HARVARD | BUSINESS | SCHOOL

DOCTORAL PROGRAMS

DISSERTATION ACCEPTANCE CERTIFICATE

(To be placed in First Library Copy)

The undersigned, approved by the Chair of the Doctoral Programs, have examined a dissertation entitled

Essays on Strategy and Management of Platforms

presented by **Anil Rajnikant Doshi**

candidate for the degree of Doctor of Business Administration and hereby certify that it is worthy of acceptance.

Signature ....

Lee Fleming, *Chair*

Signature ....

Scott Stern

Signature ....

Dennis A. Yao

Signature ....

Feng Zhu

Date April 7, 2015

WYSS | SOLDIERS FIELD | BOSTON, MA 02163 | Ph 617.495.6101 | Fx 617.496.2203 | doctoralprograms@hbs.edu | GEORGE F. BAKER FOUNDATION

Essays on Strategy and Management of Platforms

A dissertation presented

by

Anil Rajnikant Doshi

to

The Technology and Operations Management Unit

in partial fulfillment of the requirements

for the degree of

Doctor of Business Administration

in the subject of

Technology and Operations Management

Harvard University

Cambridge, Massachusetts

May 2015

Scott Stern Anil Rajnikant Doshi

Essays on Strategy and Management of Platforms

Abstract

In this thesis, I research the management of platforms by participating organizations and study the ensuing performance of both participating organizations and the platform.

In the first essay, titled "The Impact of High Performance Outliers on Two-Sided Platforms: Evidence from Crowdfunding," I study how one kind of observable on platforms affects both the subsequent entry decision of organizations and the performance of the platform. I focus on the arrival of high performing sellers and study how these "outliers" affect the subsequent growth and liquidity of the platform. In the context of the two largest rewards-based crowdfunding platforms, I find that outliers are followed by a relative increase in entry and transaction volume on the competing platform. Moreover, this average effect is stronger for marginal, or low quality, sellers. Within the platform hosting the outlier, transaction volume increases for sellers in the same product category as the outlier, but this average effect reverses for outliers in certain product categories. The results suggest that the impact of heterogeneous users depends on platform rules, and that in addition to pricing, competing platforms may selectively focus on attracting users with high performance potential to achieve the desired mix of buyers and sellers.

In the second essay, titled "Social Media, Loyalty, and Organizational Performance" (written with Shiladitya Ray), we study how the interactions between users and organizations on social media relate to organizational performance outcomes. Specifically, we explore the relationship between expressions of loyalty on social media and performance. We relate the number of followers on Twitter to television show ratings and find that change in the number of individuals following an organization's Twitter account prior to the realization of a repeated performance outcome is positively associated with that outcome. We present evidence of the heterogeneity in the effect, showing that the relationship is stronger for organizations that match the demographics

of the social network and niche product categories. We also show that higher levels of a show's initial following mitigates the relationship between followers and performance for shows in niche categories, and tentatively strengthens it for show's in non-niche categories.

In addition to considering the theoretical relationship between social media and organizational performance, we employ a parsimonious prediction model relating the two, showing that a model with social media measures outperforms both a baseline autoregressive model and a model that includes search data. In so doing, we extend recent literature that uses real-time data to predict current economic indicators by using social media data to predict organizational performance outcomes. Our results indicate that technological innovations can diminish the distance between the organization's boundaries and outside stakeholders. This reduction in distance underlines the importance for firms to attend to their non-transaction interactions on social media.

In the final essay, titled "Organizational Management of Social Media," I address how organizations manage activities on social media, beginning with the decision to adopt social media and its rate of diffusion within the organization. In the context of television show adoption of Twitter, I show that larger organizations and organizations started more recently more readily adopted social media. I also provide evidence highlighting the heterogeneity in organizational approaches to social media. By looking at the heterogeneity in the rates of diffusion, I am able to distinguish differences in approaches to social media management along three dimensions: the timing, the speed, and the centrality of management within the organization. Finally, I show the changing nature of diffusion of a technology in an industry over time. By applying theories of strategy and innovation management, I underscore the importance of considering the impact of social media on the organization and processes of the firm.

Contents

Dedicated to my Suzy

Author List

The following author contributed to Chapter 1: Anil R. Doshi.
The following authors contributed to Chapter 2: Anil R. Doshi and Shiladitya Ray.
The following author contributed to Chapter 3: Anil R. Doshi.

Acknowledgments

I am deeply indebted to many individuals, whose collective support has helped me realize this moment of accomplishment.

First, I thank Scott Stern, my advisor, for his time, mentorship, and for masterfully guiding my research. Following his example and advice has made me a better scholar.

I also thank my dissertation committee. Thank you Dennis Yao for consistently providing me feedback, providing a perspective that always helped improve my work, and asking questions that continue to shape research agenda. Thank you Lee Fleming for supporting me and participating in my growth from the beginning of the doctoral program. Thank you Feng Zhu for helping me keep a healthy perspective and for always steering my methods in a better direction.

Many other faculty have played a guiding role in my time as a doctoral student. Thanks to Andrew King and Michael Toffel for their mentoring and advice. Thanks to the Technology and Operations Management Unit for providing me with the opportunity to pursue my doctorate.

Importantly, my doctoral colleagues provided invaluable support, feedback, and advice. Sharing your rich research and brilliant approaches to thinking about very interesting questions has indelibly shaped my perspective. My appreciation and thanks to Silvia Bellezza, Sen Chai, Joseph Daves-Gavin, Andrea Hugill, Scott Lee, Frank Nagle, Pat Satterstrom, Luciana Silvestri, Haris Tabakovic, and Thomas Wollmann. I extend a special thanks to William Schmidt for being an excellent mentor.

Thank you to my co-author Shiladitya Ray, for working collaboratively with me to collect data, generate ideas, and develop a paper.

I thank John Korn and Jennifer Mucciarone of the Harvard Business Doctoral Office for their personal support. I also appreciate the Doctoral Office's generous financial support. Additional financial support was provided by an Amazon Web Services grant, where some of the data collection and computations were housed and run. This dissertation is typeset in XeTeX using a template provided by Andrew Leifer, `http://github.com/aleifer/LaTeX-template-for-Harvard-dissertation`.

Prior to joining the doctoral program, I had many teachers who influenced my thinking and impacted my desire to pursue a doctorate. I thank my high school English teacher, Ellen Epstein, for being the first teacher who encouraged my intellectual curiosity. I also thank the following professors from Dartmouth College for igniting my interest in the social sciences: David Blanchflower, Roger Masters, Matthew Slaughter, and Diederik Vandewalle.

Finally, I thank my family and friends. I stand on the shoulders of my mother and father, Lata and Rajnikant Doshi. Their sacrifice has been the springboard for my achievements. Thank you for always wanting more for me, for emphasizing the importance of education, and for showing me the value of hard work. Thank you to my brother, Neel Doshi, for being a model of success in and out of the classroom. Thank you to the parents of my wife, Kyoung Ja Oh and Sung Bong Oh, for their unwavering support and help. I acknowledge my children, Makias and Sajin, who remind me of the importance of continually learning. I hope they stand on the shoulders of their parents and effortlessly leap even further.

Most of all, thank you, Sue. You have been and continue to be everything to me.

1

The Impact of High Performance Outliers on Two-Sided Platforms: Evidence from Crowdfunding

1.1 Introduction

Two-sided platforms that facilitate transactions among distinct sets of users are becoming increasingly pervasive, particularly in the digital economy. The types of users that join each side may affect their subsequent growth and trajectory. User effects may also depend on the competitive environment faced by the platform, and may exercise a particular impact in the growth stage of platforms. Sellers may be heterogeneous in their product category, size, and performance, among other dimensions.[1] This paper investigates the impact of high performing sellers, or outliers, on the subsequent entry and liquidity of a two-sided platform, in the context of platform competition.

The impact of an outlier is illustrated by the performance of Ministry of Supply, a fashion company providing athletic business attire that began as a highly successful crowdfunding campaign.

[1] I call the users on one side of a platform *sellers* (or *creators* within the context of the empirical setting) and the users on the other side *buyers* (or *backers*).

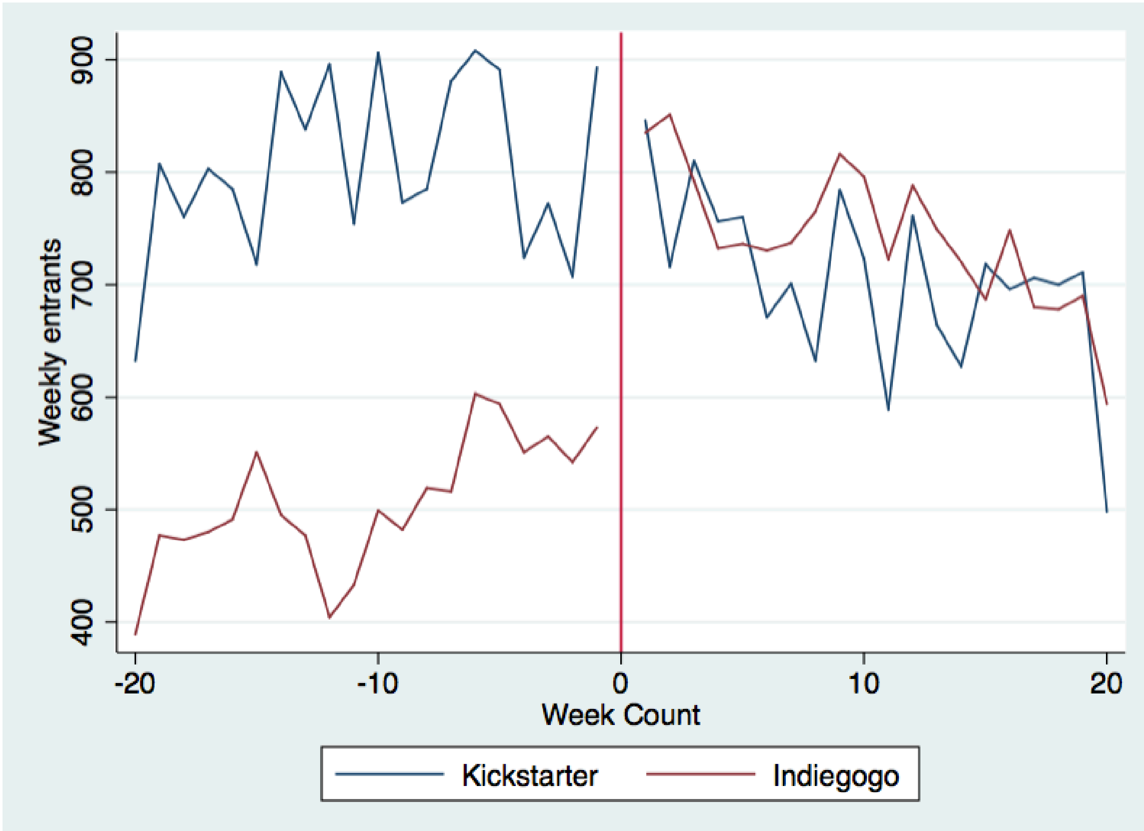


Figure 1.1: Seller Entry Before and After the Ministry of Supply Crowdfunding Campaign. This graph presents weekly entry on Kickstarter and Indiegogo for 20 weeks prior to the start and after the conclusion of the Ministry of Supply Kickstarter campaign, which ran from June 8, 2012 to July 11, 2012.

During the summer of 2012, three MIT graduate students were considering the viability of designing and selling business attire that functioned like athletic clothing. To estimate potential demand, the founders launched a Kickstarter crowdfunding campaign with a $30,000 goal over 33 days. They reached their goal within four days, and by the end of the campaign, they raised just under $430,000, or approximately 50% more than any prior crowdfunding project in the fashion category. As the three students dropped out of school to launch their company, the question remained as to how Ministry of Supply's unprecedented success affected Kickstarter. Figure 1.1 presents visual evidence that seller entry on Kickstarter appears unaffected, but entry discontinuously increases for its largest competitor, Indiegogo.

The previous example is the basis of the research question for this paper: how do high performing outlier sellers affect a two-sided platform, in the context of platform competition? I address this question at two separate levels. First, at the platform level, how do outliers impact the relative entry of subsequent sellers and the liquidity of competing platforms? Second, within the platform that hosted the outlier, how do outliers impact seller entry of and liquidity for similar sellers?

I answer these questions using crowdfunding data. Crowdfunding on online platforms has become a viable means to raise capital for firms, projects, and other causes. Kickstarter, the largest crowdfunding platform in terms of dollars pledged, recently announced it has raised a total of $1 billion on its platform since inception in 2009 (Kickstarter, 2014d). An ecosystem of crowdfunding platforms has emerged to raise funding for anything from scientific research projects to charitable causes and medical procedures. This paper utilizes data from the two largest rewards-based crowdfunding sites, Kickstarter and Indiegogo. I design an empirical approach that aggregates to the product category level over time, by platform, to arrive at estimates of relative entry and transaction volume prior to and after the outlier, for multiple outliers. The relative estimates control for crowdfunding shocks, and provides results for the entire rewards-based crowdfunding industry.

In the crowdfunding context, I find that outliers on Kickstarter are followed by a relative increase in entry and transaction volume (as measured by dollars pledged) on the competing platform, Indiegogo. This average effect is stronger for marginal, or low quality, sellers. The result suggests that the rules of the platforms are important factors in understanding the utility considerations of prospective sellers. Within Kickstarter, transaction volume increases for sellers in the same product category as the outlier, but this effect reverses for outliers in certain product categories. The differential responses by category possibly suggest that some outliers attract buyers with a taste for specific kinds of consumption, while other outliers attract more general buyers.

The competitive impact of outliers is also of interest in other settings. For example, platforms actively seek to host high performing goods to drive competitive differentiation from others. Video game console manufacturers typically highlight 'big name' developers and game titles to be released concurrently with the console to drive initial consumer adoption. Sometimes, the platform itself will seed one side of the market with prospective high performers to drive adoption on the other side. Media companies such as Netflix have started producing video content for their subscribers, featuring famous actors (e.g. Kevin Spacey in *House of Cards*). iOS, the mobile operating system platform, highlights featured developers that have created historically high performing apps for the platform. The results in this paper will provide a framework to consider these other cases of platforms, platform competition, and outliers.

1.2 Literature Review

1.2.1 Heterogeneity in Platform Participants

Models from theoretical research on two-sided markets and platforms make different assumptions about agent heterogeneity. Some assume users are homogeneous on both sides of the market (Caillaud & Jullien, 2003; Parker & Van Alstyne, 2005; Evans & Schmalensee, 2010). Others assume agent heterogeneity in either preferences over "membership" (i.e. utility from participating in the market) (Armstrong, 2006), over "interaction value" (i.e. number of users on the other side) (Rochet & Tirole, 2003; Ambrus & Argenziano, 2009), or over both (Rochet & Tirole, 2006; Weyl, 2010). Given these agent preferences, participation, and the network effects of interest, many models are solved for the optimal pricing and structure on both sides of the market.

Explicit consideration for the attractiveness of certain users is given by Rochet & Tirole (2003). They argue that the existence of "marquee" buyers (sellers) increases the desirability of the platform to sellers (buyers). Pricing for the same side as the marquee agent decreases, and increases for the other side. Ambrus & Argenziano (2009) note that users may be heterogeneous on dimensions that extend beyond preference over the size of the other side of the market. Specifically they may be heterogeneous "with respect to the network externality they generate—their 'attractiveness' to consumers on the other side." Hagiu (2009) makes the case for pricing to increase for producers when consumers demand product variety.

Empirical studies on two-sided markets have focused on pricing implications on changes in competition (Jin & Rysman, 2013), measuring indirect network effects (Rysman, 2004), and platform entry (Zhu & Iansiti, 2012; Seamans & Zhu, 2014).[2] Another stream of papers have focused on non-pricing strategies on platforms (Boudreau & Hagiu, 2009). Work that is the most closely related to this paper considers the impact of outliers on the adoption of platforms by users on the other side of the market (Binken & Stremersch, 2009; Lee, 2013). In the context of video games, outlier titles cause an increase in console purchases by video game players, though the magnitude of the change varies between papers.[3] Prior research considers neither follow on entry on the same side as the outlier, nor consumer transaction behavior following the outlier arrival.

More broadly, heterogeneity among users has been empirically studied in the context of network effects and platform competition. Boudreau (2012) showed that heterogeneity in sellers

[2] I refer the reader to Seamans & Zhu (2014) for a review of pricing studies in two-sided markets.

[3] The magnitude of the increase differs between the two papers. Binken & Stremersch (2009) estimates a 14% increase in console adoption from a hit title, while Lee (2013) estimates that counterfactual loss in platform sales from removing a hit game was a a maximum of 5.5%.

declines over time. Platform adoption is affected by distribution channel (Bresnahan & Yin, 2005) and by direct network effects i.e. number of users on the same side (Augereau et al., 2006). Understanding how differences in users drive the equilibria and tipping point of platform competition has been the focus of other empirical studies. For example, obtaining exclusive agreements with select users can help entrant platforms compete with incumbents (Lee, 2013). User characteristics (Hendel et al., 2009) and platform differentiation (Cantillon & Yin, 2011) can also explain agent adoption choices. In general, when network effects are present, the diffusion of adoption depends on the characteristics of current users. Having "boundary spanners" adopt accelerates the adoption decision of potential users (Tucker, 2008).

1.2.2 Market Entry

Research on entry has focused on entry by new firms or existing firms into new products (Gilbert & Newbery, 1982; Reinganum, 1983), new geographies (Chung & Alcacer, 2002; Alcacer & Chung, 2007) , or new industries (Rumelt, 1982; Prahalad & Bettis, 1986; Silverman, 1999). New firms considering entry typically have to make the decision, given incumbents employ policies of preemptive entry (Lieberman & Montgomery, 1988), price competition, and other barriers to entry. Resources (Helfat & Lieberman, 2002) and the development of capabilities (King & Tucci, 2002) often direct the entry choice.

Entry in a two-sided platform possesses characteristics from multiple types of entry, and does not readily fit into any one traditional type. The choice to enter a platform possibly requires changes in product and internal processes or reallocations in corporate resources, typically to match with the standards and requirements of the platform (Altman, 2015). Also, while both incumbents and entrepreneurs may enter on a platform, just as in a 'typical' market, the existence of the platform and its rules may result in a different competitive environment. Incumbent barriers to entry may not be as effective, as entry is regulated by the platform. In fact, barriers to entry may exist, but are typically imposed by the platform's regulatory regimes, rather than competing firms (and may not necessarily affect only entrepreneurial firms).

This paper contributes to the empirical literature on two-sided platforms and entry. First, in addition to the indirect network effects typically considered, I explicitly consider the impact of same side effects on the entry choice of sellers (Hagiu, 2009). Second, I consider the impact of non-pricing shocks to platforms (i.e. the arrival of heterogeneous sellers, namely performance outliers) in the context of platform competition. The paper also extends prior work that looked at how outlier sellers impacted buyer adoption (Binken & Stremersch, 2009; Lee, 2013) by looking at subsequent seller entry and platform liquidity. Third, I extend the literature on entry by considering platform entry as a different type of entry decision.

1.3 Theory

To build the theory of outlier sellers on two-sided platforms, I begin with the implications of an outlier on the utility of subsequent sellers. Then I argue that the presence of network effects will result in seller entry and liquidity being impacted in the same direction. Finally, I turn to the impact within the platform hosting the outlier, by considering how the outlier will impact similar sellers.

1.3.1 Outliers and Subsequent Seller Entry with Platform Competition

To provide a framework for the entry decision after the arrival of a performance outlier, I provide a stylized model of entry on two-sided platforms. I begin with the seller's utility function on platform, ϕ (Rochet & Tirole, 2006; Weyl, 2010):

$$U_s^\phi = V_s^\phi + (p_s^\phi - c_s^\phi)N_B^\phi \tag{1.1}$$

Users are buyers, $b \in B$ or sellers, $s \in S$. V_s^ϕ is agent s's net benefit of *participating* on the platform and consists of a benefit, P_s^ϕ, and cost of participation, C_s^ϕ, or $V_s^\phi = P_s^\phi - C_s^\phi$ The terms p_s^ϕ and c_s^ϕ are the benefit and cost of *transacting*, respectively.[4] The term N_B^ϕ represents the number of buyers in the market.

In the case of a monopoly platform, a seller will enter if the utility from entering exceeds the outside option, normalized to zero, or $U_s^\phi > 0$:

$$V_s^\phi + (p_s^\phi - c_s^\phi)N_B^\phi > 0 \tag{1.2}$$

For sellers considering multiple platforms (assuming no multi-homing), the entry decision is also a function of the utility from the competing platform. Assuming two platforms, ϕ and ω, the decision to enter platform ϕ is $U_s^\phi > \max(U_s^\omega, 0)$, or:

$$V_s^\phi + (p_s^\phi - c_s^\phi)N_B^\phi - \max[V_s^\omega + (p_s^\omega - c_s^\omega)N_B^\omega, 0] > 0 \tag{1.3}$$

[4]Weyl (2010) does not allow for heterogeneity in cost, indicating that the cost is explicitly the price charged by the platform conditional on the other side of the market. I consider a general cost term that varies by seller that includes, but is not limited to, platform fees.

Simply stated, a seller s will enter platform ϕ if her expected utility from entering that platform is greater than the utility of entering platform ω and not entering any platform.

The performance outlier signals the viability of the platform or transacting on it or it acts as diffusing agent for prospective sellers who were not previously considering entry. For this set of prospective entrants, it is possible to assess the impact on entry of a shock to platform ϕ, namely the arrival of a performance outlier using Equation 1.3. First, the arrival of an outlier increases the desirability of the platform to buyers (Rochet & Tirole, 2003) and thus increases buyer participation (Binken & Stremersch, 2009; Lee, 2013), N_B^{ϕ}. The increase in the number of buyers on the platform is an unambiguous improvement in seller utility (as long as $p_s^{\phi} - c_s^{\phi} > 0$). At the same time, the outlier increases competitive aversion among subsequent sellers (Brown, 2011), which in the model is reflected in an decrease in V_s^{ϕ}. This decrease may consist of a decrease in seller beliefs of participation benefit, P_s^{ϕ}, or an increase in actual costs, C_s^{ϕ}, the seller would incur to realize the value from transacting on the platform. Importantly, not all prospective sellers experience the same change in net benefit.

The outlier likely impacts the platform in other ways as well. First, the outlier will have received a high volume of transactions, so buyers who transacted with the outlier would experience a reduction in capital available for subsequent transactions. Capital constraints could be included in the utility function by considering expected number of transactions, rather than number of buyers, which would depend on buyer budget constraints. Also, a subset of sellers who make the consideration to enter may only consider participating on platform ϕ by considering the entry condition in Equation 1.2, as part of a 'parochial' utility maximization (e.g. if there are high platform search costs). This subset would only cause an increase in sellers on platform ϕ, increasing competition for the remaining prospective sellers. Multiple stages of response to an outlier would allow for the parochial sellers to enter first and then the remaining sellers to make their entry decisions. These additional channels by which an outlier can impact the platform are noted here but omitted from the framework to simplify the analysis and arrive at predictions. Sellers who are most susceptible to experiencing decreased net benefits of participation, V_s^{ϕ} from the outlier are likely to be most sensitive to reduced buyer budgets and competition from more sellers, so the inclusion of these additional mechanisms will drive the same predictive outcomes.

The question of entry for prospective sellers, then, involves comparing utility from platform ϕ with the outlier shock to the utility of joining the competing platform, ω. Depending on the relative value of participating and transacting across the two platforms, a seller may enter the platform hosting the outlier to capture the value of the relatively thicker market or the competing platform because of the relatively higher net benefit of participating (assuming that $\min(U_s^{\phi}, U_s^{\omega}) > 0$). In the latter case, the performance outlier effectively signaled the viability of joining a platform to engage in transactions with buyers, but the structures of the two platforms prompted the seller to choose the platform that did not experience the outlier shock. The countervailing impacts described above indicate that entry on the platform hosting the outlier may increase or decrease, relative to the competing platform. Thus, the average impact on subsequent seller entry on competing platforms is ambiguous

Regardless of the average effect, the impact will be even stronger among sellers whose costs are most likely to increase from arrival of the outlier. If, for the average seller, $U_s^{\phi} > U_s^{\omega}$, then entry on platform ϕ will increase, relative to platform ω. However, sellers with largest decrease in V_s^{ϕ} (from decreases in P_s^{ϕ} or increases in C_s^{ϕ}) from the outlier will disproportionately choose to enter platform ω. Conversely, if on average, $U_s^{\phi} < U_s^{\omega}$ and sellers disproportionately enter platform ω, then high c_s^{ϕ} sellers will select to enter on platform ω at an even greater rate.

Hypothesis 1 (H1) The impact of a performance outlier on the subsequent seller entry on a platform is stronger for sellers whose net benefits decrease the most.

1.3.2 Indirect Network Effects and Liquidity

The average effect on seller entry will be positive for either platform ϕ or ω. If platform ϕ experiences a positive impact in seller entry, then there is an increase in both number of sellers and number of buyers as a result of the outlier. If platform ω experiences a positive impact in seller entry, then there is a positive shock in arrival of number of sellers with no countervailing effect on the buyer side of the platform. In both cases, the positive shock from the outlier–whether it is to both sides or one side–affects participation without any changes to pricing to incentivize those users to participate. In this way, the outlier produces "pure" indirect network effects on the platform that experiences a relative increase in seller entry, whereas price changes create second order effects that may partially or fully offset the participation effect of those pricing changes (Rochet & Tirole, 2006). The relative increase of sellers results in an increase in buyer utility as a result of indirect network effects (Armstrong, 2006; Parker & Van Alstyne, 2005), which act as a "self-reinforcing feedback loop" (Gawer, 2014). The increase in buyer participation encourages entry by sellers and an increase in seller participation increases entry and choice for buyers. More choice is followed by better matches and more transactions, increasing liquidity, relative to the competing platform.

