programmes by matching methods. *J. Royal Statist. Soc. A* 165(1):59–82.

McAlexander JH, Schouten JW, Koenig HF (2002) Building brand community. *J. Marketing* 66(1):38–54.

Morgan RM, Hunt SD (1994) The commitment-trust theory of relationship marketing. *J. Marketing* 58(3):20–38.

Mudambi SM, Schuff D (2010) What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quart.* 34(1):185–200.

Pavlou P, Liang H, Xue Y (2007) Uncertainty and mitigating uncertainty in online exchange relationships: A principal–agent perspective. *MIS Quart.* 31(1):105–131.

Reinartz WJ, Kumar V (2000) On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *J. Marketing* 64(4):17–35.

Reinartz WJ, Kumar V (2003) The impact of customer relationship characteristics on profitable lifetime duration. *J. Marketing*. 67(1):77–99.

Rigby DK, Reichheld FR, Schefter P (2002) Avoid the four perils of CRM. *Harvard Bus. Rev.* 7(2):101–109.

Rosenbaum PR (2002) *Observational Studies* (Springer, New York).

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55.

Rubin DB (2006) *Matched Sampling for Causal Effects* (Cambridge University Press, New York).

Rust RT, Lemon KN, Zeithaml VA (2004) Return on marketing: Using customer equity to focus marketing strategy. *J. Marketing* 68(1):109–127.

Smith JA, Todd PE (2005) Does matching overcome LaLonde's critique of nonexperimental estimators? *J. Econometrics* 125(1–2): 305–353.

Stewart MB, Swaffield JK (2004) The other margin: Do minimum wages cause working hours adjustments for low-wage workers? Working paper, University of Warwick, UK.

Susarla A, Oh J-H, Tan Y (2012) Social networks and the diffusion of user-generated content: Evidence from YouTube. *Inform. Systems Res.* 23(1):23–41.

Tellis GJ, Weiss D (1995) Does TV advertising really affect sales? *J. Advertising* 24(3):1–12.

Tucker C, Zhang J, Zhu T (2012) Days on market and home sales. Working paper. MIT Sloan School of Management, Cambridge, MA.

Venkatesan R, Kumar V (2004) A customer lifetime value framework for customer selection and resource allocation strategy. *J. Marketing* 68(1):106–125.

Verhoef PC, Donkers B (2005) The effect of acquisition channels on customer loyalty and cross-buying. *J. Interactive Marketing* 19(2):31–43.

Verhoef PC, Franses PH, Hoekstra JC (2001) The impact of satisfaction and payment equity on cross-buying: A dynamic model for a multi-service provider. *J. Retailing* 77(3):359–378.

Wangenheim FV, Bayón T (2007) The chain from customer satisfaction via word-of-mouth referrals to new customer acquisition. *J. Acad. Marketing Sci.* 35(2):233–249.

Wooldridge JM (2002) *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA).

ZDNet (2011) Two thirds of small business owners use Facebook for marketing. Available at http://www.zdnet.com/blog/facebook/two-thirds-of-small-business-owners-use-facebook-for-marketing/1529. Accessed October 26, 2012.

Zhu F, Zhang X (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *J. Marketing* 74(2):133–148.

Zhu L, Benbasat I, Jiang Z (2009) Let's shop online together: An empirical investigation of collaborative online shopping support. *Inform. Systems Res.* 21(4):872–891.