As stated in Section 1.3.1, the average effect of an outlier on relative seller entry across platforms

ϕ and ω is ambiguous due to the countervailing impact of the outlier on V_s^{ϕ} and N_B^{ϕ}. Though the average impact of an outlier seller on subsequent seller entry is unclear, whichever platform enjoys the relative increase in sellers would also experience a relative increase in liquidity on the platform.

Hypothesis 2 (H2) The arrival of a performance outlier on a platform will affect liquidity for competing platforms in the same relative direction as subsequent seller entry.

1.3.3 Impact of Outlier on Entry of Similar Sellers

Sellers that are similar to the outlier may be particularly susceptible to its effects. Platforms often segment sellers based on characteristics to facilitate search and matches. To investigate the impact of the outlier particularly on subsequent entry of similar sellers, I return to Equation 1.3. The increase in buyers, N_B^{ϕ}, may be particularly beneficial to similar projects if there is heterogeneity in the new buyers and some have a preference for sellers similar to the outlier. Similar sellers may also experience an increase in p_B^{ϕ} if the outlier provides a particular signal to similar sellers of the likelihood of transacting on the platform for that type of seller. If these effects are stronger, then the outlier will be followed by a relative increase in similar sellers on the platform that hosted the outlier.

Hypothesis 3a (H3a) Following a performance outlier on a platform, subsequent entry of similar sellers will increase, relative to different sellers.

Just as the positive impact of an outlier on similar sellers may be amplified, so would the negative impacts. Sellers may be particularly averse to a similar outlier, resulting in larger declines in V_s^{ϕ}. To the extent that the outlier receives a disproportionate amount of transactions, buyers may have subsequent capital constraints, particularly for similar sellers. When increases in the above costs are the stronger effect, firms select to defer entry or enter on the competing platform to effectively differentiate from the outlier (Ellison & Fudenberg, 2003).

Hypothesis 3b (H3b) Following a performance outlier on a platform, subsequent entry of similar sellers will decrease, relative to different sellers.

1.3.4 Impact of Outlier on Liquidity of Similar Sellers

The outlier's increase in buyer participation might particularly benefit similar sellers if the increased participation is from buyers who have a particular taste for transactions with sellers similar to the outlier. In that case, the increase in buyers and buyer transaction activity would disproportionately increase for similar sellers.

Hypothesis 4a (H4a) Following a performance outlier on a platform, subsequent transaction liquidity of similar sellers will increase, relative to different sellers.

As previously stated, outliers may also deter seller entry because of increased competition, seller decisions to differentiate, or budget constraints. These might be true especially of sellers that are similar to the outlier. Competition and the need to differentiate is most salient for similar sellers. The outlier may also crowd out future investment in similar sellers, as buyers may be budget constrained for similar sellers. In this case, the arrival of an outlier would have a negative impact on entry and liquidity for sellers that are similar to the outlier.

Hypothesis 4b (H4b) Following a performance outlier on a platform, subsequent transaction liquidity of similar sellers will decrease, relative to different sellers.

1.4 Setting and Data

I use crowdfunding platforms as the setting for this paper. The two sides of the market in crowdfunding platforms are capital seekers, or creators, and capital contributors, or backers. In the current iteration, crowdfunding is the process of raising capital from multiple contributors through an online platform. The use or recipient of funds is diverse and includes creative projects, firms, specific products, political and social causes, research, and personal circumstances. Backers can receive a number of different commitments from project creators in exchange for capital, including goodwill or recognition, equity, debt, and rewards that typically consist of promises of future delivery of goods and services.

Typically, when a creator initiates a project on a platform, she will typically provide certain content and set the parameters associated with the project. The creator typically describes the the project and its current progress using text and multimedia, and provides biographies and related experience of project creators. She will also set the desired funding goal, expiration of the campaign, and contribution tiers with different awards associated with different levels of pledges.

Creators can provide updates on the fundraising campaign or the project and backers can typically publicly comment on the project.

The platform fee structure typically involves a fee charged by the platform on the total amount of successful capital raised. Effectively, the creator is charged for a successful financing, similar to many traditional capital raising intermediaries. Given the typical absence of explicit membership fees, creators and especially backers likely experience a membership benefit for participating in the community aspect of crowdfunding.

A feature that may not be be typical of other two-sided markets is the limited availability, and thus turnover, of projects by creators. Typically, at the time of project creation, a duration is specified by the creator. At the expiration, the pledges will be transferred if the terms of the campaign were met. The temporary nature of crowdfunding projects implies that the supply of projects is time variant.[5]

An emerging stream of research around crowdfunding has investigated several aspects of the phenomenon, including determinants of success (Lambert & Schwienbacher, 2010; Mollick, 2014), incentives (Agrawal et al., 2014), choice of financing (Belleflamme et al., 2010), geography (Agrawal et al., 2011), legal aspects (Kappel, 2008), choices relative to experts (Mollick & Nanda, 2015), and backer behavior (Kuppuswamy & Bayus, 2014).[6]

1.4.1 Kickstarter and Indiegogo

I use the two largest crowdfunding rewards-based platforms, Kickstarter and Indiegogo, as the setting for this study. In the context of crowdfunding, I investigate how outlier projects impact subsequent project entry and backer behavior. The impact of outliers is of great importance to the platforms themselves, as Kickstarter has written about what it calls "blockbuster" projects on multiple occasions (Kickstarter, 2012, 2013a,b).

Though the overall services of Kickstarter and Indiegogo are substantively similar, I highlight a few important differences:

[5] The implications of turnover in the participant seller projects may be generalizable outside of crowdfunding. Experience goods, such as music, typically undergo a decay in consumption over time. Additionally, perishable goods, are also available for a limited time (Sweeting, 2012). This raises the question of whether backers consider their pledges to an experience good (i.e. the experience of contributing to a campaign on a crowdfunding platform) or a limited availability pre-purchase. This distinction may go towards understanding backer motivations for contributing and the appropriate response of creators to backers after campaigns are complete and funding is transferred.

[6] For a detailed review of crowdfunding research, see Kuppuswamy & Bayus (2014).

Fundraising mechanisms. All projects on Kickstarter require the project to meet or exceed its goal in order to receive the funds. If the goal is not achieved, backers are not committed to transferring their pledges. On Indiegogo, project creators can choose the previously described 'fixed' fundraising mechanism, or they can opt for 'flexible' fundraising, where the creator receives any pledged funds at the expiration of the project even if the goal is not met. If the project creator chooses flexible funding and the goal is not met, then Indiegogo charges a higher fee on the funds. Though both options are available, only 5,478 projects in my sample had fixed goals on Indiegogo, which represents 5.4% of all Indiegogo projects in the sample.

Fee structure. Kickstarter charges a 5% fee of successful projects (project creators must also pay 3-5% in transaction fees to payment processors). Indiegogo charges a 4% fee if the goal is met, regardless of funding mechanism. For flexible funding, the fee is 9% if the goal is not met.

Curation and categories. While many classes of projects are consistent across the two platforms, Kickstarter has historically been more restrictive in the types of projects that are allowed to post on its platform. This difference is represented by the available project categories on Indiegogo that are not on Kickstarter, primarily those related to causes (which includes community, political, religious, and non-profit projects).[7]

Statistics for the two platforms during the sample period (described in Section 1.4.3) are shown in Table 1.1.

1.4.2 Outlier Projects in Crowdfunding

Outliers exist at the tails of their respective distributions by definition—scientists produce prolific research (Azoulay et al., 2010a), athletes post record statistical performances, and musicians sell substantially more albums. To operationalize a performance-based definition of outliers, I employ a 'high water mark' approach to identify outlier projects. Specifically, among projects that received funding, I identify those that raised more in pledges than any prior project within the same category, and thus changed the tail of the distribution. To exclude early projects that raised low amounts of money but fulfill the above criteria (as is the case with early projects), I drop all projects where the amount pledged was less than the median pledged amount of high-water mark projects, which was $28,650.

[7] Kickstarter has recently loosened its rules (Kickstarter, 2014a) and amended its category structure (Kickstarter, 2014b,c), making the platform more open to the types of projects that are allowed.

Table 1.1: Platform Summary Statistics

	Mean	S.D.	Median	Min	Max
Indiegogo					
pledged	1,810.67	17,636.12	100.00	0.00	1,961,862.00
backers	23.71	218.93	3.00	0.00	33,253.00
goal	31,343.91	207,418.97	5,000.00	435.75	9,000,000.00
duration	51.13	33.46	45.00	0.00	916.00
fixed funding	0.05	0.23	0.00	0.00	1.00
updates	2.27	6.09	0.00	0.00	247.00
comments	9.34	91.95	2.00	0.00	17,675.00
received money	0.62	0.49	1.00	0.00	1.00
Kickstarter					
pledged	8,110.36	68,103.43	1,404.00	0.00	10,266,846.00
backers	108.14	853.61	24.00	0.00	91,585.00
goal	16,286.75	95,987.52	5,000.00	101.00	8,961,000.00
duration	35.75	14.64	30.00	1.00	91.00
fixed funding	1.00	0.00	1.00	1.00	1.00
updates	4.76	7.94	2.00	0.00	301.00
comments	32.27	999.06	0.00	0.00	145,900.00
received money	0.48	0.50	0.00	0.00	1.00
Total					
pledged	5,161.77	51,211.90	530.00	0.00	10,266,846.00
backers	68.62	641.72	10.00	0.00	91,585.00
goal	23,334.31	158,412.02	5,000.00	101.00	9,000,000.00
duration	42.95	26.40	35.00	0.00	916.00
fixed funding	0.56	0.50	1.00	0.00	1.00
updates	3.60	7.24	1.00	0.00	301.00
comments	21.54	731.46	1.00	0.00	145,900.00
received money	0.54	0.50	1.00	0.00	1.00

Note: Kickstarter $n = 102,383$. Indiegogo $n = 116,359$. Total $n = 218,742$. Summary statistics include the 70 outliers, which are summarized separately in Table 1.2. The variable *updates* is the number of updates posted by the project creator and *comments* is the number of messages posted by backers about a project. The variable *received money* equals one if pledged capital to the project was at least one dollar and the money was transferred from backers to the creator.

Table 1.2: Outlier Summary Statistics

	Mean	S.D.	Median	Min	Max
pledged	760,129.20	1,752,733.82	165,248.50	31,028.00	10,266,846.00
backers	8,301.30	18,186.68	2,095.00	121.00	91,585.00
goal	138,938.00	298,179.94	45,000.00	2,000.00	2,000,000.00
duration	38.57	15.35	31.00	2.00	85.00
fixed funding	1.00	0.00	1.00	1.00	1.00
updates	28.14	18.31	25.00	0.00	112.00
comments	1,951.66	5,323.37	309.50	0.00	30,023.00

Note: $n = 70$. Summary statistics for 70 outlier projects identified on Kickstarter platform (listed in Table A.1).

Based on this definition, I arrive at a set of 70 'category outlier' projects that appeared on Kickstarter, summarized in Table 1.2 and listed in Table A.1.[8] Outlier projects received a median of $165,248 in pledges from a median 2,095 backers. The 70 outliers account for 5.2% of all successful capital raised during the sample period on both platforms.

1.4.3 Data and Sample

My data consist of projects initiated on the Kickstarter and Indiegogo platforms from inception through January 2014. The data was assembled through a combination of web scraping the two websites and acquiring the data from third party providers. I include only completed projects that were initiated between April 2009 and January 2014. Further, I include only projects in categories that are shared across both platforms.

Additional projects were dropped to eliminate possible test or fake projects. First, projects entitled "Untitled Draft Project" were dropped. Also, projects with duplicative names, locations, platforms, and funding types (i.e. fixed or flexible) were considered trials and all but the most recently started project were dropped. Projects where the goal was less than $100 (consistent with Mollick (2014)) or greater than $10 million were also dropped from the sample.

The resulting number of projects in the sample for Kickstarter and Indiegogo is 102,383 and 116,359, respectively. Sample statistics for project related measures are presented in Table 1.1.

[8] Three projects on Indiegogo are characterized as outliers and are discussed separately in Section 1.6.1.

Several facts emerge from the summary statistics. The median goal was $5,000 on both platforms (though the mean goal of $31,344 on Indiegogo is almost double Kickstarter's mean goal of $16,287), but the median project duration of 45 days on Indiegogo was 15 days longer than Kickstarter. On Indiegogo 61.7% of projects received some amount of capital[9] , compared to 47.7% on Kickstarter. Though a higher proportion projects receive funding on Indiegogo, the average amount pledged for a project on Kickstarter was $8,110, compared to $1,811 on Indiegogo. The most successful project on Kickstarter during the sample period was the Pebble smartwatch, which raised $10.3 million, or more than five times the most successful project on Indiegogo, the Canary home security device, which raised $2.0 million. The Kickstarter community appears to be more engaged with more mean updates made by project creators and more mean comments posted by site members. Kickstarter also appears to have a more active community, with a mean number of updates by creators and comments by users of 4.8 and 32.3, respectively. That compares with Indiegogo, where mean updates are 2.3 and mean comments are 9.3.

Creation of Sample

For each outlier, a weekly sample is constructed using 20 weeks of data prior to the start of the outlier's project campaign and 20 weeks after its conclusion. For each week, measures are aggregated (described in Section 1.4.4) at the platform-category level. Each 40 week period 'outlier period' is then stacked for all 70 outliers, creating a sample, where an observation is an outlier period-platform-category-week that is centered around the campaign period of each respective outlier. Summary statistics of the stacked sample are presented in Table 1.3 and correlations are presented in Table 1.4.

1.4.4 Measures

To capture the trends of the platforms and categories over time, I derive a number of measures related to performance, growth, timing, and project characteristics.

Dependent Variables

Growth To measure the impact on the same side of the platform as the outlier, I look at the supply of crowdfunding projects prior to the arrival and after the conclusion of an outlier project.

[9] That includes flexible campaigns that raised at least one dollar and fixed campaigns that met their goal. For projects that chose fixed funding, only 18.8% received funding i.e. met or exceeded their goal.

Table 1.3: Sample Summary Statistics

	Mean	S.D.	Median	Min	Max
entrants	39.60	52.11	20.00	1.00	360.00
pledged	179,888.71	475,954.52	33,997.00	0.00	10,929,538.00
outlier platform	0.52	0.50	1.00	0.00	1.00
outlier category	0.08	0.27	0.00	0.00	1.00
post	0.51	0.50	1.00	0.00	1.00
goal	746,899.78	2,213,497.55	168,042.00	0.00	107,177,346.00
mean duration	44.86	17.93	42.65	0.00	312.00
platform-category age	126.53	53.89	128.14	0.00	314.57
Google trend index	16.16	20.66	6.00	0.00	100.00

Note: $n = 68,122$. An observation is an outlier period-platform-category-week.

Table 1.4: Correlation Table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) entrants	1.00								
(2) pledged	0.35	1.00							
(3) outlier platform	0.05	0.25	1.00						
(4) outlier category	0.04	0.00	-0.00	1.00					
(5) post	0.05	0.05	-0.02	-0.00	1.00				
(6) goal	0.50	0.26	-0.08	0.03	0.05	1.00			
(7) mean duration	-0.06	-0.16	-0.42	0.01	-0.03	0.07	1.00		
(8) platform-category age	0.55	0.34	0.12	0.01	0.19	0.34	-0.14	1.00	
(9) Google trend index	0.23	0.43	0.59	-0.00	0.11	0.08	-0.41	0.62	1.00

Note: $n = 68,122$. An observation is an outlier period-platform-category-week.

I create *entrants*, which is the number of weekly entrants. Median weekly seller entry per category was 39.29 projects.

Liquidity I look at buyer transactions by measuring amount of money pledged. Pledged dollars are totaled weekly by platform and category to arrive at the *pledged* measure. Pledged money for non-dollar denominated projects were converted to U.S. dollars using the exchange rate between the local currency and dollars as of the final day of the project campaign. Median weekly dollars pledged was $177,994. One substantive assumption made in the creation of these measures is that all funding occurs on the concluding day of a campaign.[10] This assumption was required due to data limitations.

Focal Independent Variables

Timing relative to outlier For each of the 70 outliers, a sample 20 week window before the start date and and 20 week window after the the end date of the outlier is compiled and stacked. The variable *post* captures whether the focal week occurs after the conclusion of the outlier project campaign. The variable equals one when the focal week is after the outlier end date and zero for weeks prior to the outlier start date. The period the outlier project was live on the platform is excluded from the sample period.

Outlier platform and category I develop two measures, *outlier platform* and *outlier category* to capture two different dimensions of the outlier on the platforms. First, *outlier platform* equals one if the focal platform is the same as the one that hosted the outlier, and zero otherwise (i.e. the competing platform).[11] This measure captures the across platform affect. Second, *outlier category* equals one if the focal category is the same as the outlier's project category, and zero otherwise, which captures the within platform and across category impact.

1.4.5 Time-varying Controls

I include several measures to control for time-varying factors at the platform, category, and project level. To control for the general growth trend of each crowdfunding platform, I include *Google trend*

[10]To account for this assumption, I run a robustness test where I only consider pledges for projects where the money was transferred (i.e. flexible projects on Indiegogo that raised at least $1 or fixed projects on both platforms that met their goal). Because the transfer of money takes place on the final day of the project's campaign, the assumption of all pledges arriving on that day is more accurate when thinking of the money as 'committed pledges.' area consistent using this alternative definition of pledged capital.

[11]For the 70 outliers described in Section 1.4.2, *outlier platform* is equals 1 for Kickstarter and 0 for Indiegogo.

index for the two platforms over the sample period. The Google trend index is a measure of the relative search frequency of 'kickstarter' and 'indiegogo' on a weekly basis (and does not capture absolute search volumes). During the chosen time period, the index normalizes the most popular search term in the most popular week at 100 and presents the remaining weeks relative to that. The index has been used to control for secular trends (Ghose et al., 2012) and has been shown to forecast current business and economic activities (Choi & Varian, 2012; Wu & Brynjolfsson, 2013).

To control for the different trends in categories on each platform, I include a measure of the category's age. The measure *platform category age* was calculated as the number of weeks from the first entrant observed in each category on each platform.

For the *pledged* regression, I also include certain *mean duration* and *dollar goal*—project-level characteristics to control for differences in projects reaching the conclusion of their campaign each week. *mean duration* is the average project duration and *dollar goal* is the total dollars set in goals for all projects completing their campaigns during the focal week.

1.5 Empirical Strategy

To study how outliers arriving on a platform impact subsequent entry and capital pledged responses, I estimate the following general model[12]:

$$E(Y_{it}|X_{it}) = f[\varepsilon_{it}; \beta_1(\texttt{post}_{it} \times \texttt{outlier platform}_i) + \beta_2\texttt{post}_{it} + \beta_3 X_{it} + \theta_i + \gamma_t] \tag{1.4}$$

where i indexes each outlier period-platform-category and t indexes time in weeks. The dependent variable, Y_{it} represents *entrants* or *pledged*. *outlier platform* equals one for weeks when the focal platform is the same as the outlier's. To variable *post* equals one for weeks after the outlier concludes its campaign, and equals zero for weeks prior to its start. The term θ_i represents fixed effects for each outlier period-platform-category. Included in γ_t are year fixed effects and calendar month fixed effects. Included in X_{it} are time-varying measures of the platforms, categories, and projects. Specifically, *Google trend index* is a weekly measure of the relative search volumes for the terms "kickstarter" and "indiegogo" on Google, with a maximum index value of 100. *platform category age* controls for differences in secular trends experienced within each category. Included

[12]The empirical specification has the design of a difference in difference model. However, the model contains an important distinction from the 'standard' difference-in-difference approach, in that the platforms operate in a competitive environment and the arrival of an outlier on one platform is likely to have an impact on the other. As a result, what is estimated is the differential impact of the arrival of an outlier on the hosting platform relative to the competing platform, rather than a 'treatment' effect.

in the *pledged* equation are *dollar goal* and *mean duration*.[13] Each of the individual and interaction terms between *outlier platform* and *outlier category* are fully absorbed by the fixed effects, θ_i, and are thus excluded from the above model.

The coefficient of interest is β_1. The term β_1 represents the differential impact on the platform that hosts the outlier as compared to the competing platform, prior to the launch and after the conclusion of the outlier. Thus, β_1 reflects the platform competitive impact of the arrival of an outlier.

To test the impact of outlier within the platform and across categories the following model is estimated:

$$\begin{aligned} E(Y_{it}|X_{it}) = f[\varepsilon_{it}; \alpha_1(\texttt{outlier platform}_i \times \texttt{outlier category}_i \times \texttt{post}_{it}) \\ + \alpha_2(\texttt{post}_{it} \times \texttt{outlier platform}_i) + \alpha_3(\texttt{post}_{it} * \texttt{outlier category}_i) \\ + \alpha_4\texttt{post}_{it} + \alpha_5 X_{it} + \theta_i + \gamma_t] \end{aligned} \quad (1.5)$$

In this specification, the variable *outlier category* equals one on weeks when the category is the same as the outlier's. The coefficient of interest is α_1, which represents the differential impact on the category of the outlier within the platform that hosts the outlier prior to the launch and after the conclusion of the outlier, relative to the competing platform. In other words, α_1 indicates whether the platform competitive impact is mitigated or strengthened in the same category as the outlier project, and thus indicates the within-platform impact of the outlier.

1.5.1 Identification

The identification strategy relies on the idea that the arrival of the outlier projects is exogenous to the platform. A concern is that the outliers are endogenous—the timing of the campaign and its parameters are strategic choices set by the project creators that are eventually revealed to be outliers. By that reasoning, the project creators may merely be 'timing the market' and earning high performances because of increasing growth of the platform or a relative decrease in supply.

I address these concerns by providing anecdotal evidence of outlier exogeneity from informal interviews with seven creators of outliers and by addressing the endogeneity concern into the empirical specification. Informal interviews with outlier project creators provides some evidence that outliers are exogenous. If outliers are endogenous, then creators would have high *ex ante* expec-

[13] Also included in models that contain *dollar goal* and *mean duration* is *zero exits*, a dummy variable set to 1 when the number of exits during the week equals 0. This term is included to account for the weeks during which there were no exits and mean duration is set to zero, rather than undefined.

tations about the amount of capital that will be raised and timing choices would be tied to the platform. None of the creators communicated both sentiments.

Some project creators expressed concern over the success of their project, implying that there was no prior expectation of success that changed the distribution of pledges for the category. One creator indicated, “I didn’t think it would succeed, and I wasn’t looking forward to the humiliation of failure in such a public forum,” while another said, “We hope we’d reach our goal.” Creators also indicated that alternative methods of financing were not an option: “There was no other way to raise money. Regular routes of investment, loans, *et cetera* were closed.”

Regarding timing, duration for projects were typically set to 30 days, based on a general Kickstarter recommendation (made to all projects), rather than a strategic choice to time the length or conclusion of the project. Another creator indicated timing was dictated by outside factors, namely actor schedules: “We had to raise the money by [the end date] or we’d lose our lead actor to his second season of ‘True Blood’.” A music group set their start date on a significant anniversary: “November 22nd [the project’s start date] is a fan celebrated day because it was the date of our last show when we originally broke up.”

The empirical specification provides further consideration of secular trends and timing, by accounting for the trends in the crowdfunding industry, each platform, categories within each platform, and of the projects that complete each week. First, the difference-in-difference specification with year and month fixed effects should partially account for the trends in the crowdfunding industry and trends over time. Second, the Google Trends Index for each separate platform acts as a control for the periods during which each platform was more popular. Third, by including the age of each platform-category, I account for the different timing of growth trends that occurred among different product groups within each platform. Finally, weekly pledged capital may be a function of the project that are completing in each week, so aggregate project characteristics (total goal dollars and mean duration) control for different projects coming due each week.

1.6 Empirical Results

The nonnegative and highly skewed distributions of both dependent variables motivate the choice of using the Poisson model with quasi-maximum likelihood standard errors (Azoulay et al., 2010b) to estimate Equations 1.4 and 1.5. Mean and median statistics for the dependent variables in Table 1.3 illustrate the skew in the data. Further evidence is provided by the skewness measure for each variable, which for *entrants* and *pledged* are 2.44 and 9.46, respectively. There was no entry in 5.4% of weeks and no exits (hence no pledges) in 9.1% of weeks.

Table 1.5: Outlier Impact On Entry

DV *entrants*	(1)	(2)	(3)	(4)
entrants type:	All	All	All	Low Quality
Outlier platform x Post	-0.165**	-0.194**	-0.206**	-0.370**
	(0.03)	(0.02)	(0.03)	(0.05)
Post	0.286**	0.098**	0.112**	0.065
	(0.03)	(0.02)	(0.02)	(0.04)
Year FE	No	Yes	Yes	Yes
Month FE	No	Yes	Yes	Yes
Time-Varying Controls	No	No	Yes	Yes
Outlier period-platform-category FE	Yes	Yes	Yes	Yes
Observations	72125	72125	68117	68105
Outlier period-platform-categories	1806	1806	1801	1796

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variable is count of project entrants in Columns 1 to 3 and count of projects that raised no pledges in Column 4. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for 70 Kickstarter outliers are stacked to produce the sample. An observation is an outlier period-platform-category-week.

1.6.1 Results

Results for the impact of 70 Kickstarter outliers on entry and pledged dollars are reported in Tables 1.5 and 1.6, respectively. In each table, Column 1 excludes time fixed effects and controls and Column 2 includes fixed effects. Both columns are included for reference.

Seller Entry with Platform Competition

Following an outlier on Kickstarter, entry on that platform declined by 18.6% (Table 1.5, Column 3, $\exp(\beta) - 1 = \exp(-0.206) - 1 = -0.186, p < 0.001$), relative to the competing platform, Indiegogo. Given a mean of 33.5 weekly entrants during the pre-period, the result implies a net

Table 1.6: Outlier Impact On Pledges

DV: *pledged*	(1)	(2)	(3)
Outlier platform x Post	-0.220**	-0.242**	-0.238**
	(0.04)	(0.04)	(0.04)
Post	0.515**	0.165**	0.187**
	(0.03)	(0.04)	(0.04)
Year FE	No	Yes	Yes
Month FE	No	Yes	Yes
Time-Varying Controls	No	No	Yes
Outlier period-platform-category FE	Yes	Yes	Yes
Observations	70975	70975	67877
Outlier period-platform-categories	1776	1776	1769

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variable is dollars pledged. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for 70 Kickstarter outliers are stacked to produce the sample. An observation is an outlier period-platform-category-week.

Table 1.7: Indiegogo outlier Impact

	(1)	(2)	(3)
DV:	entrants	Low Quality entrants	pledged
Outlier platform x Post	0.113*	0.316	0.212*
	(0.06)	(0.25)	(0.10)
Post	-0.140**	-0.375+	-0.000
	(0.02)	(0.22)	(0.05)
Year FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Time-Varying Controls	Yes	Yes	Yes
Outlier period-platform-category FE	Yes	Yes	Yes
Observations	3120	3120	3120
Outlier period-platform-categories	78	78	78

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variable is count of project entrants in Column 1, dollars pledged in Column 2, and count of projects that raised no pledges in Column 3. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for three Indiegogo outliers are stacked to produce the sample. An observation is an outlier period-platform-category-week.

difference of 6.2 ($33.5 \times -0.186 = -6.2$) projects per week entering on the competing platform (Indiegogo) after the arrival of an outlier.

By comparison, the three outliers on Indiegogo result in a relative increase in entry on Indiegogo, relative to Kickstarter (Table 1.7, Column 1, $\beta = 0.113, p < 0.048$). Regardless of the platform hosting the outlier, Indiegogo experienced the relative improvement in entry. These results suggest that after the arrival of an outlier, subsequent entrants on average faced high enough costs on Kickstarter relative to Indiegogo, so as to offset the benefit from an increase in buyers.

To test the hypothesis that prospective sellers with relatively higher costs will experience a stronger than average effect, I decompose entry into types to delineate project quality. Projects of lower quality are more likely to face higher costs after an outlier, in terms of both incurring costs to improve quality to successfully transact with buyers and experiencing a greater degree of competitive aversion. I consider entry only for projects that eventually do not raise any pledges during their campaigns. I then look at entry of these 'lower quality' projects to distinguish in entry trends between them and all projects. Results are presented in Column 4 of Table 1.5. Entry of low quality projects decrease on Kickstarter at a much greater rate than do all projects ($\beta = -0.370, p < 0.001$). The increase in magnitude of negative relative entry for low quality projects provides evidence in support of H1.

Indirect Network Effects

Looking at Column 3 of Tables 1.5 and 1.6 shows that the direction of seller entry and amount pledged is the same (negative) across platforms ($\beta = -0.204, p < 0.001$ and $\beta = -0.238, p < 0.001$, respectively). The same result of consistent directions for the three outliers on Indiegogo holds, where outlier impact on both seller entry and amount pledged is positive (Table 1.7, $\beta = 0.113, p < 0.048$ and $\beta = 0.212, p < 0.032$, respectively). The matched direction of effects on both outcomes of interest provides evidence for H2.

Within Platform

Results for outlier impact on similar sellers are reported in Table 1.8. The average effect on project entry in the same category as the outlier is not significantly different from projects outside the category (Column 1, $\beta = 0.054, p < 0.159$). There is no evidence in support of H3a or H3b. The observed negative platform effect on transactions is mitigated for projects in the same product category as the outlier within the platform that hosted it (Column 2, $\beta = 0.195, p < 0.009$). The positive and statistically significant impact on dollars pledged indicates that there is a capital 'spill in' effect for similar projects after the arrival of an outlier. The within platform results on pledged dollars provide support for H4a.

Disaggregating results by the category of the outlier reveals interesting heterogeneity in the outlier's impact within the platform hosting the outlier. Tables 1.9 and 1.10 contain category by category results for entry and pledged dollars, respectively.[14]

[14]Each table includes results for six categories, where there were statistically significant results for either entry or pledged (or both): film and video, theater, music, dance, design, and technology. The remaining seven categories contained no statistically significant results for either entry or pledged and are not reported. Those categories are art, comics, fashion, food, games, photography, and publishing.

Table 1.8: Outlier Impact Within Hosting Platform

	(1)	(2)
DV:	entrants	pledged
Outlier platform x Outlier category x Post	0.054	0.195**
	(0.04)	(0.07)
Outlier platform x Post	-0.211**	-0.255**
	(0.03)	(0.04)
Outlier category x Post	-0.031	-0.122
	(0.03)	(0.08)
Post	0.115**	0.199**
	(0.02)	(0.04)
Year FE	Yes	Yes
Month FE	Yes	Yes
Time-Varying Controls	Yes	Yes
Outlier period-platform-category FE	Yes	Yes
Observations	68117	67877
Outlier period-platform-categories	1801	1769

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variable is count of project entrants in Column 1 and dollars pledged in Column 2. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for 70 Kickstarter outliers are stacked to produce the sample. An observation is an outlier period-platform-category-week.

Table 1.9: Outliers Impact By Category On Entry

Category	Film and Video	Theater	Music	Dance	Design	Technology
DV: *entrants*	(1)	(2)	(3)	(4)	(5)	(6)
Outlier platform x Outlier category x Post	0.133^{*}	-0.136^{**}	-0.135^{**}	-0.137^{**}	-0.387^{**}	0.174
	(0.06)	(0.04)	(0.02)	(0.01)	(0.09)	(0.13)
Outlier platform x Post	-0.288^{**}	-0.290^{**}	-0.156	-0.260^{**}	-0.221	-0.153^{**}
	(0.07)	(0.08)	(0.10)	(0.09)	(0.16)	(0.06)
Outlier category x Post	-0.166^{**}	-0.070	0.116^{**}	0.070	0.376^{**}	0.094
	(0.04)	(0.08)	(0.04)	(0.08)	(0.09)	(0.06)
Post	0.212^{**}	0.164^{**}	0.171^{*}	0.070	0.078	0.012
	(0.06)	(0.06)	(0.08)	(0.06)	(0.10)	(0.07)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-Varying Controls	Yes	Yes	Yes	Yes	Yes	Yes
Outlier period-platform-category FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8459	4016	7668	3120	4802	4888
Outlier period-platform-categories	230	104	206	78	130	128

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variable is count of project entrants. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for nine film and video, four theater, eight music, three dance, five design, and five technology Kickstarter outliers are stacked to produce the sample in each column, respectively. An observation is an outlier period-platform-category-week. Categories where both within platform entry and pledge effects (in Table 1.10) are not statistically significant are not reported (art, comics, fashion, food, games, photography, and publishing). To facilitate convergence in all regressions except the design equation, year fixed effects were aggregated for each two year period.

Table 1.10: Outliers Impact By Category On Amount Pledged

Category	Film and Video	Theater	Music	Dance	Design	Technology
DV: *pledged*	(1)	(2)	(3)	(4)	(5)	(6)
Outlier platform x Outlier category x Post	0.375**	0.366**	-0.188+	-0.065	-0.720+	-0.485**
	(0.06)	(0.12)	(0.11)	(0.26)	(0.38)	(0.14)
Outlier platform x Post	-0.504**	-0.532**	0.009	-0.385**	-0.187**	-0.474**
	(0.15)	(0.12)	(0.06)	(0.06)	(0.06)	(0.12)
Outlier category x Post	-0.458**	-0.569**	-0.086	-0.137	1.001**	0.910**
	(0.12)	(0.13)	(0.10)	(0.30)	(0.12)	(0.14)
Post	0.483**	0.672**	-0.069	0.345**	0.087	0.424**
	(0.13)	(0.12)	(0.06)	(0.03)	(0.08)	(0.09)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-Varying Controls	Yes	Yes	Yes	Yes	Yes	Yes
Outlier period-platform-category FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8420	4016	7659	3120	4786	4888
Outlier period-platform-categories	225	104	203	78	127	128

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variable is count of project entrants. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for nine film and video, four theater, eight music, three dance, five design, and five technology Kickstarter outliers are stacked to produce the sample in each column, respectively. An observation is an outlier period-platform-category-week. Categories where both within platform entry (in Table 1.9) and pledge effects are not statistically significant are not reported (art, comics, fashion, food, games, photography, and publishing). To facilitate convergence in all regressions except the design equation, year fixed effects were aggregated for each two year period.

Entry by category shows how sellers in different categories respond to outliers. Entry by film and video sellers increase relative to other categories following an outlier in the category (Column 1, $\beta = 0.133, p < 0.022$). This contrasts with entry in a number of categories, including theater (Column 2, $\beta = -0.136, p < 0.002$), music (Column 3, $\beta = -0.135, p < 0.001$), dance (Column 4, $\beta = -0.137, p < 0.001$), and design (Column 5, $\beta = -0.387, p < 0.001$) where subsequent similar entry decreases relative to other categories within Kickstarter.

The results by category for pledged dollars suggests interesting differences in the behavior of buyers. Outliers in the film and video and theater categories produce a positive effect on similar projects within Kickstarter (Column 1, $\beta = 0.375, p < 0.001$ and Column 2, $\beta = 0.366, p < 0.003$). The direction of these effects are consistent with the average results from Table 1.6. The within category impact of outliers in the design and technology categories are reversed (column 5, $\beta = -0.720, p < 0.057$ and column 4, $\beta = -0.485, p < 0.001$).

Whereas outliers in technology and design resulted in a substantial spillover in pledges to projects in other categories, outliers in the film and video and theater categories provided increased pledges to other projects in the same categories. One possible explanation for these results is that outliers in certain categories (i.e. the arts) attract pledges from backers who have a taste for art, and that those backers go on to support other arts projects on the platform. Outliers in other categories (i.e. hardware) attract pledges from backers who are essentially transacting a pre-purchase. Those backers have no predisposition for transacting with other projects in that category, and they go on to act as general Kickstarter or crowdfunding users.

1.6.2 Robustness Tests

I explore the sensitivity of the results to design decisions and assumptions, including definition of outliers, specification, and definition of a transaction. First, I implemented an alternative method for defining outliers, by identifying projects in the top five for dollars pledged by category each year. The definition resulted in 244 outliers on Kickstarter and 14 on Indiegogo. Results from Equations 1.4 and 1.5 with this alternative definition of outlier are reported in Table A.2 and are directionally the same as the primary results. Triple interactions for *entrants* and *pledged* are positive, but not statistically significant.

Returning to the primary definition of outliers, I account for two concerns in its implementation. I rerun the results excluding any outlier samples if the outlier's campaign was live on the platform at the same time as another outlier in the same category. The reason is to prevent repeating what is effectively the same sample when two outliers campaigns overlap. Results are reported in Table A.3 and are consistent with the main results.

Second, in the main specification, I allow an outlier to appear in the entry and pledged measures in the pre- and post-periods in other outlier samples. To test whether the results are dependent on the presence of outliers in any sample, I run an alternative sample that excludes any of the outliers from any weekly counts of entrants or dollars pledged. Results are reported in Table A.4 and are consistent with the primary results.

To test the dependence of the results on the model specification, I estimate the results using OLS with fixed effects with *ln entrants* and *ln pledged* as the dependent variables.[15] Results are reported in Table A.5 and are consistent, except the estimate of the triple interaction in the *ln pledged* pledged regression is positive, but not statistically significant.

The use of *pledged* as a measure of liquidity may raise to concerns. First is the assumption of timing of pledges. Due to data constraints, all pledges were assumed to arrive on the final day of of a project's campaign, but in reality, pledges accrue over the duration of the project. Second, for fixed funding model projects, pledges are not actually transferred to projects unless the project meets its goal. Because liquidity is dependent on the actual flows of capital, rather than the promised flows, including pledged dollars that were ultimately not transferred may not accurately reflect liquidity. To test whether the results are sensitive to the timing and transfer of pledges, I estimate a model where I only include dollars that were transferred at the conclusion of the project (as measured by *transferred pledged*). This measure accounts for both the timing problem, because transfers take place on the final day of the campaign making the assumption more closely tied to actual behavior, and the measurement problem. The results using this alternative definition of pledged capital are reported in Table A.6 and are consistent with the primary results.

Finally, sensitivity to the dependent variables is also considered. One concern is the use of Poisson estimation for a dependent variable that measures money. To address this concern, I estimate the results for an alternative measure of transactions, using *backers*, defined as the number of individuals that committed pledges for projects each week and the results (reported in Table A.7) are consistent with the *pledged* regressions. I also estimate the results only for weeks when *entrants* and *pledged* values are greater than zero to test whether the results are driven by periods when the platforms were not active and the results (reported in Table A.8) are consistent.

[15] Both variables are logged after adding one to include weeks with no entry or pledges.

1.7 Discussion

This paper determines the impact of high performing outlier sellers on a platform's growth and liquidity when platform competition exists. After the arrival of an outlier, entry disproportionately benefits the platform where sellers experience higher utility, after considering the outlier's impact on the host platform's utility. The average effect is moderated by heterogeneity in the sellers' cost sensitivity to the outlier. Moreover, within the platform, outliers in certain product categories are followed by a positive increase in entry and transactions for similar projects, whereas outliers in other categories created spillovers outside of that category.

In the strategy literature, market entry and the relationship between existing sellers and the subsequent entry strategies are well-studied phenomena (Fudenberg & Tirole, 1984; Lieberman & Montgomery, 1988). Certainly, high performance outliers are likely to affect market entry. The success of a product in a new product category, for example, might spur follow-on entry. Would-be entrants consider the prospective value of entering, given a growing market and competition with an incumbent firm, and compared to the outside option of not entering.

Entry in two-sided markets adds an additional dimension to the entry decision, because often times there are competing platforms to choose among. The entry choice of users impacts the equilibrium arrived at by the two platforms. Entry is not only a function of the market environment, but also of the platform's structure and rules. This paper extends the entry literature into two-sided markets.

The important role of platform rules is implicit in the impact of outliers on platforms. The rules of the platforms significantly impact their respective paths towards equilibrium, especially for platforms in a growth stage. Given there are likely multiple equilibria among competing platforms (Ellison & Fudenberg, 2003), outliers (and other events on the platforms that affect the growth) influence the path on which both platforms compete, and thus, which equilibrium is ultimately reached. That influence, in turn, is dependent upon the regulatory regimes of both platforms.

1.7.1 Managerial Implications

How should a platform manager incorporate outliers into developing a growth strategy? The results suggest that understanding the user mix within the platform would contribute to discerning when outliers would benefit the platform overall, or some subset of users. By considering the likely differences in kinds of prospective outliers, a manager could seek out sellers that are potential high performers to facilitate the appropriate spillovers within the platform.

Managers may also consider the rules of their platforms, and those of competitors, with respect to the types of users arriving on the platform. To the extent that existing and prospective users on both sides of the platform respond to outliers and other types of heterogeneity, and the response is shaped by the rules of the platform, the rules of the platform help guide the path it takes during the growth phase. In general, considering user heterogeneity and the interacting dynamics within one side of a platform, across sides of a platform, and across platforms provides insight into potential success.

Given the rules of the platforms and the composition of sellers and buyers, outlier users may create temporary surpluses for the platform, in which case there may be incentives to deviate from equilibrium pricing to capture rents (Armstrong, 2006). Moreover, if the second order effect of an outlier seller is to drive subsequent sellers to a competing platform, managers may also consider pricing deviations to drive sellers to their platform immediately following the outlier.[16]

1.7.2 Implications for Crowdfunding

It is worth considering the implications of the findings for the crowdfunding setting, given its increasing prominence as a phenomenon and area of academic study. For the two largest rewards-based crowdfunding platforms, Kickstarter and Indiegogo, the arrival an outlier on either platform was followed by a disproportionate increase in entry of subsequent sellers on Indiegogo. That entry was driven by marginal sellers.

One explanation for the relatively higher entry on Indiegogo relates to the regulatory differences on the two platforms. Kickstarter was historically more closed in terms of types of projects which could could raise funding and its fixed funding requirement calls for projects to reach its goal prior to receiving any funds. Project entry was more restricted due to the available categories of projects that were accepted and the minimum requirements project creators needed to meet. The fixed funding structure creates a collective investment decision among backers and allows for initial backers to pledge capital with the knowledge that pledges are committed once the goal is reached. To the extent that pledges later in the project are dependent upon early pledges, the fixed structure allows for this momentum investing to occur. It also allows for the goal to convey a more meaningful signal to prospective backers—goals set too low relative to the aims of the project could be considered unrealistic, and goals set too high would not be met thus freeing backers from committing to overly ambitious projects. Alternatively, the selection into a fixed funding regime may signal creator quality, where those users might have private information about their quality or

[16]One might imagine promotional pricing where the platform is 'celebrating' the success of the outlier by offering discounted prices to subsequent sellers.

confidence in raising funding, or are users who are better able to estimate the costs associated with their project. Indiegogo, by contrast, has been a more open platform for types of projects allowed on the site, as well as the choice of fixed or flexible funding. With flexible funding, the goal no longer conveys the same meaningful information to the backers, because creators could set higher goals (to appear realistic) without incurring any substantial costs, namely the risk of not acquiring the funds. In this regime, initial backers might be more hesitant to pledge to a project because their funding might go towards a project that experiences a shortfall when its goal is not reached. This precludes the possibility of deriving momentum from later backers.

These differences could be interpreted as the regulations for Kickstarter being 'buyer friendly' and those for Indiegogo being 'seller friendly.' While not the focus of this paper, Kickstarter arguably hosted almost all outliers due to its buyer friendly rules of capital raising. Initial investors could submit pledges with the confidence that their pledges would not be binding if the project did not receive its goal. Thus, more projects would thus receive those initial pledges, giving projects on Kickstarter a higher chance of exceeding or greatly exceeding their goals. Indiegogo's project friendly rules, then, allowed it to capture the subsequent entrants attempting to use crowdfunding to raise capital.

1.7.3 Conclusion

How platforms grow and compete partly depends on the effect that heterogeneous users have on subsequent activities of the platform. The mix of users that arrive on the platform and their impact are dependent upon the rules of the platform. The choice of a more open or closed regime by a platform will impact its growth and competitive positioning. Under certain regulatory regimes, a positive shock on one platform may be mitigated by the response of subsequent users adopting or transacting on the competing platform. Awareness of relationship between platform growth, rules, and user mix and behavior are important features in the management of a platform.

2

Social Media, Loyalty, and Organizational Performance

2.1 Introduction

Social media as a phenomenon is having a substantial impact on both management research (Aral et al., 2013) and managers. Questions such as understanding how firms manage social media (discussed further in Chapter 3) and its impact on performance (Rishika et al., 2013) are increasingly important as organizations interact more with consumers and other stakeholders on the medium. We explore the relationship between social media and organizational performance and ask how loyalty expressed on social media is associated with performance outcomes.

From the perspective of managers, organizations have access to increasing levels of data that is created both within and outside the organization's boundaries. For organizations with fewer resources (Wernerfelt, 1984) or a lack of capability to incorporate and process external information (Cohen & Levinthal, 1990), the lacking of appropriate human and technology capital may act as a barrier to incorporating the high levels of data. How can organizations derive value from data, such as social media, with minimal investment?

We look at the relationship between Twitter activity for shows in their initial season and their performance. In recent years, the idea of 'social TV' has emerged, which involves viewing television content and concurrently posting to social media services about the television content. Twitter is one of the leading platforms in the social TV space, as the television industry and Twitter have become increasingly institutionally interdependent.[1]

We make two contributions in this paper. First, we theorize about the relationship between loyalty expressed on social media and organizational performance. We find that there is a positive relationship between social media loyalty and organizational performance. That relationship is even stronger for niche organizations or those delivering niche products. We also find that higher levels of initial following mitigates the relationship between new followers and performance for niche products. Our results suggest that social media effectively diminishes the distance between the organization's boundary and its customers and other stakeholders, by allowing for these non-transaction interactions. Managing interactions with these outside parties is then a question of strategic importance as consumer behavior is not limited to purchases, but also communication and alignment with the organization and other stakeholders. This networked following represents a strategic asset, or resource, of the firm (Shankar & Bayus, 2003) that influences organizational performance.

Second, we provide evidence of how social media can be employed by resource and attention constrained organizations (Ocasio, 1997) in a straightforward manner to better understand future performance. Moreover, we show that in most instances, social media data performs better than general search data at predicting outcomes. We extend the growing nowcasting literature by utilizing social media data produced by both organizations and consumers to predict organization-level outcomes. The ability to use publicly available social media data to predict organization-level outcomes has important competitive implications. Organizations with fewer resources can improve the assessment of their goods and services. They can also observe the social media activity of their competitors and respond in both their organization and social media strategies.

2.2 Literature Review

The impact of digitized "social interactions" (Godes et al., 2005) on organizations and organizational outcomes has been studied in various aspects in the management literature. Online social interactions—or user generated content or digitized word of mouth (Dellarocas, 2003)—have been related to different organizational outcomes. The relationship between online reviews and

[1]Twitter's efforts to promote social TV include providing networks with guides on integrating Twitter into show content, acquisitions of television social media analytics firms, and its partnership with Nielsen to provide Twitter television ratings.

sales outcomes have been studied for books (Chevalier & Mayzlin, 2006), movies (Duan et al., 2008; Chintagunta et al., 2010), and video games (Zhu & Zhang, 2010). Less structured commentary and discussion, as occurs in message boards and forums, have been studied as well (Bickart & Schindler, 2001; Godes & Mayzlin, 2004; Sonnier et al., 2011).

More recently, social interactions online are taking place on social media, a phenomenon that has drawn recent academic attention (Aral et al., 2013). Certain types of social media are distinct from prior incarnations of social interactions in two main ways.[2] First, organizations themselves, and individuals within the organization, systematically participate on social media platforms. Consequently, customer interaction with firm social media efforts has been shown to positively impact the intensity of in-person interactions (Rishika et al., 2013) and consumption (Goh et al., 2013).

Second, the nature of the production of social media as real-time and streaming causes this kind of communication to fall within the general term of 'big data' from both the researcher's and the manager's point of view. Big data, including social media and other aggregated digital data, has been widely used in prediction papers, and in a particular subset of those papers called nowcasting. Giannone et al. (2008) is one of the earliest papers to use the term 'nowcasting' to mean applying intra-period data to improve forecasts of periodically released economic indicators. Much of the work since has investigated the predictive power of search data online in a variety of settings, including public health (Ginsberg et al., 2009), macroeconomic indicators and industry performance (Choi & Varian, 2012; Wu & Brynjolfsson, 2013), and product sales (Goel et al., 2010).

Social media as a source of data has been used to predict the stock market (Bollen et al., 2011; Nagle, 2015), electoral outcomes (Tumasjan et al., 2010), natural phenomena (Achrekar et al., 2011; Lampos & Cristianini, 2012), and firm sales (Asur & Huberman, 2010). For a survey of how social media has been used to predict various outcomes (some of which are nowcasting models), see Kalampokis et al. (2013).

Our paper continues the line of research about the relationship between digitized social interactions and organizational outcomes. We focus on the declared loyalty of users to organizations and their service offerings, a construct that is complementary to, but different from, most types of engagement studied in the literature. We theorize the relationship between loyalty and performance and use that as the basis for creating a parsimonious nowcasting model. Thus, we extend the nowcasting literature to social media and organizational outcomes, by demonstrating the value of nowcasting in providing insight into micro-level future performance. Moreover, we show how social media data compares to search in predicting organizational level outcomes.

[2]See Section 3.2 for more information on the typology of social media.

2.3 Hypotheses

We first hypothesize about the relationship between loyalty on social media and organizational performance. Then, we discuss moderators of that relationship based on characteristics of the organization, its products, and its prior social media activity.

2.3.1 Social Media Loyalty

We base our theory about the relationship between social media and organizational on Hirschman (1970). He argued that loyalty to an organization mitigates the likelihood of consumers choosing to cease consumption ('exit') amid concerns of declining quality. Rather, they would disproportionately attempt to communicate their grievances with the organization ('voice').

On social media, consumers can directly and publicly declare their loyalty to an organization, so we start with the above premise and adapt given the new medium of interaction that social media represents. We highlight four characteristics about social media that would further strengthen the relationship between loyalty and organizational performance. First, the core function of social media is a platform for the production and sharing of information and communication. Expressions of loyalty on a platform that is designed for communication further reinforces the tendency that loyalty promotes voice. Second, voice on social media does not necessarily have to be limited to occasions where there are decreases in quality. Rather, consumers can express positive and neutral sentiments, and thus influence others. Third, social media is a forum where the organizations themselves are often participants, so expressions of voice have a positive likelihood of being responded to by the organizations. Finally, expressions of loyalty are observed by other members of social media, thus allowing for the creation of *persistent* communities around an organization or its product.

Expression of loyalty on social media engenders a community around an organization or product, increasing engagement among social media users and between users and the organization. Increased engagement results in stronger relationships (Rishika et al., 2013) and increased performance (Goh et al., 2013). Taken together, the interaction between loyalty, community and engagement implies that increases in loyalty would be associated with increases in the organization's performance.

Hypothesis 1 (H1): Loyalty Increases in loyalty to an organization on social media are associated with increased organizational performance outcomes.

2.3.2 Matched Organizations and Niche Products

Political science theory of interest groups contends that lobbying efforts are likely to coalesce around interests that are held by a concentrated minority (Mitchell & Munger, 1991). Given a plurality of interests across individuals, those with the most concentrated economic interest in any issue are the most likely to be active and responsive to that interest. We hypothesize the same mechanism holds in the impact of communities around organizations and products on social media.

In most industries, organizations are differentiated along many customer dimensions, including geography, socioeconomic status, and taste. As a result, the impact of social media on organizations is likely most salient when consumers of the organizations goods are matched with users of social media. Given the heterogeneity in firms, social media is most likely to be impactful and related to outcomes of those organizations that are most similar to the users of social media

Hypothesis 2 (H2): Matched organizations Increases in loyalty to an organization on social media are associated with increased organizational performance outcomes, particularly for organizations whose customers match more closely with users of social media.

Product categories that target concentrated sets of individuals are more likely to be impactful than those that target the general population. Those users have a more concentrated interest, facilitating community formation and strengthening the community. Moreover, those users are more likely to have an interest in the outcome and performance of the product–they know their voice counts most heavily in those instances, so they are more likely to express loyalty, and interact with the organization along the dimensions described above than in the case of a product that targets the general population.

Thus, loyalty matters more for products that target niche populations when considering organizational outcomes.

Hypothesis 3 (H3): Niche Categories Increases in loyalty to an organization on social media are associated with increased organizational performance outcomes particularly for niche products.

2.3.3 Initial Levels of Loyalty

Next, we consider the importance of the initial levels, or 'stock,' of loyalty (rather than the ongoing 'flows') on organizational performance. Two countervailing effects that we call the 'implanted network effect' and the 'lead consumer effect,' impact how initial levels of loyalty affect the relationship between subsequent changes in loyalty and performance. We disentangle these effects by hypothesizing conditions in which one effect would dominate the other, based on whether the product is a niche good.

Users on social media experience network effects–their utility of participating is positively increasing in the number of other users, with whom they can produce content for, interact with, and consume more and varied content (Katz & Shapiro, 1985) . To the extent that there is a subset of users that coalesce around an organization or product, the same network effects will apply locally to that subset of users (Sundararajan, 2007). For example, if there is a set of individuals that enjoys discussing Stata on social media that those users utility will increase with each additional social media user that enjoys that topic of interaction. Social media effectively strengthens a product's network effects (Dou et al., 2013) or implants network effects on an organization or product that does not inherently possess such characteristics. Thus, any organization or product can potentially capture the benefits of network effects from social media by harnessing latent interest in that organization or product.

Social media acts as a complementary offering for any organization or product, providing additional value to consumers, as they can consume the product and engage with the network of users that does the same. As more people declare loyalty to the organization over social media, those users derive benefit from engagement with one another and with the organization itself. The utility of subsequent consumers is benefited by the existence of the local network relating to the organization or product on social media (Shankar & Bayus, 2003). In this case, then the larger the initial indication of loyalty, the larger the social media network effects and the better it would be for the organization.

This contrasts with the high marginal impact of early consumers for certain goods. In the innovation literature, lead users are early adopters of an innovative good that derive high utility from the good's consumption to solve a problem that most users generally have (Von Hippel, 1986). I apply the idea of differentiating among timing of consumption more broadly to identify lead consumers that are the most passionate consumers of a product. For many goods, early, or lead, consumers typically derive the highest utility from its consumption, and thus are the most likely to be interested in the product and its success. These users are also most likely to be engaged with other users, future users, and the organization about the organization or product. These users act as the seed users to generate the above described network effects. Through these many functions, lead users can act as bellwethers of a product's success.

On social media, lead consumers can be observed because of the early timing of their declarations of loyalty to an organization. Those consumers, then would have the highest marginal impact on the organization and the community. So incremental followers when the initial level of following is low are lead consumers and thus have the most marginal value to the organization. In this case, the smaller the initial indication of loyalty, the larger the value of the lead consumers and the better it would be for the organization.

Under which conditions would the positive effect of a larger and smaller initial following dominate? We posit that in conditions where there is a small community, the marginal value of the next follower is much higher. Based partly on our discussion of niche categories above, those narrowly focused communities are more likely to form around niche product categories. For these categories, where social media users are more likely to coalesce, the earliest users are its most ardent supporters. Thus, marginal declarations of loyalty when the initial level of loyalty is low are likely to be the most impactful on the network, for the organization, and for performance. For those niche product categories, then, the relationship between increases in levels of loyalty and performance is likely to be stronger when the initial following is smaller.

For non-niche products, the value to expressing loyalty and community participation would be larger, if the initial community were larger. The value of marginal users in the case of non-niche products is to strengthen the network effects, rather than the passion of their voice. Thus, for non-niche products, the relationship between increases in loyalty and organizational performance would be more pronounced when the initial level of loyalty is larger.

Hypothesis 4 (H4): Implanted Network Effects and Lead Consumers The moderating effect of initial loyalty levels on the relationship between increases in loyalty to an organization on social media and organizational performance is negative (stronger at lower levels) for niche organizations and positive (stronger at higher levels) for non-niche organizations.

2.4 Data and Empirical Approach

Our dataset is a weekly panel of 28 primetime shows premiering in the Fall 2013 television season across six networks. We collect weekly performance and daily social media data for each show. Table A.9 provides a listing of all shows in the dataset and Tables 2.1 and 2.2 provide summary statistics and correlations of our sample, respectively. We match the television episode data with social media data (described below) and our resulting dataset spans from the earliest premiere on September 16 to December 16.[3]

2.4.1 Twitter and television

We acquire our social media data from Twitter, a "microblogging" service launched in 2006. Twitter effectively acts as a newswire for all users (individuals or organizations), who can publicly broadcast 140 character messages, called tweets.[4] Users can select which accounts they wish to follow, effectively creating a personal feed of tweets from those selected accounts. Twitter ranks amongst the top 10 of most visited websites in the United States, with 284 million monthly visitors (Twitter, 2015) and a billion tweets produced every two to four days (Tsukayama, 2013).

Users post nearly anything including thoughts and opinions, references to online links, and replications of other posts (called retweets). Users also post tweets directly to other users, resulting in a publicly observable 'conversation.' Tweets also typically include hashtags ('#') on specific words or phrases, effectively creating a publicly observable conversation topic that other users can then include in their tweets.

The institutional setting in which we study social media performance is television. We collect tweets from and about each show for the six days prior to the episode's airing on an hourly basis using the Twitter application programming interface (API). As an illustrative example, for the show "The Millers" we collect number of tweets and replies by the show account, 'themillerscbs.' To capture the discussion on Twitter of the show, we compile all tweets and replies that include the show's account in the text (i.e. 'themillerscbs'), the primary hashtag associated with the show (i.e. '#themillers'), or the show's exact name (i.e. 'the millers').

We also collect data about the show's Twitter account profile. Included in the profile information is the number of accounts the show follows (called friends) and the number of accounts that follow the show (called followers).

[3]Many shows continued running the season into 2015, but we limit our analysis to 2014 due to constraints in our collection of social media data.

[4]Users have the option to make their tweet history private, viewable only by approved users.

Table 2.1: Sample Summary Statistics

	Mean	S.D.	Median	Min	Max
Logged measures					
rating	0.34	0.55	0.35	-1.20	1.55
viewers	1.51	0.63	1.52	-0.07	2.83
followers	6.40	1.42	6.37	2.48	10.13
community replies	3.01	2.00	3.11	0.00	6.12
community tweets	4.46	2.34	4.97	0.00	8.56
community hashtags	4.41	2.48	5.07	0.00	8.99
community account mentions	4.90	2.66	5.48	0.00	8.28
friends	0.73	1.09	0.00	0.00	5.57
show replies	1.29	1.58	0.00	0.00	4.61
show tweets	3.18	0.87	3.33	0.00	4.80
show hashtags	3.28	1.16	3.45	0.00	5.09
show account mentions	3.25	1.25	3.57	0.00	5.32
google trend	2.04	1.00	1.95	0.00	4.61
Raw measures					
rating	1.61	0.81	1.43	0.30	4.70
viewers	5.41	3.15	4.55	0.93	17.01
followers	1,605.98	2,858.34	581.50	12.00	25,008.00
community replies	78.45	107.68	21.50	0.00	454.00
community tweets	387.53	630.76	143.00	0.00	5,193.00
community hashtags	467.19	931.05	157.50	0.00	8,060.00
community accounts	738.07	950.26	240.00	0.00	3,952.00
friends	5.39	21.71	0.00	-1.00	262.00
show replies	12.61	23.16	0.00	0.00	99.00
show tweets	31.38	22.46	27.00	0.00	121.00
show hashtags	40.99	33.59	30.50	0.00	162.00
show accounts	42.76	38.57	34.50	0.00	203.00
google trend	12.85	14.64	7.00	1.00	100.00

Note: $n = 242$. An observation is a show-week.

Table 2.2: Correlation Table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) rating	1.00								
(2) viewers	0.92	1.00							
(3) followers	0.00	-0.08	1.00						
(4) community replies	-0.05	-0.00	0.36	1.00					
(5) community tweets	-0.05	0.00	0.36	0.92	1.00				
(6) friends	0.13	0.06	0.25	-0.01	-0.03	1.00			
(7) show replies	0.09	-0.00	0.13	-0.02	-0.04	0.50	1.00		
(8) show tweets	-0.10	-0.11	0.32	0.47	0.46	0.20	0.34	1.00	
(9) google trend	0.08	0.03	0.78	0.24	0.22	0.05	-0.16	0.06	1.00

Note: $n = 242$. An observation is a show-week.

2.4.2 Description Of Measures

Organizational Performance

Our television performance data is tabulated and released by Nielsen, a company that provides market data primarily for the television industry. To assess television episode performance, Nielsen calculates its rating score as the percentage of total television households watching a given episode. The firm also collects total viewership data for each episode.

Our data consists of Nielsen ratings and viewers for each television episode collected from the website TV by the Numbers[5] Our primary measure of organizational performance is the Nielsen television ratings, which we log to arrive at *ratings*. As a robustness test, we also look at the total number of viewers to compare absolute viewership to the market penetration measure that ratings represents. We take the log of number of viewers in millions to arrive at the *viewers* measure.

[5] The website for TV by the Numbers is `www.tvbythenumbers.com`.

Independent Variables Of Interest

Loyalty On Twitter, users have the choice to select other accounts they choose to 'follow.' Following another account is a one-sided interaction (it does not need to be reciprocated, nor does it need to be confirmed) that places the feed of tweets from that account onto the focal user's feed of tweets.[6] As such, it is an indication of interest in the content produced by that account. Moreover, the action of following another account is not costless, as it does consume real estate on the user's personalized feed. In the limit, if a user chose to follow every other Twitter account, that user's feed would become congested and ineffective. As a result, we define this feature on Twitter as being a measure of loyalty–it demonstrates the user's interest in the information being provided by the account being followed. Thus, we define *followers* as the log number of new followers of the show's Twitter account during the six days prior to the focal show-episode air date. A median 624 new followers (unlogged) accrued to accounts during a show-week.

Organizational Focus We identify matched organizations as the television networks whose demographics largely mirror Twitter's. Among the networks represented in our dataset, we identify one network, the CW, as being skewed towards a younger audience, similar to Twitter (Duggan et al., 2015). We thus define *matched network* as a binary variable that equals one when the show airs on the CW network, and zero otherwise. Based on this definition, there are three shows that air on a matched network.

Product Focus Product category type is based on the genre of the television show in this setting. Genre data for the shows is obtained from Wikipedia. In our data, we observe nine genres and we identify five as being directed to more niche audiences. Specifically, we define *niche genre* as a binary variable that equals one when the genre of the show is fantasy, game, horror, period, or scifi. The variable equals zero when the genre is action, crime, sitcom, or drama. We have a total of seven shows that are in niche genres.

Initial Loyalty To assess the initial loyalty of a television show, we look at the total number of followers acquired one day prior to the airing of the premiere episode. The median number of followers prior to the premiere was 3,235 and ranges from 418 to 112,860. We then divide the number of initial followers into quartiles to arrive at our *initial followers* categorical measure.

[6]Another aspect of following an account is that it allows that account to send direct, private messages to the user. While we do not have any data on this, anecdotal evidence would suggest that this is not the primary means of communication by the television show accounts in our study.

Table 2.3: Description of Social Media Variables

Variable	Count of	Action: Performed by	Action: Directed to
followers	New followers	Community	Show
community engagement	Retweets (Hashtags, Account mentions)	Community	Community
community tweets	Tweets	Community	Publicly broadcast
friends	New friends	Show	Community
show engagement	Retweets (Hashtags, Account mentions)	Show	Community
show tweets	Tweets	Show	Publicly broadcast

Social Media

Using the above data sources and data collection strategies, we derive five additional measures for the activity of both the focal television and for the Twitter community activity with respect to the focal show for the six days prior to the show's airing. All variables are calculated as the natural log of the raw measures (after adding one) and are summarized in Table 2.3.[7]

Community Activity In addition to assessing number of accounts that follow the show account, we capture other measures of the communities broadcasting and engagement behaviors. First, *community tweets* is the count of all non-reply tweets that include the show account in the tweet. The variable *community replies* is the total number of replies made by all users that include the television show in the text.

[7] In addition to the variables described in this section, we derive additional measures of show and community engagement based on the content of the tweets. All four measures are logged after adding one. First, we count the number of other accounts included within a tweet as a measure of the number of other users engaged by messages from the show, in *show account mentions*, and from the community, in *community account mentions*. Second, we count the number of hashtags included within a tweet as a measure of the number of conversations engaged by messages from the show, in *show hashtags*, and from the community, in *community hashtags*. These additional measures of engagement are included in a robustness test in Table A.13.

Show Activity We track the change in accounts the focal television show account chooses to follow during the six days prior to the episode air date, as measured by *friends*. We next aggregate all of the shows engagement with the community in *show replies*, which is the sum of its replies. Finally, total number of original non-reply tweets composed by the account is captured in *show tweets*.

Show Specific Secular Trends

One concern of our independent variable of interest, *followers*, is that the measure is merely capturing the secular popularity of a show over time, rather than a deliberate expression of commitment. We control for this by including the one period lag of the performance dependent variable.

We also include a time-varying measure of search popularity of each show over time. We include the show's Google Trend index each week, a score from 1 to 100 of the relative search frequency of one or more search terms or categories over a set period. The index is set such that the most popular search term-week is 100, thus providing relative search for all terms over the search period. The most popular search term-week was for "Agents of S.H.I.E.L.D." during the week of September 22, its premiere week. Each search term is assigned a Google search category to allow for variations on specific search terms to be attributed to the appropriate show (e.g. searching "brooklyn 99" or "brooklyn nine nine").[8] We log the weekly Google Trend index values to arrive at our measure *google trend*. Because the data is aggregated to the week level, we include a one week lag of *google trend*, so as not to include search data for days after a show aired in the regression.

2.4.3 Empirical Approach

Our data consists of weekly panel data of the performance and social media activity related to television shows. Both visual inspection of performance over time (as presented in Figure A.1 to A.5 for each show) and statistical testing[9] indicate the presence of first-order autocorrelation in the data. That is not surprising, given that outcomes are realized relatively frequently and that, intuitively, the performance outcome of a television show one week is a function of who watched the prior week.

[8] The only exception to the category search was for the television show "Dracula." No search category was identified by Google, so we include the index values for the search term "dracula nbc."

[9] We show the results from estimating a fixed effects OLS model in Table A.10 with show, calendar week, and age (in weeks) fixed effects. For each model, a test of first-order autocorrelation indicates the presence of autocorrelation.

As a result, we estimate the following model of organizational performance:

$$Y_{it} = \beta_0 Y_{it-1} + \beta_1 X_{it} + \nu_i + \gamma_t + \varepsilon_{it} \tag{2.1}$$

where Y_{it} is our organizational performance outcome of interest and X_{it} is a set of covariates, ν_i is an organizational fixed effect, γ_t is a time fixed effect, and ε_{it} is an idiosyncratic error term.

In our regressions, Y_{it} is the performance outcome of interest, *rating*.[10] Included in X_{it} is our variable of interest, *followers*, which measures the log number of new followers of the show's account during the six days prior to the show's airing, and five additional social media measures. There are three to capture actions by the focal show, *friends*, *show replies*, and *show tweets*, and two measures in addition to *followers* to capture community actions, *community replies* and *community tweets*.

Similar to Zhu & Zhang (2010), we interact *followers* with a fixed characteristic of the show, *matched network*, *niche genre*, and quartiles of initial followers. For those models, we separately include each interaction. Note that because the characteristics do not vary over time, those variables are considered part of the organizational fixed effect and not separately included in the model. We also include a one period lag of *google trend* as a control.

2.5 Results

In this section we consider the empirical support for the hypotheses described above. We also develop a prediction model and consider the nowcasting implications of the model.

2.5.1 Primary Results

In the dynamic panel model described in Equation 2.1, we mechanically introduced endogeneity between the lagged dependent variable, Y_{it-1} and the organizational fixed effect, ν_i (Nickell, 1981). Thus, we employ Arellano & Bond (1991)'s GMM estimation approach with instrument variables to address the endogeneity.[11] Each of our regressions includes *skipped week*, an indicator variable

[10] We run a robustness test looking at logged number of viewers in millions as the dependent variable, *viewers*. The results are discussed in more detail in Section 2.5.3.

[11] Arellano & Bond (1991) specify a first-difference model (to address organizational fixed effects) instrumenting the first difference lagged dependent variable with prior levels of the dependent variable, $Y_{i1}, \ldots, Y_{it-2}$. Similarly, where X_{it} might be weakly exogenous to the lagged error, $\varepsilon it - 1$, instruments for the differenced independent variable are used, $X_{i1}, \ldots, X_{it-1}$. Where X_{it} is strictly exogenous, the variable instruments itself. In practice, the Arellano & Bond (1991) estimator is included in popular statistical packages. For example, in Stata, the procedure is implemented in the `xtabond` or `xtdpd` commands.

set to one for weeks where the episode air time was not consecutive. Standard errors are clustered at the show level. Table 2.4 presents our results testing the hypotheses. Column 1 includes only the control variables and fixed effects and is included for reference.

Table 2.4: Relationship Between Social Media and Ratings

DV: *rating*	(1)	(2)	(3)	(4)	(5)	(6)
Genre Subsample:					Niche	Not Niche
Community actions						
H1. followers		0.055*	0.052*	0.045*	0.384***	0.041*
		(0.02)	(0.02)	(0.02)	(0.07)	(0.02)
H2. followers x matched network			0.125***			
			(0.03)			
H3. followers x niche genre				0.069*		
				(0.03)		
H4. followers x initial followers						
second quartile					-0.322**	-0.006
					(0.12)	(0.02)
third quartile					-0.280*	-0.035
					(0.13)	(0.03)
top quartile					-0.392***	0.042
					(0.10)	(0.03)
community replies	-0.012	-0.013	-0.019	-0.016	-0.094	0.000
	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)	(0.01)
community tweets	0.002	0.002	0.005	0.003	0.067	-0.004
	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.01)
Show actions						
friends	-0.005	-0.004	-0.005	-0.006	0.006	-0.016**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
show replies	-0.002	0.003	0.005	-0.001	-0.010	-0.003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
show tweets	0.031*	0.018	0.017	0.022	-0.060**	0.025
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)

Continued on next page

Table 2.4 – *Continued from previous page*

DV: *rating*	(1)	(2)	(3)	(4)	(5)	(6)
Genre Subsample:					Niche	Not Niche
google trend(t-1)	0.156***	0.110**	0.113**	0.123***	0.127	0.111**
	(0.03)	(0.04)	(0.04)	(0.04)	(0.17)	(0.04)
rating(t-1)	0.141*	0.106	0.123*	0.097	0.338*	0.072
	(0.07)	(0.06)	(0.05)	(0.05)	(0.15)	(0.04)
Constant	-0.129	-0.316	-0.466*	-0.512*	-1.128*	-0.137
	(0.14)	(0.20)	(0.22)	(0.23)	(0.51)	(0.14)
Calendar Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Show-weeks	186	186	186	186	47	139
Shows	26	26	26	26	7	19
Instruments	176	181	185	185	48	140
Wald Chi-Square	948.0	2871.1	9609.7	6131.7	36.24	5255.9
Arellano Bond $\mathcal{Z}_1$	-3.809	-3.580	-3.568	-3.598	-2.154	-3.112
Arellano Bond $\mathcal{Z}_2$	0.482	-0.183	0.0269	-0.0860	0.599	-1.024

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Note: Models estimated using Arellano & Bond (1991) generalized method of moments (GMM) with standard errors clustered at the show level (in parentheses). The lagged dependent variable is instrumented by all prior levels until period $t-2$. Each social media measure is differenced and is instrumented by all prior levels of the covariate until period $t-1$. Presented are the Arellano & Bond (1991) test for autocorrelation, where the null hypothesis is no autocorrelation (model specification is supported when the first order test is statistically significant, while the second order is not). The dependent variable is the log of Nielsen's rating. Each regression includes a dummy variable, *skipped week*, that equals one for weeks when the show's air date was not consecutive.

Organizations, Communities, And Engagement

To assess the relationship between loyalty and organizational performance, we look at the coefficient on *followers* in Column 2 and find a positive and statistically relationship between changes in new followers on Twitter and television show ratings ($\beta = 0.055, p < 0.013$), providing support for H1. As an indication of the magnitude of the effect, a doubling in the number of new followers during the six days prior to the show's airing is associated with a ratings increase of 5.5% ($\beta * 100 = 5.4$). Given the median rating of 1.61, a 5.5% increase represents a ratings improvement of 0.089, which translates into just over an additional 100,000 television households (given approximately 115.6 million television households, calculated as change in rating / 100 * 115.6 million).

Heterogeneity in Networks, Genres, and Initial Followers

We now explore heterogeneity in ratings responsiveness to changes in levels of loyalty, based on the type of organization or product being offered and the initial followership of the show. Results for these interacted models are presented in Columns 3 to 6 of Table 2.4. We find that the positive association between increases in followers and increased ratings is stronger for shows aired on the matched network, the CW, and shows in niche genres (fantasy, game, horror, period, science fiction).

The interaction of *followers* with *matched network* is positive, and statistically significant (Column 3, $\beta = 0.125, p < 0.001$), providing support for H2. Similarly, the interaction of *followers* with *niche genre* is positive and statistically significant (Column 4, $\beta = 0.069, p < 0.033$), providing support for H3. We graphically present the results for the above interactions with matched network and niche genre in Figures 2.1 and 2.2, respectively. The graphs visually present the stronger relationship between followers and ratings experienced by the matched network and niche genres, as compared to the appropriate other category.

To test H4, we divide the sample into two groups based on whether a shows is in a niche genre, and present the interaction between each quartile of *initial followers* (omitting the bottom quartile) and *followers* in two separate regressions, shown in Columns 5 and 6. Among shows in niche genres, we find evidence that levels of initial followers in the top three quartiles mitigate the relationship between followers and ratings (Column 5, second quartile $\beta = -0.322, p < 0.009$, third quartile $\beta = -0.280, p < 0.037$, top quartile $\beta = -0.392, p < 0.001$). Conversely, among shows that are not in niche genres, levels of initial followers in the top quartile strengthen the relationship between followers and ratings, though the increase relative to the bottom quartile is weakly statistically significant (Column 6, $\beta = 0.042, p < 0.095$). Taken together, we see that lower initial following levels strengthen the relationship between followers and ratings in niche genres, and higher initial following levels tentatively strengthen the relationship in non-niche genres, providing support for H4.

We present a graph summarizing these results in Figure 2.3. Each panel shows the differential trends between ratings and followers for shows in the bottom and top quartile of initial followers.

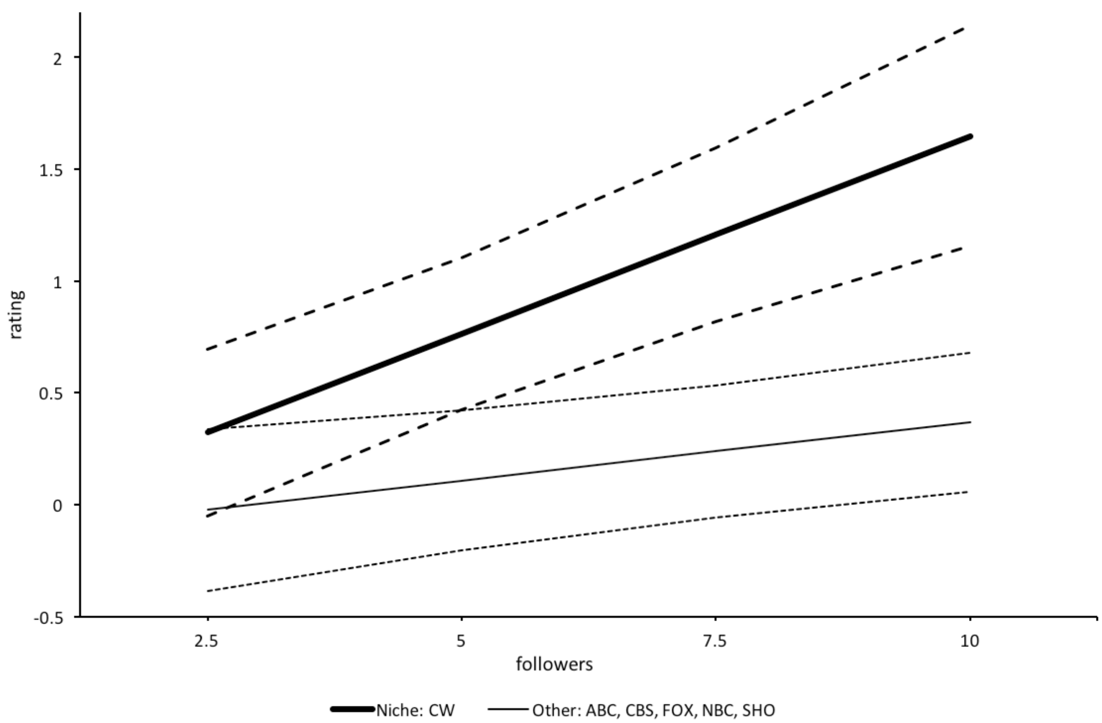


Figure 2.1: Differential Results Based on *matched network*. The above graph depicts the results from the interaction term presented in Column 3 of Table 2.4.

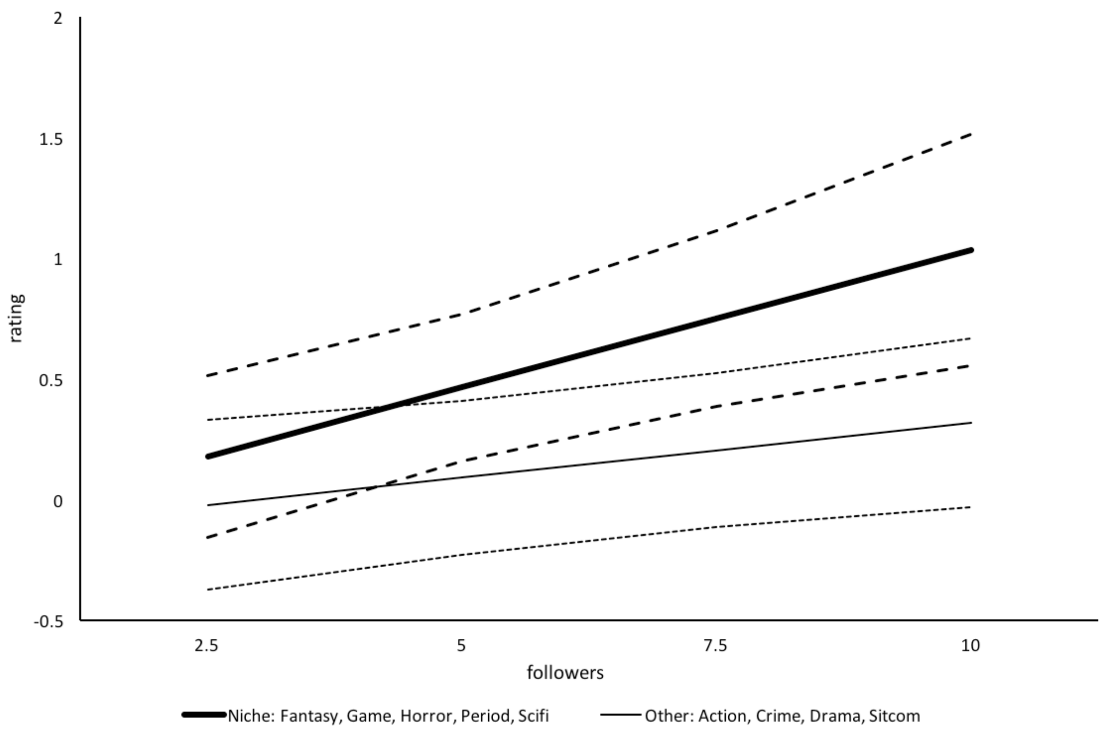


Figure 2.2: Differential Results Based on *niche genre*. The above graph depicts the results from the interaction term presented in Column 4 of Table 2.4.

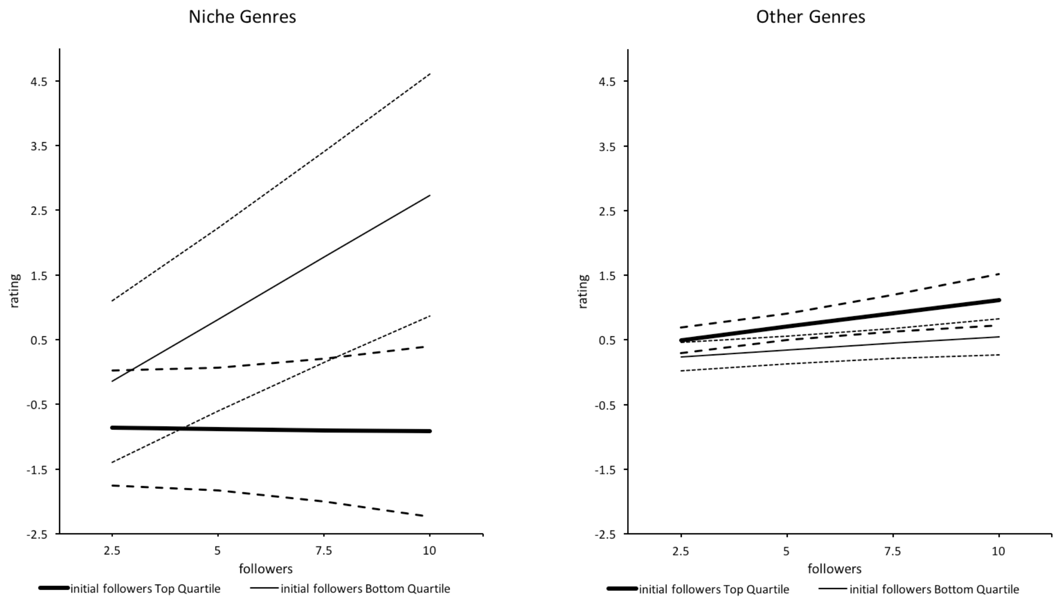


Figure 2.3: Differential Results Based on *initial followers*. The above graph depicts the results from the interaction term presented in Columns 5 and 6 of Table 2.4.

The panel on the left depicts the relatively stronger positive relationship between followers and rating for bottom quartile of initial followers compared to the top quartile in niche genres, while the panel on the right depicts the reversal of that relationship among shows in non-niche genres.

2.5.2 Constructing and Interpreting a Nowcasting Model

Thus far, in this section, we have looked at the theoretical relationship between expressions of loyalty on social media and subsequent organizational performance. We now take another approach based on the idea that social media activity in the periods between the realization of firm outcomes may provide additional insight during these "between periods." Thus, we take on an exercise in nowcasting firm performance outcomes to see whether activity on social media can provide more information to managers about the performance of their organization.

To do this, we build a parsimonious model using prior work in prediction contexts (Choi & Varian, 2012; Wu & Brynjolfsson, 2013) as a foundation for our organizational setting. To begin, we employ one period autoregressive model to provide a baseline as follows:

$$Y_{it} = a_0 + a_1 Y_{it-1} + \gamma_t + \nu_i + e_{it} \tag{2.2}$$

The above specification includes a one period lag for the performance dependent variable of interest, *rating*.

We extend the above model to include social media data, X_{it}, as follows:

$$Y_{it} = b_0 + b_1 Y_{it-1} + b_2 X_{it} + \gamma_t + \nu_i + \epsilon it \tag{2.3}$$

Using the above specifications allow us to also compare various social media models employing Equation 2.3 to the baseline autoregressive model described in Equation 2.2.

Again, due to the first-order autoregressive structure with organizational fixed effects, our estimation of all prediction models follows Arellano & Bond (1991). We present the results of the prediction models in Table 2.5.

Table 2.5: Social Media Prediction Models of Show Ratings

DV: *rating*	(1)	(2)	(3)	(4)	(5)
Community actions					
followers			0.076***	0.083***	
			(0.02)	(0.02)	
community replies			-0.013	-0.013	-0.016*
			(0.01)	(0.01)	(0.01)
community tweets			-0.007		
			(0.01)		
community hashtags			0.010		
			(0.01)		
community account mentions			-0.002		
			(0.01)		
Show actions					
friends			-0.004		
			(0.01)		
show replies			0.001		

Continued on next page

Table 2.5 – *Continued from previous page*

DV: *rating*	(1)	(2)	(3)	(4)	(5)
			(0.01)		
show tweets			0.007	-0.013	
			(0.03)	(0.03)	
show hashtags			0.034	0.023	0.019*
			(0.03)	(0.02)	(0.01)
show account mentions			-0.023		
			(0.02)		
google trend(t-1)		0.121**			
		(0.05)			
rating(t-1)	0.248**	0.162	0.220***	0.196***	0.323***
	(0.08)	(0.10)	(0.04)	(0.04)	(0.07)
Constant	0.355**	0.063	-0.249	-0.272	0.272**
	(0.12)	(0.13)	(0.18)	(0.19)	(0.09)
Calendar Week FE	Yes	Yes	Yes	Yes	Yes
Show-weeks	186	186	186	186	186
Shows	26	26	26	26	26
Instruments	61	62	187	167	134
Wald Chi-Square	137.0	407.9	15307.7	498.9	305.6
Arellano Bond $\mathcal{Z}_1$	-3.290	-3.283	-3.570	-3.598	-3.659
Arellano Bond $\mathcal{Z}_2$	0.827	0.431	-0.0211	0.0572	1.052
MAE	0.321	0.364	0.341	0.351	0.290
MSE	0.158	0.208	0.182	0.193	0.130
RMSE	0.397	0.456	0.426	0.440	0.361

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Note: Models estimated using Arellano & Bond (1991) generalized method of moments (GMM) with standard errors clustered at the show level (in parentheses). The lagged dependent variable is instrumented by all prior levels until period $t-2$. Each social media measure is differenced and is instrumented by all prior levels of the covariate until period $t-1$. Presented are Arellano & Bond (1991) test for autocorrelation, where the null hypothesis is no autocorrelation (model specification is supported when the first order test is statistically significant, while the second order is not). The dependent variable is the log of Nielsen's rating. Each regression includes a dummy variable, *skipped week*, that equals one for weeks when the show's air date was not consecutive.

Column 1 presents the baseline model with lagged *rating* only. Our method for selecting a prediction model followed two steps, given the relative few number of covariates.[12] First, we assessed each variable separately in a univariate regression (presented in Table A.11) and identified four variables that were statistically significant at a p-value less than 0.200, *followers*, *community replies*, *show tweets*, and *show hashtags*.

In Column 4, we present a model that includes those four variables in a predictive model. Then, we identified the model with the lowest Mean Absolute Error (MAE) among all combinations of the four covariates.

That model is presented in Column 5 and includes two variables, *community replies* and *show hashtags*. The MAE of 0.290 represents a 9.7% improvement over the baseline model in Column 1. The MAE improves for 22 of the 28 shows in our model. We present this evidence in Figure 2.4, which shows the difference in MAE of the social media model in Column 5 and the baseline model in Column 1, by show. In Figures A.1 to A.5, we show the actual performance of the show and compare it to the baseline and social media predictions, on a show by show basis.[13]

We conduct an out of sample analysis by looking at a series of one step ahead predictions (Choi & Varian, 2012), beginning with predictions of week 5 ratings. We find that the MAE for all predictions from the social media model is 11.3% higher (i.e. worse) than the baseline. The lower relative performance is driven by early predictions in episode numbers 5 to 7, where the baseline model was superior. In the next five weeks (episode numbers 8 to 12), the social media model outperformed the baseline model four out of five times, indicating better performance later in the time series.

Comparing Sources of Nowcasting Data

Prior papers in nowcasting utilize the Google Trends Index as a source of real-time data to improve economic indicators of interest (Choi & Varian, 2012; Wu & Brynjolfsson, 2013). To provide a second comparison with our social media model, we used lagged values of the Google trend index to predict television show ratings in the subsequent period. Results for the Google Trends model are reported in Column 2 of Table 2.5. In aggregate, the model performs 14.0% worse than the baseline

[12]For models with many covariates a stepwise procedure would need to be implemented. We present a model with all ten covariates included in Column 3, for reference.

[13]Note that due to the use of first-differences, two shows, "Lucky 7" and "We Are Men," only have one predicted week and are thus not included in the individual show graphs.

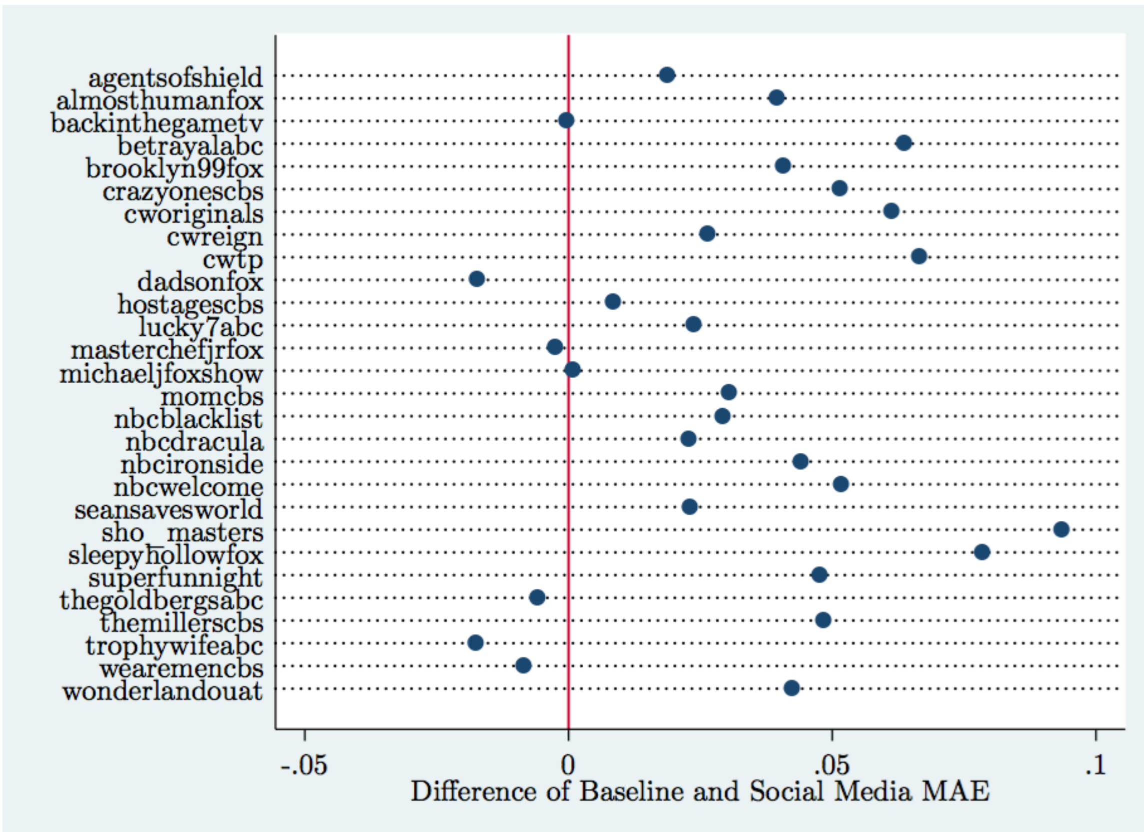


Figure 2.4: Comparative Performance of Social Media and Baseline Prediction Models by Television Show. Each line represents one television show and indicates the performance of the social media prediction model relative to the baseline autoregressive prediction model. Each point is calculated as the difference of the baseline model Mean Absolute Error (MAE) from the social media and Google Trend model, respectively MAE (Wu & Brynjolfsson, 2013). Values greater than zero indicate that the model performs better than the baseline at predicting *rating* for that television show.

model. Interestingly, when looking on a show by show basis, the Google trends prediction model demonstrates a substantially larger variance than the social media model, as shown in Figure 2.5. For nine shows, the Google trends model out performs the social media model, but for 17 shows, it performs worse than both the baseline and social media models (the performance of the social media and Google trends models are almost equivalent in two cases).

2.5.3 Robustness Tests

We check the robustness of our results to the choice of dependent variable as well as the empirical specification.

viewers As Alternate Dependent Variable

A show's television ratings are based on a calculation made by a private firm, Nielsen. We utilized the measure, as it is the primary variable of importance to television shows in making renewal decisions and to attract advertising revenue. However, we wanted to test the sensitivity of our results to the construction of the ratings measure, by looking at a direct measure of consumption, viewership. We construct the measure *viewers* by taking the natural log of the number of viewers, in millions.

Table A.12 reproduces the primary results in Table 2.4 using *viewers* as the dependent variable. Overall, the results corroborate the direction we observed in our outcomes of interest from our primary findings. Specifically, we find support for H2 (Column 2, $\beta = .131, p < 0.020$) and H3 (Column 3, $\beta = 0.093, p < 0.006$).

For H1, we find a positive relationship, but one that is not statistically significant (Column 1, $\beta = 0.030, p < 0.357$)). For H4, there are some inconsistencies across the quartiles of initial followers. Among niche genres, shows in the third quartile of initial followers in niche genres show a positive, though not statistically significant effect, counter to our main findings (Column 4, $\beta = 0.015, p < 0.878$). In non-niche genres, shows in the top quartile show a positive relationship, as with the main results, but not statistically significant (Column 5, $\beta = 0.013, p < 0.729$)

2.5.4 Alternative Measures of Social Media Interactions

In our primary specification, we use *community replies* and *show replies* to measure the level of interaction both Twitter users and television shows have with the wider Twitter community. We

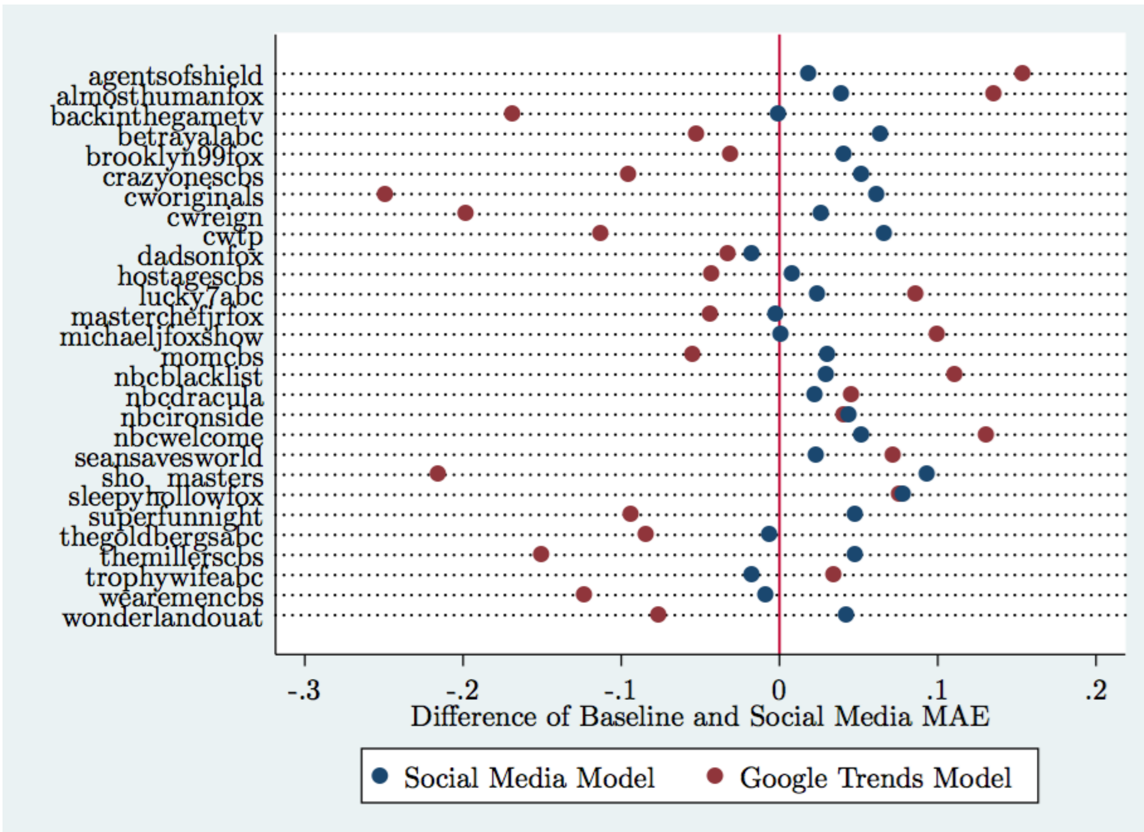


Figure 2.5: Comparative Performance of Social Media and Google Trend Prediction Models by Television Show. Each line represents one television show and indicates the performance of the social media and Google Trends prediction models relative to the baseline autoregressive prediction model. Each point is calculated as the difference of the baseline model Mean Absolute Error (MAE) from the social media and Google Trend model, respectively MAE (Wu & Brynjolfsson, 2013). Values greater than zero indicate that the model performs better than the baseline at predicting *rating* for that television show.

also have alternative measures of interaction, not based on the number of messages (or replies), but rather the content of the messages.

Specifically, we count the number of hashtags used in messages by both Twitter users and the television show as an estimate of the number of conversations messages in a given six day period were directed at. We create *community hashtags* and *show hashtags* as the log number of hashtags included in tweets posted by the community and the television show, respectively, during the six days prior to the next episode's airing. Similarly, we count the number of accounts mentioned in messages as an estimate of the number of people messages were directed to. We create *community accounts mentioned* and *show accounts mentioned* as the log number of accounts mentioned by the community and television show, respectively. Both of these measures are imperfect measures of interaction. These measures are proxies for the intensity in intent to interact with the community.

Given these additional measures, we reproduce our results from Table 2.4 using all these additional measures of community and show activity. The results are presented in Table A.13 and they are largely consistent with our primary findings, except for the results testing H4. In niche genres, the third quartile shows a negative effect like our main results, but just outside the bounds of statistical significance (Column 4, $\beta = -0.254, p < 0.069$). In non-niche genres, the top quartile shows positive effect like the main results, but it is not statistically significant (Column 5, $\beta = 0.038, p < 0.116$).

2.6 Discussion

We investigate how social media activity is related to organizational performance and find a positive relationship between increased indications of loyalty on social media and subsequent organizational performance. We also find that relationship is stronger for niche organizations and for niche products. Finally, we show that having a higher endowment of followers mitigates the relationship between loyalty and performance for niche products, but strengthens it for non-niche products.

We also consider social media as a source of data for a nowcasting model to predict organizational level outcomes. The improvement in the predictive power of the nowcasting model, relative to the baseline, implies that social media data provides important information that predicts subsequent organizational outcomes. Resource constrained organizations can utilize social media to have a better understanding of subsequent performance of their service offerings.

We highlight three implications of our findings. First, whereas much of the open innovation literature has focused on the erosion of organization boundaries due to technological innovations (von Hippel, 2005), we observe an apparent erosion of the *distance* between the organization's boundaries and outside stakeholders. The behavior of individuals on social media and interactions between organizations and social media communities are both associated with and predictive of increased performance outcomes. Thus, organizations are well served by considering their social media strategy and by understanding the value contained in social media information. Because of the importance of social media as a communication and broadcast platform amongst many consumers and stakeholders, organizations must pay special attention to these non-transaction interactions with consumers and other stakeholders.

Second, social media introduces characteristics of 'lead users' for any organization. For certain organizations, increases in following were more strongly associated with higher performance when initial following was low. This points to the idea that early users are of high value to organizations for their increased attention value derived by these lead users.

Finally, we demonstrate the application of nowcasting techniques to organization level outcomes and provide evidence that social media contains consequential information to those outcomes. Interestingly, our hypothesized variable, *followers*, did not contribute additional prediction power and was not included in our social media prediction model. This highlights the distinction between hypothesizing relationships and finding predictive power.

Though our findings and implications are likely easily generalized outside of our television setting, our study is limited by not providing a definitive causal pathway between loyalty on social media and organizational performance. Future work might investigate the causal effect of loyalty and other social media behavior on organizational performance. With respect to the nowcasting model, we can make claims about the applicability of our model to neither shows beyond their initial season, nor other organizational and product settings. Here, generalizability is a greater challenge. However, we point out that in the case of television ratings and social media, we are able to provide more prediction power using certain social media measures than merely historical performance alone. This indicates the value of social media towards understanding real-time performance changes over time. The models predicting other organizational level outcomes would likely need to be specified in accordance to differing contexts.

The mass adoption of social media by so many individuals and organizations carries with it interesting theoretical and managerial implications. How we think of networked goods, interactions with constituencies outside the boundaries of the organization, and the importance of real-time communications in our assessment of performance are all affected. Continued research on the management of social media will further distill its impact on the management landscape.

3

Organizational Management of Social Media

3.1 Introduction

Social media is an emergent phenomenon that is increasingly central to a number of managerial and academic activities. Social media is a broadly defined phenomenon, typically encompassing forums, blogs, ratings and reviews, social networks, and content sharing. Considering the subset of social media related to content sharing alone, it has impacted almost all content creation markets, from news to entertainment, as models of producing and distributing content shift toward platforms where single users, often individuals, produce content for their local networks, and in some cases, for a wider audience.

In academic research, social media's roots lie in user generated content and word of mouth (Godes & Mayzlin, 2004), or more generally "social interactions" (Godes et al., 2005), facilitated by technology and digitization (Dellarocas, 2003). Many aspects of social media research have been investigated including motivations and incentives for social production of media (Zeng & Wei, 2013), impact of social media on consumer behavior and organizational performance (Rishika et al., 2013), and impact of changes in engagement (Claussen et al., 2013).

With respect to how firms manage social media, the focus has been on how firms implement social media activities (Aral & Walker, 2011), seed and grow networks (Dou et al., 2013), or how firms communicate on social media (Miller & Tucker, 2013). These types of activities can be thought of as external management of social media. What has been less studied is how social media is internally managed within the firm, including how the firm organizes around social media, the evolution of business processes that result from social media, and strategic decisions related to social media, including the initial decision to participate on social media.

In this paper, I focus on how social media is managed by the organization, by focusing on the decision to adopt and the pattern of diffusion of social media within an organization. First, I briefly outline dimensions on social media formats differ to better categorize the wide scope of the phenomenon. Second, I interpret the management of social media as it relates to the organization's internal processes, learning, and adaptability by looking at organizational adoption, diffusion, and subsequent behavior on social media.

To study this aspect of social media, I look at how television networks and shows integrated Twitter into their respective organizations. First, I look at the adoption decision by individual shows and find that larger organizations with more existing resources more readily adopted social media. Moreover, television shows that started more recently—as social media increasingly proliferated and its relationship with television in particular strengthened—adopted social media more quickly. I also distinguish among network strategies of social media management by identifying three dimensions of the strategy: timing of adoption (i.e. early or late), rate of adoption, and centrality of management within the organization. Finally, I look at differential diffusion patterns of social media, based on cohorts of shows premiering in the same year.

3.2 Towards a Typology of Social Media

Research on social media covers a wide array of settings[1] including product reviews (Chevalier & Mayzlin, 2006; Archak et al., 2011), message boards (Das & Chen, 2007), social networks (Miller & Tucker, 2013), photo sharing sites (Zeng & Wei, 2013), and others. Many settings that fall under the rubric of social media have very different dynamics of platform, organizational, and consumer behavior. To better distinguish among types of social media, I present a basic typology of the phenomenon, providing some dimensions along which these settings might differ. Differences have theoretical and practical implications for researchers and organizations. I discuss three dimensions in this typology, namely the hosting service, types of participants, and the timing of the

[1]A few recent papers have provided a fairly thorough catalog of work done in social media. I refer readers to Aral et al. (2013), Kalampokis et al. (2013), and Wu et al. (2013) for an overview of research on the phenomenon.

content delivery.

Social media can be physically hosted by a centralized source (such as a corporate or individual blog) or by third party providers, including service providers and platforms.[2] When hosted by a platform, a participating organization must conform to the rules of that platform (Altman, 2015) and thus cede some control over its interactions with users. This contrasts with an organization hosting its own social media activities, for example, where it would maintain control over every aspect of their interaction with users and user interactions with one another. Moreover, operating social media on a platform implies that many principals of two-sided platforms or markets (Rochet & Tirole, 2006; Hagiu, 2009; Rysman, 2009)—with producers of content on one side and consumers on the other—would govern the operation of the service and the behavior of participants on the service.

Another distinction between the two hosting alternatives is that users would have to choose to engage with a social media offering provided directly by an organization, but on a platform, users participate on the platform (likely for reasons that are not related to the organization) and then engage with organizations that maintain a presence on the platform. Thus, on a social media platform a heterogeneity of user tastes, preferences, and types are present, whereas on social media hosted by the firm firm, users would have a specialized interest in interacting with that organization. The heterogeneity in users partly mitigates a selection effect that may arise from users who choose to engage on social media with a firm on its site versus users who are engaging with social media, some activity of which is with the organization.

A second characteristic that differentiates social media services is the types of participants that engage. On most services, individuals are the primary, and sometimes only, type of user, while in others, organizations participate at effectively the same level as individuals.[3] At the extreme, an organizational account is no different than any individual account.

Organizational participation on social media implies that expectations of responses from organizations by individuals discussing the organization, its brands, or its performance are more likely to be observed or responded to by the organization. Knowing that organizations are participating on a given social media platform might increase expectations of firm responsiveness, given the

[2]Many examples of platforms hosting social media are readily available and include include Twitter, YouTube, and Yelp. Examples of centralized hosting include the Lego Creator site, where users can create, share and discuss user designs and product reviews at a corporate store. An example that captures some elements of both types is a corporate page on a social media site, such as Facebook, where the corporation manages the content and interactions with users, but the content is ultimately controlled by the platform.

[3]Message boards are an example, where organizations typically do not participate, whereas on some product review sites the organization will participate to respond to reviews.

lower barriers to interact with users.

Organizational presence on platforms also exists on the same footing as individual user accounts. This means that the production of content originates from many users and is consumed by many users, rather than it being a more traditional model where the content is produced largely by a single entity and generally consumed (as it is with blogs, for example). Another characteristic of organizations participating on platforms is that communication between individuals and the organization are typically publicly observable by all users.

Finally, social media services differ in how they produced and interacted with with some being more real-time and others being more archival. Some social media sites act as real-time newswires, where historical releases can be located, but the primary mode of consumption is the live feed. This contrasts with archival types of services, where some interactions may occur in real-time, but a primary way of interacting with the content is in an asynchronous manner with search and traditional page visits being the interaction model. The tradeoff between the two content delivery are that real-time broadcast feeds provide a better sense of current information, whereas archival pages provide better historical access and search.[4]

3.3 Applying Theories of Strategy and Innovation to Social Media

3.3.1 Adoption

For individual users, social media is a prototypical network good. Users produce and consume content and utility is a function of the number of users (Katz & Shapiro, 1985; Liebowitz & Margolis, 1994). An individual's decision to adopt is a function of factors such as social ties (Fang et al., 2013), user heterogeneity (Sundararajan, 2007), current and future expectations (Katz & Shapiro, 1986) and others. When considered to be a question of innovation or technology diffusion, factors like network structure (Abrahamson & Rosenkopf, 1997), timing (Gort & Klepper, 1982), or the characteristics of the innovation (Rogers, 1995) have been considered. In all these cases, though users may be heterogeneous in terms of their network centrality, location, or influence, they are of the same type (i.e. individuals).

Organizations may differ from individual users by considering social media primarily as a means

[4]For example, a doctoral student looking for an answer to a question about Stata would go about it in different ways on Twitter and Stack Exchange, as examples of real-time and archival formats, respectively. On Twitter, she might pose her question with the #stata topic indicator and get a response by someone who follows that topic. That compares to StackExchange, where she might search the boards for keywords relating to her problem and find someone else who had a similar issue and the solution that was provided at that time and post a question if the search did not yield helpful results.

to produce and distribute news and content, and not to consume it For those users of social media, the adoption choice would be dependent upon different factors. For organizations acting as producers, social media is effectively a two-sided platform, with content producers, including organizations, on one side, and social media consumers on the other.

Utility is the sum of benefit from participating and transacting, less the sum of costs of participating and transacting (Rochet & Tirole, 2006). The benefit of transacting would increase with the number of social media consumers, because of the presence of indirect network effects (Armstrong, 2006; Parker & Van Alstyne, 2005). The costs of participation are the resources required to manage social media. As is the case in innovation environments, firms engage in varying levels of experimentation with different complementary services to promote their main goods to differing extents (March, 1991). Prior to social media being a 'dominant design' (Tushman & Anderson, 1986) for organizational interactions with outside stakeholders, firms must first have the capability to identify it as a prospective viable alternative to adopt and implement (Rivkin & Siggelkow, 2003). Upon adoption, the organization must have the resources to allocate to the implementation and develop the capabilities to properly integrate new efforts (Wernerfelt, 1995; Teece et al., 1997; Eisenhardt & Martin, 2000).

Organizational Size

One characteristic that is related to the utility of adoption is the size of the organization. Whereas small firms might derive more value from participating, larger firms would be more likely to adopt based on costs. Smaller firms derive a disproportionately greater benefit from participation on social media, particularly if there is a match between the social media audience and the service offering of the organization (see Chapter 2). Larger firms with more resources can more readily identify, adopt, and deploy them in new experimental ways, while small firms with fewer resources would likely need to focus on the execution of their current activities.

Organizational Inception

As social media matures, organizational adoption rates are likely to change as well. The context in which organizations originate impact its subsequent behavior and trajectory (Meyer & Rowan, 1977). Organizations that are established when social media is more salient among competitors and among the general public are more likely to organize its resources to integrate social media as part of its core activities. The increased salience may also be attributable to the actions of the social media platform. Doing so partly ensures that the organization will better fit with the environmental landscape that exists at that time (Levinthal, 1997).

Further substantiating the argument that recently initiated organizations will adopt social media at an accelerated rate is that older organizations may be less likely to do so. Older firms demonstrate inertia in their innovative processes (Henderson & Clark, 1990; Henderson, 1993) and are less likely to be technologically innovative (Balasubramanian & Lee, 2008). This anti-innovative bias on the part of older firms may carry into their adoption of innovations as well. From both perspectives, recent organizations would more readily adopt social media.

3.3.2 Diffusion

For organizations, participation in social media can occur at multiple levels. First, the organization itself can establish and maintain an account. Second, within the organization, departments or divisions may also have a social media presence. In addition, organizations may establish social media identities for separate brands or individual products. Third, individuals, including executives, managers, and employees may have accounts (that may have different levels of relevance to the organization). Organizing, coordinating, and managing across these levels is something organizations coordinate and execute upon differently.

Given the decision to adopt social media at sub-organizational levels, we can study how social media diffuses (Bass, 1969; Rogers, 1995) within an organization and infer its social media strategy from differences in diffusion patterns. An organization's timing, speed, and centrality of social media management can be distinguished by looking at differences in the rate of diffusion of social media within an organization. First, organizations where the various stages of implementation takes place sooner are early adopters. Second, the speed with which the diffusion takes place relates to the level of coordination and resources dedicated to the adoption and implementation of the organization's social media presence. Third, the level of centrality relates to the extent to which social media is managed at sub-organizational levels as opposed to the corporate level.

Similarly, diffusion of social media can be studied over time, as it grew among general users. While diffusion theory covers instances where successive iterations of a technology diffuse through the population (Norton & Bass, 1987; Bresnahan & Yin, 2005), I look at the diffusion of a technology through subsequent sets of susceptible populations.

3.4 Setting and Data

The social media setting for this paper is Twitter, which is typically described as a "microblogging" or "social networking" service within the larger context of social media. Twitter is a service that allows users to publicly broadcast 140 character messages, called tweets. Users have the option to select other accounts to include as part of a personal newswire of content being published by those specific accounts. Given the categories to distinguish between types of social media in Section 3.2, Twitter is a social media platform, where individuals and organizations participate in real-time broadcasts and communication.

In recent years, Twitter and the television industry have become increasingly interdependent, as evinced by a number of behaviors by each institution. First, Twitter has taken a number of actions to develop interdependencies with the television and other industries. The company operates content sites geared towards managers of television Twitter accounts, including Twitter for Business, Twitter Media, and The Twitter Media Blog. The latter two are aimed specifically towards media companies using Twitter as part of their content production portfolio. For television shows in particular, representatives from the firm meet with actors and other individuals associated with new shows to educate and reinforce the importance to the show and their personal brands to actively maintain a presence on Twitter. Second, to bolster the relationship between viewers, television shows, and social media, Twitter has made a number of personnel and firm acquisitions related to the television industry and analytics.

Concurrent with the efforts of Twitter, television shows also frequently include hashtags during the airing of their show to generate conversation on Twitter. Also, actors and other individuals associated with the show may send messages and communicate with viewers live during the show. Finally, Nielsen, the provider of television ratings data, announced in 2012 that would begin to collect Twitter conversation metrics about television shows.

The set of television show comes from the online database, The Futon Critic.[5] I narrow the total set of shows to include only those shows that aired after 2006, the year Twitter launched.[6] Given the list of television shows, I then matched those shows to Twitter accounts if one was available as of the time of data collection. Matching Twitter accounts to television shows was a labor intensive process, as there does not exist (to the best of my knowledge) any data sources that collect Twitter accounts for television shows. I utilized Amazon Turk's service to obtain a preliminary list of

[5]The website for The Futon Critic is `www.thefutoncritic.com`.

[6]This includes shows that premiered prior to the the start of Twitter, but were on the air for at least a portion of the time after and shows that premiered after the start date.

accounts, by soliciting three laborers on Amazon Turk to find the Twitter account associated with the show. If their choice was unanimous, I did not perform a secondary check. For each show that did not have a unanimous account identified by the Turk laborers, and for shows that returned no results, I did a search for the account to reconcile the difference or to confirm the lack of an account.

After collecting the account names, I used the Twitter Application Programming Interface (API) to acquire information for all the profiles. Data for each show included date of account creation, number of tweets and date of most recent tweet, number of followers (other accounts that add the focal show's tweets to their timeline), number of friends (accounts the focal show includes on its timeline), number of times the account was added to a list (a way for users to organize accounts into categories), number of favorited tweets (messages the focal show account wanted to acknowledge or keep a record of), location (if provided), and whether the account was verified. The data is censored as of the date of the API request in January 2014.[7]

Crucially, the set of television shows in this study represents a well-defined and complete population of at-risk or target members. By using the full population in each study, no systematic bias can arise from the identification of members of the population based on the adoption decision. Thus, for both the adoption (Table 3.2) and diffusion Models (Table 3.3 and 3.5), identifying the appropriate population allows for a precise, unbiased estimate of hazard rates and adoption shares, respectively.

I separately address adoption, diffusion, and behavior using three different empirical approaches. For adoption, I utilize a hazard model to measure time to adopting Twitter by television shows. I then employ a logistic function to assess the diffusion of Twitter within each network over time. Finally, I provide simple reduced form regressions to assess the relationships among various Twitter activities by and related to the television show accounts. For each method, I outline the approach and describe the results below.

[7]For a subset of shows, the Twitter API was accessed in March 2015 to acquire the profile data. This difference in time of data collection may affect the descriptive regressions presented in Table 3.9. For those regressions, I include an indicator variable, *acquired2014* that equals one if the profile data was collected in 2014 and zero if 2015. The variable is not statistically significant at a 5% level of confidence for any of the regressions.

3.5 Results

3.5.1 Adoption

I employ a Weibull proportional hazard model to assess the duration from the premiere of the show to the social media adoption decision, as measured by the date on which an account was created for the show.[8] Shows are considered to be 'at risk' on the later of Twitter's launch in July 2006 (for shows that premiered prior to Twitter's launch) or the premiere of the show's first episode. The hazard model accounts for various characteristics of the data, including the different times at which shows become at risk and the right censoring that occurs if a show has not adopted social media by the date of data collection (January 2014).

The 'at-risk' population is the 4,057 television shows that aired after the launch of Twitter in July 2006. Of those shows, I identify 1,209 shows that obtain a Twitter account as of the time of data collection (in January 2014).[9] A show was no longer at risk at the earlier of the show cancellation or the data collection date.

The model is:

$$h(t, X_{it}, \beta, p) = (p + 1)t^p \exp\left(X_{it} \ln \beta\right) \tag{3.1}$$

In Equation 3.1, X_i includes characteristics about the network, premiere year, and genre of the television show. First, *big 4 network* is a binary variable that equals one of the show aired on one of the big 4 networks, ABC, CBS, FOX, or NBC. Second, *premiere year* is included in the hazard model non-parametrically and captures the year the show first aired. Finally, *genre* of the show can be either reality, animated comedy, serial, news/talk, or other. Categories include shows that aired prior to 2001, between 2001 and 2005, or annually from 2006 to 2013. The term p represents the ancillary parameter of the Weibull distribution and can be interpreted as the elasticity of time in the hazard model.

An important assumption had to be made for a subset of shows where an account was established prior to the episode's first airing. In those instances, the shows are never at risk and have an undefined duration to event time. In order to include these "fast adopter" shows in the analy-

[8] Though a semiparametric hazard model, such as the Cox proportionate hazard model, would not make any assumption on the baseline hazard, social media adoption rates are likely a function of time. The Schoenfeld (1982) test of proportionality of hazard shows a statistically significant difference between time and the residuals ($p < 0.001$).

[9] I identify a show account as an account that was created specifically for that particular show. In certain instances, accounts that were not created specifically for the show would be the focal point on Twitter for that show's activity. For example, hosts or personalities on shows may have managed the Twitter conversation about the show from those individual accounts. I did not include these types of accounts in the analysis, as that account was associated with an individual (or other institution) and the choice to join was not the organization's.

sis, I made the simplifying assumption that the show became at risk one day prior to the account creation date .

Because each of the explanatory variables is categorical, I present summary data as tabulations in Table 3.1. The table presents adopters and non-adopters, by network, premiere year, and genre. Numbers of adopters is broken up by whether the account was created after the show premiere ("Adopted") or after ("Fast Adopter"). The tabulations of fast adopters provide an indication of the distinctions among members different types of shows. With respect to network, 74.2% of adopters were fast adopters on the big 4 networks, compared to 65.4% for shows on other networks. Based on the year in which a show premieres, fast adopters represent an increasing share of all adopters over time, increasing from 15.9% in 2008 to 94.9% in 2013. Descriptive data indicating that the propensity to adopt is faster for shows on the big 4 networks and over time is supported by the hazard model.

Results from estimation of the hazard model model is presented in Table 3.2. In all models, standard errors are clustered at the network level. Reference categories are "reality" for the *genre* variable and July 2006 to 2007 for *premiere year*, the period in 2006 after Twitter's launch (July 15, 2006) and the full year 2007. Columns 1 through 3 report the univariate results for network, genre and date, respectively. Column 4 shows results for all variables in the model, and is the focus of the discussion below.

First, looking at the rate of adoption by network, I note that the rate of adoption among shows on the Big 4 networks was seven times higher than shows on other networks (HR = $\exp(\beta) - 1 = \exp 2.078 - 1 = 6.988, p < 0.001$).[10] Looking at the impact of premiere year on the rate of Twitter adoption, there is a strong relationship between increased rates of social media adoption and the year the show premiered on television. Moreover, the year over year difference in the effect from 2007 on is statistically significant, indicating that newer shows are ever more rapidly adopting social media. Overall, shows that begin during periods when social media has become increasingly more pervasive are more likely to rapidly incorporate social media into the show's marketing and development plan.

[10] Interviews corroborated this finding, with one executive indicated that the big 4 networks "understand that you can market to the Twitter audience. The Twitter audience is more loyal and engaged. And has a higher intent to view and will watch longer. It is a valuable audience to market to."

Table 3.1: Tabulation of Twitter Adoption by Show Characteristics

Network	Did Not Adopt	Adopted	Fast Adopters	Total
Network				
Big 4	220	74	213	507
Other	2,628	319	603	3,550
Premiere Year				
Pre-Twitter	238	120	0	358
2006-2007	373	46	1	420
2008	308	37	7	352
2009	341	49	52	442
2010	367	43	106	516
2011	393	44	187	624
2012	420	41	220	681
2013	408	13	243	664
Genre				
reality	2,077	173	375	2,625
serial	486	149	339	974
animated comedy	121	18	31	170
news/talk	79	31	48	158
other	85	22	23	130
Total	2,848	393	816	4,057

Note: Tabulation of shows in sample by characteristic of show (network, premiere year, and genre) and adoption status. Adoption is divided into two types–shows that adopted after the premiere of the show and shows that adopted prior to the premiere ("Fast Adopters").

Table 3.2: Hazard Model of Twitter Account Adoption

	(1)	(2)	(3)	(4)
big 4 network	1.502***			2.073***
	(0.20)			(0.43)
Premiere Year				
Pre-Twitter		0.069		0.020
		(0.21)		(0.25)
2008		0.663***		1.150***
		(0.17)		(0.26)
2009		1.504***		2.105***
		(0.17)		(0.35)
2010		1.790***		2.815***
		(0.20)		(0.51)
2011		2.509***		3.588***
		(0.21)		(0.50)
2012		2.694***		3.976***
		(0.19)		(0.59)
2013		3.241***		4.592***
		(0.19)		(0.58)
Genre				
serial			1.253***	1.291***
			(0.19)	(0.19)
animated comedy			-0.124	0.094
			(0.28)	(0.24)
news/talk			0.726***	1.312***
			(0.15)	(0.29)
other			0.193	0.880***
			(0.26)	(0.25)
Constant	-3.356***	-1.763***	-3.388***	-3.631***
	(0.36)	(0.20)	(0.19)	(0.63)
ln(p)	0.643***	-0.562***	0.603***	-0.449***
	(0.09)	(0.12)	(0.05)	(0.12)
Shows	4057	4057	4057	4057
Twitter signups	1209	1209	1209	1209
Networks (clusters)	111	111	111	111

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Note: Models estimated using Weibull proportionate hazard model. The term *ln(p)* is time elasticity parameter. Columns 1 to 3 are univariate regressions. Colunmn 4 includes all variables. Reported standard errors are clustered at the network level. The hazard rate for any factor, HR, is derived by the following formula: $HR = \exp(\beta)$. The omitted category for *genre* is "reality" and for *premiere year* is Jul 2006-2007, the period in 2006 after Twitter's launch (July 15, 2006) and 2007.

3.5.2 Diffusion

To estimate the television show adoption rates of Twitter by networks and over time, I begin with the logistic function, a standard approach in the diffusion literature (Griliches, 1957; Geroski, 2000). For a given network or premiere year cohort, i, at time, t:

$$s_{it} = \frac{S_i}{1 + \exp\left[-(\beta_0 + \beta_1 X_t)\right]} \tag{3.2}$$

where s_{it} is the share of shows in cohort i that have adopted Twitter at time t, S_i is the maximum Twitter adoption rate for cohort i, and X_t indicates time (measured in calendar quarters). The parameters β_0 and β_1 reflect the shifting and the steepness of the curve, respectively.

To account for differences in adoption timing and rates across cohorts, I estimate the following linear model:

$$\ln\left(\frac{s_{it}}{S_i - s_{it}}\right) = \beta_0 + \beta_1 X_t + \theta_i + \gamma_i X_t + \varepsilon_{it} \tag{3.3}$$

In the above specification, θ_i is a cohort level fixed-effect that accounts for differences in the starting period of adoption for each cohort, and γ_i accounts for cohort-level differences in the average rate of adoption, β_1.

Results for Network Cohorts

Figures 3.1 and 3.2 show the diffusion of Twitter for the big four networks (ABC, CBS, FOX and NBC) and select cable networks, respectively. Together, the graphs show the heterogeneity in timing, rate, and total adoption across networks.

Estimates from the mixed linear model described in Equation 3.3 are reported in Table 3.3, Column 2 (Column 1 presents a fixed effects model and is included for reference). Errors are clustered at the network level and no structure was imposed on the covariance matrix of the random effects. Networks with fewer than two shows airing during the study period were excluded from the analysis, as were syndicated television shows.

Table 3.4 utilizes the parameters from Table 3.3 to arrive at network-specific estimates for the diffusion of social media. Three parameters characterize the diffusion curve. First, maximum adoption, S_i, is the highest share of shows for network i that adopted Twitter during the study

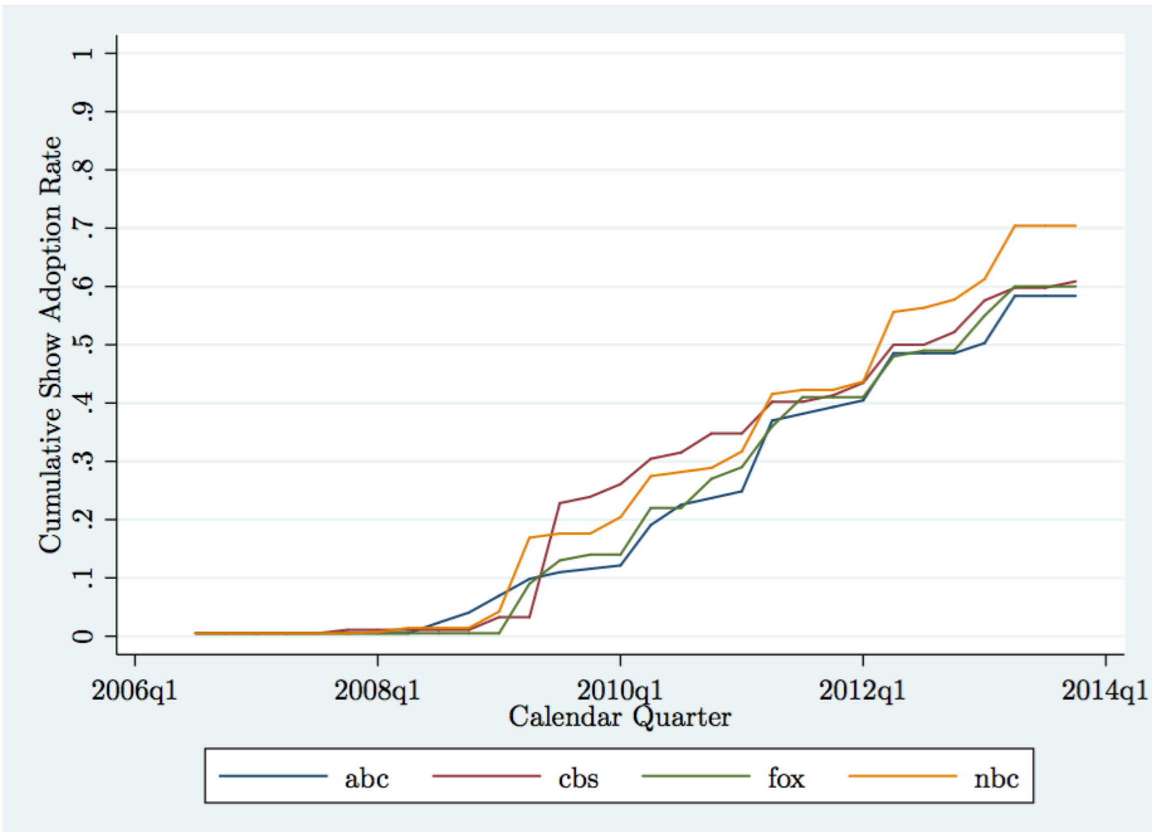


Figure 3.1: Twitter Diffusion within Big 4 Networks. Each line depicts s_{it} for the given network, which is calculated as the cumulative number of shows that adopted Twitter at time t divided by the total shows on the network during the sample period.

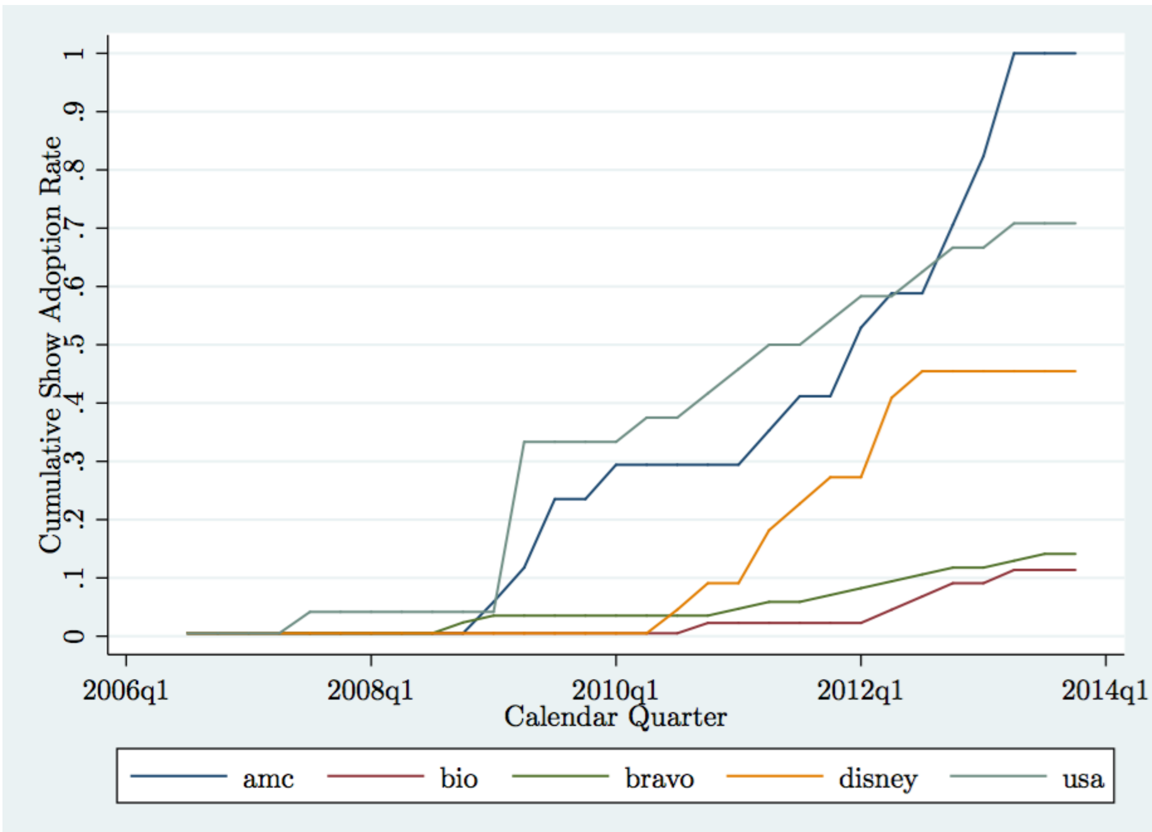


Figure 3.2: Twitter Diffusion within Select Cable Networks. Each line depicts s_{it} for the given network, which is calculated as the cumulative number of shows that adopted Twitter at time t divided by the total shows on the network during the sample period.

Table 3.3: Twitter Diffusion by Network

	(1)	(2)
Time	0.298***	0.312***
	(0.008)	(0.010)
Constant	-60.643***	-63.553***
	(1.580)	(2.099)
var(Time)		0.008***
		(0.004)
var(Constant)		343.304***
		(176.309)
cov(Time,Constant)		-1.665***
		(0.826)
var(Residual)		1.278*
		(0.123)
Model	FE	MLM
Network-Quarters	2410	2410
Networks (clusters)	93	93

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Note: Column 1 is a fixed effects (FE) model estimated using OLS and column 2 is a mixed linear model (MLM) estimated using maximum likelihood. The dependent variable is a function of the share of shows that adopted Twitter on network i at time t (measured in calendar quarters) and the maximum share for that network, $\ln(s_{it}/(S_i - s_{it}))$. Standard errors, reported in parentheses, are clustered at the network level. For the mixed linear model in Column 2, no structure is imposed on the covariance matrix for the random effects.

period. Higher levels of maximum adoption imply lower levels of centralization. Second, the midpoint is the point of inflection on the logistic curve where the curve shifts from monotonically increasing to monotonically decreasing. For network i, the midpoint is calculated as:

$$\texttt{Midpoint}_i = \frac{\beta_0 + \theta_i}{\beta_1 + \gamma_i} \tag{3.4}$$

and is interpreted as the number of quarters after the launch of Twitter in Q3 2006 the midpoint of diffusion was achieved.[11] Third, the steepness of the diffusion curve is a parameter that measures how fast the diffusion took place within network, i and is calculated as:

$$\texttt{Steepness}_i = \beta_1 + \gamma_i \tag{3.5}$$

Higher values of steepness are associated with a more rapid rate of diffusion of Twitter within the network.

Table 3.4: Best Linear Unbiased Predictors by Network

Network	First Adoption	Twitter Accounts	Total Shows	Maximum Adoption	Midpoint	Steepness
a&e	2009q1	16	72	0.222	17.282	0.279
abc	2008q3	101	173	0.584	17.577	0.330
abc family	2007q2	25	33	0.758	16.063	0.292
adult swim	2008q3	5	26	0.192	15.054	0.300
amazon	2013q2	2	3	0.667	21.106	0.284
amc	2009q1	17	17	1.000	18.943	0.349
animal planet	2009q3	14	100	0.140	19.871	0.239
bbc america	2008q3	23	96	0.240	16.127	0.314
bet	2009q1	20	43	0.465	17.148	0.353
bio	2010q4	5	44	0.114	22.233	0.211
bounce tv	2012q2	3	7	0.429	15.507	0.360
bravo	2008q4	12	85	0.141	18.734	0.226
cartoon	2009q1	8	47	0.170	16.654	0.332

Continued on next page

[11]To arrive at an easily interpreted value for the midpoint, the calculation in Equation 3.4 is adjusted by 186, which represents the system value for Q3 2006.

Table 3.4 – *Continued from previous page*

Network	First Adoption	Twitter Accounts	Total Shows	Maximum Adoption	Midpoint	Steepness
cbs	2007q4	56	92	0.609	17.398	0.326
centric	na	0	3	0.000	na	na
cinemax	2012q3	2	3	0.667	23.261	0.640
cmt	2009q1	3	55	0.055	12.115	0.227
cnbc	2013q3	1	7	0.143	36.647	0.135
comedy central	2008q4	12	53	0.226	16.510	0.296
cooking	2010q2	7	29	0.241	19.328	0.377
current tv	2011q1	2	3	0.667	1.419	0.209
cw	2007q1	34	63	0.540	18.473	0.271
da	2009q4	3	20	0.150	18.752	0.328
directv	2009q2	7	16	0.438	16.630	0.359
discovery	2008q2	29	155	0.187	21.276	0.221
disney	2010q3	10	22	0.455	18.775	0.370
disney xd	2009q1	10	26	0.385	20.704	0.317
diy	2011q4	1	30	0.033	20.037	0.288
e	2009q1	32	57	0.561	19.131	0.307
espn	2009q2	2	5	0.400	12.171	0.402
esquire	2013q3	8	8	1.000	15.470	0.368
food	2010q1	6	57	0.105	17.937	0.253
fox	2009q2	60	100	0.600	17.953	0.348
fox reality	na	0	15	0.000	na	na
fx	2010q2	9	20	0.450	20.540	0.323
fxx	2009q4	5	5	1.000	17.666	0.437
g4	2009q2	9	23	0.391	15.134	0.413
golf	2011q4	2	6	0.333	20.924	0.545
gsn	2009q4	5	11	0.455	19.379	0.316
h2	2010q2	1	11	0.091	15.088	0.316
hallmark	2013q1	1	8	0.125	28.476	0.164
hbo	2009q1	14	51	0.275	18.483	0.268
health	2010q2	1	24	0.042	14.588	0.218

Continued on next page

Table 3.4 – *Continued from previous page*

Network	First Adoption	Twitter Accounts	Total Shows	Maximum Adoption	Midpoint	Steepness
hgtv	2009q1	16	85	0.188	17.674	0.283
history	2009q2	18	82	0.220	21.556	0.242
hub	2010q4	11	29	0.379	16.513	0.360
id	2010q1	4	98	0.041	18.765	0.179
ifc	2009q1	15	29	0.517	17.927	0.320
ion	2010q1	3	5	0.600	18.386	0.472
lifetime	2009q1	26	71	0.366	18.974	0.287
logo	2009q2	11	26	0.423	15.123	0.387
military	2012q4	2	26	0.077	28.415	0.140
mtv	2007q2	44	122	0.361	17.292	0.284
mtv2	2009q2	5	18	0.278	18.831	0.268
mynetworktv	na	0	21	0.000	na	na
nat geo wild	2008q4	8	77	0.104	19.556	0.218
nbc	2008q1	100	142	0.704	17.687	0.337
nbcsn	2012q3	2	3	0.667	13.307	0.349
netflix	2011q3	6	9	0.667	25.494	0.275
ngc	2009q2	18	98	0.184	20.001	0.252
nickelodeon	2008q4	27	50	0.540	18.245	0.327
nickmom	2012q4	2	5	0.400	14.875	0.343
nicktoons	2009q2	3	13	0.231	14.900	0.354
ovation	2011q3	3	7	0.429	22.718	0.355
own	2009q1	16	60	0.267	18.620	0.304
oxygen	2009q1	26	44	0.591	18.273	0.350
pbs	2007q4	12	29	0.414	14.117	0.304
pivot	2013q3	4	5	0.800	15.699	0.362
planet green	2010q1	2	36	0.056	16.371	0.249
reelz	2011q2	6	10	0.600	22.396	0.506
science	2009q1	4	45	0.089	13.335	0.274
showtime	2008q3	28	54	0.519	16.890	0.319
soapnet	2009q3	1	9	0.111	12.766	0.303

Continued on next page

Table 3.4 – *Continued from previous page*

Network	First Adoption	Twitter Accounts	Total Shows	Maximum Adoption	Midpoint	Steepness
speed	2009q1	2	6	0.333	15.626	0.361
spike tv	2009q2	15	61	0.246	16.048	0.321
starz	2009q2	13	13	1.000	18.618	0.385
style	2009q2	9	38	0.237	17.510	0.312
sundance	2008q2	5	34	0.147	15.882	0.254
syfy	2007q3	30	73	0.411	17.632	0.283
tbs	2008q3	14	20	0.700	18.046	0.354
teen nick	2009q2	5	10	0.500	17.564	0.366
tlc	2009q1	24	164	0.146	19.219	0.243
tnt	2008q4	21	28	0.750	18.848	0.342
travel	2008q2	10	78	0.128	15.415	0.219
trutv	2009q2	7	50	0.140	18.651	0.244
tv guide	2012q3	1	9	0.111	22.749	0.248
tv land	2011q2	5	21	0.238	22.567	0.259
tv one	2011q4	4	22	0.182	24.204	0.262
usa	2007q3	17	24	0.708	15.687	0.321
velocity	2009q1	3	15	0.200	11.639	0.333
versus	2010q4	1	4	0.250	15.955	0.378
vh1	2009q1	27	127	0.213	19.225	0.268
vh1 classic	2009q1	1	4	0.250	-6.005	0.136
we	2009q2	12	66	0.182	17.907	0.278
weather	2011q4	4	28	0.143	18.778	0.360
wgn	2011q1	1	3	0.333	18.029	0.553

Note: Best linear unbiased predictors (BLUPs) provided by network and based on estimates of Equation 3.3 presented in Column 2 of Table 3.3. Network diffusion curves are characterized by Maximum Adoption, Midpoint, and Steepness. Maximum adoption, S_i is the highest share of shows for network i that adopted Twitter. Midpoint represents the number of quarters after the launch of Twitter (Q3 2006) that diffusion curve reaches the inflection point and begins monotonically decreasing. It is calculated as $\texttt{Midpoint}_i = ((\beta_0 + \theta_i)/(\beta_1 + \gamma_i)) - 186$ (the value 186 is the stored value of Twitter's launch quarter). Steepness is calculated as $\texttt{Steepness}_i = \beta_1 + \gamma_i$. Quarter of first adoption, total shows, and total adoptions are provided for reference. Three networks (centric, fox reality, and mynetworktv) did not have any adoptions and were excluded from the regression, but are included in this table for reference.

Noting the variance in the three parameters shows the diverse strategies adopted by different networks. The midpoint (and date of initial adoption), steepness, and maximum adoption provide a meaningful assessment of a network's timing, speed, and centrality of its social media strategy, respectively. For example, the big four networks were early adopters where social media diffused rapidly through the organization as part of a more decentralized strategy. Some cable networks implemented a similar strategy, such as Showtime and the USA Network, whereas others differed on one or more dimensions. Networks such as Bravo and the Discovery Network adopted later at a slower rate and with a greater level of centralization.[12] Interviews with at least one network confirms that the low maximum adoption rate is consistent with a centralized approach to social media management.

Results for Premiere Year Cohorts

Similar to results by network cohorts in Section 3.5.2, I show the raw data as a graph, present the results from the diffusion model and then present and discuss estimates of parameters for each premiere year.

Figure 3.3 presents diffusion of Twitter by a television show's premiere year. A number of interesting results emerge from inspection of the diffusion curves. First, shows that premiered prior to the launch of Twitter demonstrate a prototypical "S-curve", starting by design at a share of adoption of zero. This compares to shows that premiered after 2009 where the rate of adoption does not show any visual evidence of the initial accelerated increase in adoption over time (i.e. the first 'half' of the logistic curve, where both the first and second derivatives of the curve are greater than zero) Moreover, beginning in 2009 the initial rate of adoption increases, indicating a general increase in the baseline level of adoption each year. Finally, shows that premiered during the recent years (2011 to 2013) reach equilibrium adoption rates that are markedly similar to shows that premiered prior to the launch of Twitter.

Results from estimating Equation 3.3 by premiere year are presented in Table 3.5.[13] The results from the mixed linear model are then used to arrive at best linear unbiased predictors for each

[12]The decentralized and centralized approaches to social media management are not inconsistent with the theory put forth in Section 3.3.1. Firms that opt for a centralized strategy may do so partly due to resource constraints, and the converse would be true for firms adopting a decentralized strategy.

[13]Though, the logistic curve may not be the best model for estimating diffusion by premiere year cohorts, particularly in recent years (based on the raw data presented in Figure 3.3), I show the results here to be consistent with results in Table 3.3.

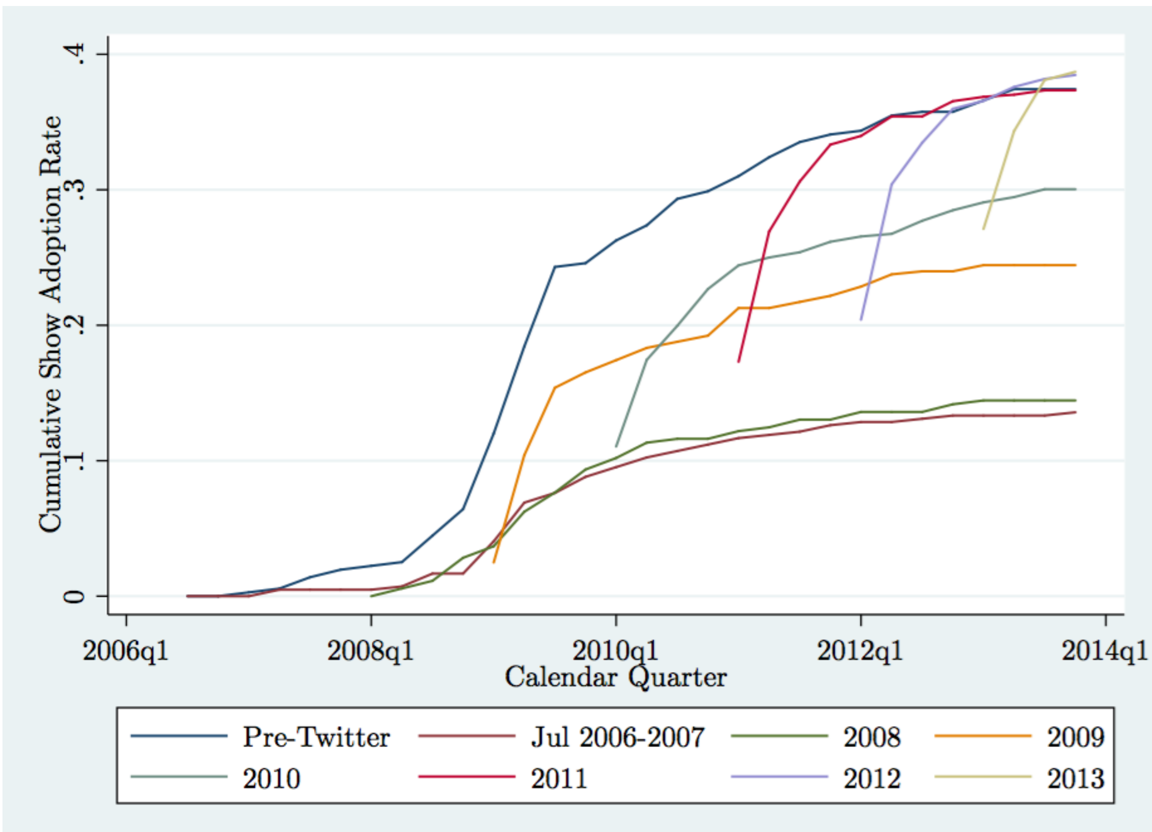


Figure 3.3: Twitter Diffusion by Premiere Year of Show. Each line depicts s_{it} for a given cohort of shows that premiered that year, which is calculated as the cumulative number of shows that adopted Twitter at time t divided by the total shows in each respective cohort during the entire sample period. "Pre-Twitter" includes shows that premiered before and continued airing, in part, after the launch of Twitter (July 15, 2006). "Jul 2006-2007" cohort includes 2006 shows that premiered after Twitter's July launch and all shows premiering in 2007.

Table 3.5: Twitter Diffusion by Premiere Year

	(1)	(2)
Time	0.290***	0.376***
	(0.012)	(0.066)
Constant	-58.060***	-76.663***
	(2.460)	(14.169)
var(Time)		0.028
		(0.080)
var(Constant)		1279.032*
		(3704.004)
cov(Time,Constant)		-5.935**
		(17.233)
var(Residual)		0.267***
		(0.057)
Model	FE	MLM
Start Year-Quarters	138	138
Start Years (clusters)	8	8

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Note: Column 1 is a fixed effects (FE) model estimated using OLS and column 2 is a mixed linear model (MLM) estimated using maximum likelihood. The dependent variable is a function of the share of shows that adopted Twitter on network *i* at time *t* (measured in calendar quarters) and the maximum share for that network, $\ln(s_{it}/(S_i - s_{it}))$. Standard errors, reported in parentheses, are clustered at the network level. For the mixed linear model in Column 2, no structure is imposed on the covariance matrix for the random effects.

premiere year (presented in Table 3.6) Premiere year cohorts include shows that premiered prior to the launch of Twitter on July 15, 2006, shows that premiered after Twitter's launch in 2006 and 2007, and annually thereafter.

Estimates for premiere year cohorts demonstrate similar trends observed in the raw data. Beginning in 2009, both midpoint and steepness increase each year, indicating the higher levels of initial adoption, the more rapid rates of diffusion. Maximum adoption levels are similar for shows premiering prior to Twitter's launch and from 2011 to 2013. The evidence from the raw data in Figure 3.3 and annual estimates in Table 3.6 reveal the changing nature of diffusion of Twitter as a communication technology over time within the television industry.

Table 3.6: Best Linear Unbiased Predictors by Year

Network	Twitter Accounts	Total Shows	Maximum Adoption	Midpoint	Steepness
Pre-Twitter	134	358	0.374	13.957	0.304
Jul 2006-2007	57	420	0.136	14.211	0.283
2008	51	353	0.144	13.641	0.272
2009	108	442	0.244	12.273	0.260
2010	155	516	0.300	14.582	0.287
2011	233	624	0.373	16.200	0.346
2012	262	681	0.385	20.684	0.509
2013	257	664	0.387	23.739	0.750

Note: Best linear unbiased predictors (BLUPs) provided by premiere year of show and based on estimates of Equation 3.3 presented in Column 2 of Table 3.5. Premiere year cohort diffusion curves are characterized by Maximum Adoption, Midpoint, and Steepness. Maximum adoption, S_i is the highest share of shows of premiere year cohort i that adopted Twitter. Midpoint represents the number of quarters after the launch of Twitter (Q3 2006) that diffusion curve reaches the inflection point and begins monotonically decreasing. It is calculated as $\texttt{Midpoint}_i = ((\beta_0 + \theta_i)/(\beta_1 + \gamma_i)) - 186$ (the value 186 is the stored value of Twitter's launch quarter). Steepness is calculated as $\texttt{Steepness}_i = \beta_1 + \gamma_i$. The Pre-Twitter cohort includes shows that premiered before and continued airing, in part, after the launch of Twitter (July 15, 2006). The Jul 2006-2007 cohort includes 2006 shows that premiered after Twitter's July launch and all shows premiering in 2007. Total shows and total adoptions are provided for reference.

3.5.3 Activity on Social Media

Part of the data acquired in the profile data collection process was the social media activity associated with the account up to the date of data collection. Social media data include actions performed by the account itself including number of tweets number of friends (other accounts the focal show account chooses to monitor), and number of favorites (noted messages) as well as actions by other users with respect to the focal show account including number of followers (users who choose to monitor the show's tweets in their timeline) and number of times listed (how many times an account placed the focal show account in an organizational list e.g. "TV shows" or "Shows that I watch"). Using this cross-sectional data, I created logged daily average measures to adjust for the different spans of time each account has been active.

Summary statistics and correlations for these average daily Twitter behavioral measures are

Table 3.7: Summary Statistics of Average Daily Twitter Behavior

	Mean	S.D.	Median	Min	Max
Logged measures					
tweets	-0.95	2.04	-0.63	-7.59	3.35
friends	-2.39	2.15	-2.36	-7.95	4.48
favorites	-4.58	2.71	-5.36	-7.95	2.89
followers	1.58	2.39	1.62	-5.59	9.23
listed	-3.07	2.05	-3.04	-7.67	3.69
Raw measures					
tweets	1.54	2.78	0.53	0.00	28.59
friends	0.97	4.62	0.09	0.00	87.97
favorites	0.23	0.87	0.00	0.00	17.92
followers	61.92	349.65	5.06	0.00	10,174.31
listed	0.32	1.47	0.05	0.00	39.91

Note: $n = 1,209$. An observation is a television show. Summary statistics above shown for all television shows that adopted Twitter during study period. Each raw variable is a daily average of each social media activity, and each logged variable is a log of that daily average.

presented in Tables 3.7 and 3.8, respectively. The average number of daily tweets by a television show account was approximately 1.5 and the television show added about 1 new friend each day. Approximately 62 Twitter users followed each television show account daily.

The data allows for descriptive regressions showing the associations between those various Twitter activities and characteristics of shows. Those results are presented in Table 3.9. All regressions are estimated using OLS with standard errors clustered at the network level. Columns 1 and 3 only include show characteristics, while columns 2 and 4 include other Twitter behaviors.

Table 3.8: Correlation Table for Average Daily Twitter Behavior Variables

	(1)	(2)	(3)	(4)	(5)
(1) tweets	1.00				
(2) friends	0.58	1.00			
(3) favorites	0.64	0.42	1.00		
(4) followers	0.71	0.40	0.51	1.00	
(5) listed	0.65	0.36	0.41	0.91	1.00

Note: $n = 1,209$. Correlation shown for logged average daily measures of each variable for all shows that adopted Twitter during the study period.

Table 3.9: Twitter Actions by Show and Community

DV:	(1) Tweets	(2) Tweets	(3) Followers	(4) Followers
Twitter Activity / Characteristics				
Tweets				0.111***
				(0.03)
Friends		0.241***		-0.033
		(0.03)		(0.02)
Favorites		0.228***		0.046***
		(0.02)		(0.01)
Followers		0.228***		
		(0.06)		
Listed		0.167*		0.977***
		(0.08)		(0.03)
account age	0.209***	0.132**	0.150*	-0.248***
	(0.04)	(0.04)	(0.06)	(0.03)
verified	2.127***	0.200*	3.152***	0.499***
	(0.14)	(0.10)	(0.12)	(0.08)
big 4 network	-0.929***	-0.372**	-0.416**	-0.288***
	(0.14)	(0.12)	(0.14)	(0.07)
Premiere Year				
Pre-Twitter	0.544*	0.006	0.556*	-0.024
	(0.27)	(0.20)	(0.26)	(0.12)
2008	-0.421	-0.504	-0.221	-0.245
	(0.38)	(0.29)	(0.36)	(0.15)
2009	-0.215	-0.224	-0.631*	-0.409**
	(0.27)	(0.19)	(0.31)	(0.14)
2010	0.051	0.139	-0.790**	-0.487***
	(0.28)	(0.20)	(0.25)	(0.13)
2011	0.525	0.396	-0.297	-0.248*

Continued on next page

Table 3.9 – *Continued from previous page*

	(1)	(2)	(3)	(4)
DV:	Tweets	Tweets	Followers	Followers
	(0.30)	(0.21)	(0.26)	(0.12)
2012	0.667*	0.600**	-0.626*	-0.374**
	(0.27)	(0.22)	(0.28)	(0.13)
2013	0.969**	0.591**	-0.595*	-0.536***
	(0.30)	(0.22)	(0.30)	(0.15)
Constant	-3.039***	0.072	0.066	6.112***
	(0.39)	(0.46)	(0.40)	(0.23)
Genre dummies	Yes	Yes	Yes	Yes
Shows	1209	1209	1209	1209
Networks (clusters)	97	97	97	97
F-test	41.0	126.1	108.3	354.9
Adj R-squared	0.33	0.69	0.51	0.89

* $p < 0.05$; ** $p < 0.01$; *** < 0.001.

Note: All models are estimated using OLS with standard errors (displayed in parentheses) clustered at the network level. Included in all regressions is a binary variable, *acquired 2014* that equals one if the activity data for the show was collected in 2014 and zero if the data was collected at a later time in 2015. The variable is not statistically significant at a 5% level of significance in any regression, indicating that averages did not significantly differ between shows whose data was collected at a later time.

Some interesting facts emerge from the regressions. First, while shows on the big 4 networks adopt Twitter at a faster rate than other shows, they are associated with fewer log daily tweets (Column 2, $\beta = -0.368, p < 0.003$) and fewer log daily followers (Column 4, $\beta = -0.289, p < 0.001$). This result is corroborated by an interview with an executive who indicated that the big 4 networks were likely to be more prepared to market to viewers on Twitter, but that niche shows better align with Twitter's demographic and are likely to see higher levels of engagement.

Also, compared to shows that premiered just after the launch of Twitter, shows premiering in 2012 (Column 2, $\beta = 0.600, p < 0.008$) and 2013 ($\beta = 0.591, p < 0.008$) utilize the service more. Surprisingly, shows premiering from 2009 to 2013 experience lower average daily new followers

(Column 4).

The activity regressions also highlight the importance of some of Twitter's actions to improve safety for organizations operating on the platform.In June 2009, the company introduced a feature for accounts called "Verified Accounts." Popular individuals and institutions frequently had fake accounts set up by non-affiliated individuals in those names, thus creating an environment where it was difficult for the actual individuals and institutions to interact. A verified account contains a logo besides the name, indicating the user behind the account is, in fact, real and not fraudulent, increasing the confidence for other users that the owner was not being impersonated and tweets are from that actual person described by the account. This action improved the safety of the platform, from a market design perspective.

In my dataset of television shows, of the 1,209 shows that adopted Twitter, 575 shows were verified as of the time of data collection. The importance of verified accounts is underscored by Column 4 of Table 3.9. This regression shows that there is a positive and statistically significant relationship between whether a show is verified and the average daily change of number of followers of the show ($\beta = 0.501, p < 0.001$), a behavior by users that is associated with improved organizational performance (see Chapter 2).

3.6 Discussion

In this paper, I provide evidence about how organizations adopt social media. First, larger and newer organizations adopt at a faster rate than other organizations. Second, I show that organizational heterogeneity in diffusion reveals differences in the management of social media with respect to the timing, speed and centrality of its management. The case of social media in the television industry show how as a new innovation or technology matures and is widely adopted, it evolved from being a dynamic capability to a 'minimally required capability.'

Organizational adoption of social media emerged largely out of a need to broadcast to users. However, given that organizations interact on social media through multiple levels (i.e. organizations, subunits, and individuals), there are many points through which they can develop the capabilities (Tripsas & Gavetti, 2000) to incorporate and process the information that is present on social media (Cohen & Levinthal, 1990). By holding many points of contact with social media, information, sentiment, and popular *zeitgeist* can be absorbed by multiple areas of the organization, thus facilitating its utility within the organization. As the utilization of information on social media increases, organizations shift from solely focusing on producing content for users to consuming and processing available information within the organization (O'Reilly & Tushman, 2008).

Due to the path dependent origin of social media management, as organizations increasingly absorb the information from social media, learn how to interact with social media, and develop the capabilities around its management, they adjust internal processes (Huber, 1991; Levinthal, 1991; Nelson, 1995).

Interviews with managers in the television industry at least provide some anecdotal evidence of this shift. Discussions with television managers revealed that social media influences various activities of the organization, including writers using it to inspire ideas, executives using Twitter feedback to bring a canceled show back on a different network, and producers and executives receiving social media reports to help assess the success of a show or parts of a show. In one instance, an organization reoriented its management of social media around the multiple functions it serves within the organization. The group was separated from the marketing department and has responsibilities coordinating with all other departments to ensure social media is properly utilized for those various functions, including research, marketing, production, and management.

This paper provides a framework for thinking about how social media is currently handled by the firm at the organizational, and brand or divisional levels. As social media business models develop, organizations increase their presence on social media, and organization-individual interactions become more pervasive, management of social media is likely to evolve from content production to learning and influencing organizational decisions.

A
Supplemental Tables and Figures

Table A.1: Description Of 70 Kickstarter Outlier Projects

Project	Category	Start Date	End Date	Goal	Pledged
The Vanderbilt Republic Foundation: "Masters"	Art	04aug2009	04oct2009	50,000	50,265
Designing Obama	Art	16sep2009	05nov2009	65,000	84,614
Greenlight the PATROL BASE JAKER Movie	Film and Video	16jan2010	22feb2010	45,000	45,535
STUFFER	Film and Video	05jan2010	31mar2010	32,500	57,160
The Chris Knox Benefit Concert	Music	30mar2010	01apr2010	37,500	40,555
Decentralize the web with Diaspora	Technology	24apr2010	02jun2010	10,000	200,642
Mystery Brewing Company: A Non-Traditional Approach to Artisanal Ales	Food	16may2010	24jul2010	40,000	44,259
Musopen: Record and release free music without copyrights.	Music	16aug2010	15sep2010	11,000	68,360
Lockpicks by Open Locksport	Design	15jul2010	24sep2010	6,000	87,408
GAMEFUL, a Secret HQ for Worldchanging Game Developers	Games	22aug2010	29sep2010	2,000	64,966
Cursed Pirate Girl: "Our Generation's Alice in Wonderland" Jeremy Bastian comic book	Comics	14sep2010	14oct2010	2,500	36,018
SAVE Blue Like Jazz! (the movie)	Film and Video	24sep2010	26oct2010	125,000	345,992
REVERENCE	Photography	30sep2010	31oct2010	50,000	50,016
Glif - iPhone 4 Tripod Mount & Stand	Design	04oct2010	03nov2010	10,000	137,417
Reopen the Parkway Theater!	Theater	04oct2010	03dec2010	50,000	56,832
Search & Restore documents and unites the new jazz scene!	Music	05oct2010	06dec2010	75,000	76,823
TikTok+LunaTik Multi-Touch Watch Kits	Design	17nov2010	17dec2010	15,000	942,578
The TRANSMETROPOLITAN art book	Comics	10dec2010	15feb2011	26,000	46,690

Continued on next page

Table A.1 – *Continued from previous page*

Project	Category	Start Date	End Date	Goal	Pledged
Vere Sandal Company, USA - 2011 Collection	Fashion	03jan2011	01mar2011	12,000	56,619
The Manual	Publishing	08feb2011	10mar2011	40,000	53,291
The Brotherhood of the Screaming Abyss!	Publishing	07apr2011	06jun2011	80,000	85,750
Naked Sea – Spencer Tunick Dead Sea Installation	Art	26apr2011	06jun2011	60,000	116,270
Build DC Public School Kids a FoodPrints Teaching Kitchen!	Food	18may2011	15jun2011	60,000	60,409
New Broadway Musical: ONE FOR MY BABY	Theater	20apr2011	19jun2011	50,000	67,606
Julia Nunes would be nothing without me	Music	11jun2011	11jul2011	15,000	77,888
HexBright, an Open Source Light	Technology	21may2011	19jul2011	31,000	259,294
Womanthology; Massive All Female Comic Anthology!	Comics	07jul2011	08aug2011	25,000	109,302
Glory To Rome [[Black Box Edition]] Rome Demands BEAUTY!	Games	01aug2011	22aug2011	21,000	73,103
Nataly Dawn's first solo album	Music	18jul2011	06sep2011	20,000	104,788
Alien Frontiers: Factions	Games	01sep2011	02oct2011	15,000	76,078
An Evening With Neil Gaiman & Amanda Palmer	Music	06sep2011	03oct2011	20,000	133,342
VENUS PATROL: charting a new course for videogame culture	Games	07sep2011	07oct2011	50,000	105,398
Brand New Windowfarms- Vertical Food Gardens	Food	17nov2011	08dec2011	50,000	257,308
D-Day Dice Board Game	Games	30oct2011	09dec2011	13,000	171,805
Printrbot: Your First 3D Printer	Technology	17nov2011	17dec2011	25,000	830,828
The Versalette by {r}evolution apparel	Fashion	17nov2011	23dec2011	20,000	64,246
New Five Iron Frenzy Album!!!!	Music	23nov2011	22jan2012	30,000	207,980
Elevation Dock: The Best Dock For iPhone	Design	13dec2011	11feb2012	75,000	1,464,707

Continued on next page

Table A.1 – *Continued from previous page*

Project	Category	Start Date	End Date	Goal	Pledged
The Order of the Stick Reprint Drive	Comics	22jan2012	21feb2012	57,750	1,254,120
Double Fine Adventure	Games	09feb2012	14mar2012	400,000	3,336,371
Idle Thumbs Video Game Podcast	Publishing	20feb2012	22mar2012	30,000	136,924
Pebble: E-Paper Watch for iPhone and Android	Design	11apr2012	19may2012	100,000	10,266,846
Flint and Tinder: Premium Men's Underwear	Fashion	22apr2012	22may2012	30,000	291,493
Amanda Palmer: The new RECORD, ART BOOK, and TOUR	Music	30apr2012	01jun2012	100,000	1,192,793
The Olympic City	Photography	30may2012	29jun2012	45,000	66,162
Ministry of Supply: The Future of Dress Shirts.	Fashion	08jun2012	11jul2012	30,000	429,277
THE ICARUS DECEPTION: WHY MAKE ART? New from Seth Godin	Publishing	18jun2012	17jul2012	40,000	287,342
Nomiku: bring sous vide into your kitchen.	Food	18jun2012	18jul2012	200,000	586,061
BRIDEGROOM - An American Love Story	Film and Video	19jun2012	19jul2012	300,000	384,376
Standard Time - The Workshop	Dance	27jun2012	27jul2012	12,000	31,028
Save the Lyric!	Theater	06jul2012	07aug2012	150,000	158,692
OUYA: A New Kind of Video Game Console	Games	10jul2012	09aug2012	950,000	8,596,475
Ukiyo-e Heroes	Art	01aug2012	31aug2012	10,400	313,341
Oculus Rift: Step Into the Game	Technology	01aug2012	01sep2012	250,000	2,437,430
The Gamers: Hands of Fate	Film and Video	18jul2012	08sep2012	320,000	405,917
Charlie Kaufman's Anomalisa	Film and Video	11jul2012	09sep2012	200,000	406,237
Rescue The Historic Catlow Theater From Extinction	Theater	26jul2012	24sep2012	100,000	175,395
STILL MOTION presents "Moments Defined"	Dance	17aug2012	16oct2012	25,000	38,570
FORM 1: An affordable, professional 3D printer	Technology	26sep2012	26oct2012	100,000	2,945,885

Continued on next page

Table A.1 – *Continued from previous page*

Project	Category	Start Date	End Date	Goal	Pledged
"The Goon" Movie... let's KICKSTART this sucker!!!	Film and Video	12oct2012	11nov2012	400,000	441,900
YAGP's "Ballet's Greatest Hits" Gala	Dance	21nov2012	17dec2012	35,000	38,752
To Be Or Not To Be: That Is The Adventure	Publishing	21nov2012	21dec2012	20,000	580,906
GUSTIN: Redefining premium menswear, starting with denim.	Fashion	07jan2013	09feb2013	20,000	449,654
Video Game High School: Season Two	Film and Video	11jan2013	12feb2013	636,010	808,341
The Veronica Mars Movie Project	Film and Video	13mar2013	13apr2013	2,000,000	5,702,153
THE 10-YEAR HOODIE: Built for Life, Backed for a Decade!	Fashion	07mar2013	21apr2013	50,000	1,053,831
Planet Money T-shirt	Publishing	30apr2013	14may2013	50,000	590,807
ARKYD: A Space Telescope for Everyone	Photography	29may2013	01jul2013	1,000,000	1,505,367
Marina Abramovic Institute: The Founders	Art	26jul2013	25aug2013	600,000	661,452
Sansaire Sous Vide Circulator	Food	07aug2013	06sep2013	100,000	823,003

Table A.2: Robustness Test: Alternative Definition Of Outliers

	(1)	(2)	(3)	(4)
DV:	entrants	pledged	entrants	pledged
Outlier platform x Outlier category x Post			0.001	0.035
			(0.02)	(0.05)
Outlier platform x Post	-0.101**	-0.123**	-0.101**	-0.126**
	(0.01)	(0.02)	(0.01)	(0.02)
Outlier category x Post			-0.012	-0.006
			(0.02)	(0.05)
Post	0.040**	0.094**	0.041**	0.095**
	(0.01)	(0.02)	(0.01)	(0.02)
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Time-Varying Controls	Yes	Yes	Yes	Yes
Outlier period-platform-category FE	Yes	Yes	Yes	Yes
Observations	238067	237528	238067	237528
Outlier period-platform-categories	6292	6208	6292	6208

Note: (+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$) Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variable is count of project entrants in Columns 1 and 3 and dollars pledged in Columns 2 and 4. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for 244 Kickstarter outliers are stacked to produce the sample. Outliers are defined as the top 5 pledged projects within each category each year. An observation is an outlier period-platform-category-week.

Table A.3: Robustness Test: No Overlapping Outlier Periods

	(1)	(2)	(3)	(4)
DV:	entrants	pledged	entrants	pledged
Outlier platform x Outlier category x Post			0.068	0.185*
			(0.04)	(0.09)
Outlier platform x Post	-0.198**	-0.208**	-0.203**	-0.221**
	(0.03)	(0.04)	(0.03)	(0.04)
Outlier category x Post			-0.019	-0.063
			(0.04)	(0.10)
Post	0.109**	0.160**	0.110**	0.165**
	(0.02)	(0.04)	(0.03)	(0.04)
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Time-Varying Controls	Yes	Yes	Yes	Yes
Outlier period-platform-category FE	Yes	Yes	Yes	Yes
Observations	61835	61696	61835	61696
Outlier period-platform-categories	1625	1604	1625	1604

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variable is count of project entrants in Columns 1 and 3 and dollars pledged in Columns 2 and 4. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for 64 Kickstarter outliers are stacked to produce the sample. The sample excludes any outlier periods whose campaigns overlapped with another outlier in the same category. An observation is an outlier period-platform-category-week.

Table A.4: Robustness Test: Excluding Outliers from All Samples

	(1)	(2)	(3)	(4)
DV:	entrants	pledged	entrants	pledged
Outlier platform x Outlier category x Post			0.055	0.148*
			(0.04)	(0.07)
Outlier platform x Post	-0.206**	-0.219**	-0.211**	-0.232**
	(0.03)	(0.04)	(0.03)	(0.04)
Outlier category x Post			-0.031	-0.129
			(0.03)	(0.08)
Post	0.112**	0.175**	0.115**	0.187**
	(0.02)	(0.03)	(0.02)	(0.04)
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Time-Varying Controls	Yes	Yes	Yes	Yes
Outlier period-platform-category FE	Yes	Yes	Yes	Yes
Observations	68117	67877	68117	67877
Outlier period-platform-categories	1801	1769	1801	1769

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variable is count of project entrants in Columns 1 and 3 and dollars pledged in Columns 2 and 4. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for 70 Kickstarter outliers are stacked to produce the sample. In all outlier periods, entry and pledged data from the 70 outliers are excluded. An observation is an outlier period-platform-category-week.

Table A.5: Robustness Test: OLS Results

	(1)	(2)	(3)	(4)
DV:	ln entrants	ln pledged	ln entrants	ln pledged
Outlier platform x Outlier category x Post			0.009	0.012
			(0.02)	(0.11)
Outlier platform x Post	-0.183**	-0.223**	-0.184**	-0.223**
	(0.03)	(0.06)	(0.03)	(0.06)
Outlier category x Post			0.047	0.027
			(0.03)	(0.11)
Post	0.088**	0.101**	0.084**	0.099**
	(0.02)	(0.04)	(0.02)	(0.04)
Constant	0.481**	-1.551**	0.483**	-1.550**
	(0.08)	(0.25)	(0.08)	(0.25)
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Time-Varying Controls	Yes	Yes	Yes	Yes
Outlier period-platform-category FE	Yes	Yes	Yes	Yes
Observations	68122	68122	68122	68122
R-squared	0.317	0.434	0.318	0.434

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using OLS regression with standard errors clustered at the outlier period (in parentheses). The dependent variable is the log of entrants (plus 1) in Columns 1 and 3 and log of pledged (plus one) in Columns 2 and 4. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for 70 Kickstarter outliers are stacked to produce the sample. An observation is an outlier period-platform-category-week.

Table A.6: Robustness Test: Including Transferred Pledges Only

DV: *transferred pledged*	(1)	(2)
Outlier platform x Outlier category x Post		0.198*
		(0.08)
Outlier platform x Post	-0.241**	-0.258**
	(0.04)	(0.05)
Outlier category x Post		-0.123
		(0.08)
Post	0.190**	0.202**
	(0.04)	(0.04)
Year FE	Yes	Yes
Month FE	Yes	Yes
Time-Varying Controls	Yes	Yes
Outlier period-platform-category FE	Yes	Yes
Observations	67877	67877
Outlier period-platform-categories	1769	1769

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variable is committed dollars pledged, which are pledges that were transferred to the project creator at the conclusion of the campaign (i.e. pledges for projects that met or exceeded the goal in fixed campaigns and any positive pledges in flexible campaigns). Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for 70 Kickstarter outliers are stacked to produce the sample. An observation is an outlier period-platform-category-week.

Table A.7: Robustness Test: Backers as a Measure of Liquidity

DV: *backers*	(1)	(2)
Outlier platform x Outlier category x Post		0.185*
		(0.09)
Outlier platform x Post	-0.248**	-0.263**
	(0.05)	(0.05)
Outlier category x Post		-0.101
		(0.09)
Post	0.204**	0.213**
	(0.05)	(0.05)
Year FE	Yes	Yes
Month FE	Yes	Yes
Time-Varying Controls	Yes	Yes
Outlier period-platform-category FE	Yes	Yes
Observations	67877	67877
Outlier period-platform-categories	1769	1769

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variable is number of backers. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for 70 Kickstarter outliers are stacked to produce the sample. An observation is an outlier period-platform-category-week.

Table A.8: Robustness Test: *entry* and *pledged* Greater Than Zero

	(1)	(2)	(3)	(4)
DV:	entrants	pledged	entrants	pledged
Outlier platform x Outlier category x Post			0.054	0.195**
			(0.04)	(0.07)
Outlier platform x Post	-0.206**	-0.235**	-0.211**	-0.252**
	(0.03)	(0.04)	(0.03)	(0.04)
Outlier category x Post			-0.031	-0.122
			(0.03)	(0.08)
Post	0.112**	0.184**	0.115**	0.196**
	(0.02)	(0.04)	(0.02)	(0.04)
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Time-Varying Controls	Yes	Yes	Yes	Yes
Outlier period-platform-category FE	Yes	Yes	Yes	Yes
Observations	68117	63698	68117	63698
Outlier period-platform-categories	1801	1756	1801	1756

(+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$)

Note: Models estimated using Poisson regression with quasi-maximum likelihood standard errors, clustered at the outlier period (in parentheses). The dependent variables are number of project entrants in Columns 1 and 3 and dollars pledged in Columns 2 and 4. The sample is limited to weeks when the dependent variable is greater than zero. Each outlier period constitutes a 20 week window prior to launch and a 20 week window after the completion of each outlier. Outlier periods for 70 Kickstarter outliers are stacked to produce the sample. An observation is an outlier period-platform-category-week.

Table A.9: Description Of Shows

				Premiere			*initial*	
Show	Twitter account	Network	Genre	Date	Rating	Viewers	*followers*	Episodes
Agents of S.H.I.E.L.D.	agentsofshield	ABC	Action	9/24/13	4.7	17.01	112,860	10
Almost Human	almosthumanfox	FOX	Crime	11/18/13	3.1	9.38	5,284	5
Back in the Game	backinthegametv	ABC	Sitcom	9/25/13	2.2	8.01	1,349	10
Betrayal	betrayalabc	ABC	Drama	9/29/13	1.5	5.16	2,717	10
Blacklist	nbcblacklist	NBC	Crime	9/23/13	3.8	12.58	9,591	10
Brooklyn 99	brooklyn99fox	FOX	Sitcom	9/17/13	2.5	6.17	2,034	11
Dads	dadsonfox	FOX	Sitcom	9/17/13	2.2	5.76	1,988	11
Dracula	nbcdracula	NBC	Horror	10/25/13	1.8	5.26	24,319	6
Hostages	hostagescbs	CBS	Drama	9/23/13	1.8	7.41	1,682	13
Ironside	nbcironside	NBC	Drama	10/02/13	1.3	6.81	13,741	4
Lucky 7	lucky7abc	ABC	Drama	9/24/13	1.3	4.43	418	2
Master Chef Jr.	masterchefjrfox	FOX	Game	9/27/13	1.6	4.29	1,892	7
Masters of Sex	sho_masters	SHO	Period	9/29/13	0.4	1.00	942	12
Mom	momcbs	CBS	Sitcom	9/23/13	2.5	7.99	1,155	12
Once Upon a Time in Wonderland	wonderlandouat	ABC	Fantasy	10/10/13	1.7	5.82	25,461	8
Reign	cwreign	CW	Drama	10/17/13	0.8	1.98	5,622	8
Sean Saves the World	seansavesworld	NBC	Sitcom	10/03/13	1.4	4.43	8,367	9
Sleepy Hollow	sleepyhollowfox	FOX	Horror	9/16/13	3.5	10.10	4,329	10
Super Fun Night	superfunnight	ABC	Sitcom	10/02/13	3.2	8.23	2,354	9
The Crazy Ones	crazyonescbs	CBS	Sitcom	9/26/13	3.9	15.52	3,326	11

Continued on next page

Table A.9 – *Continued from previous page*

				Premiere			*initial*	
Show	Twitter account	Network	Genre	Date	Rating	Viewers	*followers*	Episodes
The Goldbergs	thegoldbergsabc	ABC	Sitcom	9/24/13	3.1	8.94	1,427	11
The Michael J. Fox Show	michaeljfoxshow	NBC	Sitcom	9/26/13	2.2	7.50	18,921	10
The Millers	themillerscbs	CBS	Sitcom	10/03/13	3.3	13.09	1,003	10
The Originals	cworiginals	CW	Horror	10/03/13	1.0	2.21	108,912	9
The Tomorrow People	cwtp	CW	Scifi	10/09/13	0.9	2.32	7,991	9
Trophy Wife	trophywifeabc	ABC	Sitcom	9/24/13	2.3	6.69	3,144	10
We Are Men	wearemencbs	CBS	Sitcom	9/30/13	2.0	6.61	921	2
Welcome to the Family	nbcwelcome	NBC	Sitcom	10/03/13	1.1	2.99	7,718	3

Table A.10: Results from the Fixed Effects OLS Model

DV: *rating*	(1)	(2)	(3)	(4)	(5)
Genre Subsample:				Niche	Not Niche
Community actions					
H1. followers	0.041	0.043*	0.036	0.351*	0.034
	(0.02)	(0.02)	(0.02)	(0.13)	(0.02)
H2. followers x matched network		0.140***			
		(0.04)			
H3. followers x niche genre			0.078		
			(0.05)		
H4. followers x initial followers					
second quartile				-0.454*	-0.010
				(0.13)	(0.03)
third quartile				-0.167	-0.049
				(0.26)	(0.05)
top quartile				-0.340	0.031
				(0.24)	(0.03)
community replies	-0.011	-0.018	-0.014	-0.104	-0.002
	(0.01)	(0.01)	(0.01)	(0.06)	(0.02)
community tweets	0.000	0.004	0.002	0.059	-0.003
	(0.01)	(0.01)	(0.01)	(0.05)	(0.01)
Show actions					
friends	-0.011	-0.012	-0.012	-0.007	-0.020**
	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)
show replies	0.006	0.007	0.001	-0.037	-0.006
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)
show tweets	0.018	0.019	0.022	-0.074	0.018
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
google trend(t-1)	0.159***	0.163***	0.166***	0.118	0.151**
	(0.03)	(0.03)	(0.03)	(0.28)	(0.04)
Constant	-0.240	-0.379*	-0.494*	-0.867	-0.016

Continued on next page

Table A.10 – *Continued from previous page*

DV: *rating*	(1)	(2)	(3)	(4)	(5)
Genre Subsample:				Niche	Not Niche
	(0.15)	(0.14)	(0.18)	(0.46)	(0.11)
Episode Number FE	Yes	Yes	Yes	Yes	Yes
Calendar Week FE	Yes	Yes	Yes	Yes	Yes
Show FE	Yes	Yes	Yes	Yes	Yes
Show-weeks	214	214	214	54	160
Shows	28	28	28	7	21
F-stat	64.7	68.8	61.5	15.2	158.4
R-squared	0.59	0.60	0.60	0.85	0.67
Autocorrelation F-stat	3.47	3.18	3.07	12.0	5.56

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Note: Models estimated using OLS with standard errors clustered at the show level (in parentheses). F-test represents joint test of significance for all covariates, excluding any time fixed effects. The dependent variable is the log of Nielsen's rating. Each regression includes a dummy variable, *skipped week*, that equals one for weeks when the show's air date was not consecutive. Autocorrelation F-test is a test for first-order autocorrelation in panel data devised by Woolridge (2002) and implemented by Drukker (2003).

Table A.11: Univariate Autoregressive Models Predicting Show Ratings

DV: *rating*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
followers		0.070*									
		(0.02)									
community replies			-0.014++								
			(0.01)								
community tweets				-0.004							
				(0.01)							
community hashtags					0.001						
					(0.01)						
community account mentions						-0.004					
						(0.00)					
friends							-0.004				
							(0.01)				
show replies								-0.005			
								(0.01)			
show tweets									0.028++		
									(0.01)		
show hashtags										0.014+	
										(0.01)	
show account mentions											0.011
											(0.02)
rating(t-1)	0.248*	0.165*	0.291*	0.285*	0.282*	0.280*	0.289*	0.263*	0.221*	0.270*	0.219*
	(0.08)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.09)	(0.10)	(0.09)	(0.09)

Continued on next page

Table A.11 – *Continued from previous page*

DV: *rating*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Constant	0.355*	-0.146	0.355*	0.334*	0.311*	0.330*	0.280*	0.340*	0.245++	0.259*	0.317*
	(0.12)	(0.19)	(0.10)	(0.10)	(0.10)	(0.09)	(0.11)	(0.11)	(0.14)	(0.12)	(0.13)
Calendar Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Show-weeks	186	186	186	186	186	186	186	186	186	186	186
Shows	26	26	26	26	26	26	26	26	26	26	26
Instruments	61	103	103	103	103	103	103	103	103	103	103
Wald Chi-Square	137.0	216.3	254.7	299.3	206.6	207.4	213.3	168.9	147.5	198.0	218.0
Arellano Bond $\mathcal{Z}_1$	-3.290	-3.634	-3.521	-3.534	-3.511	-3.532	-3.659	-3.582	-3.339	-3.437	-3.298
Arellano Bond $\mathcal{Z}_2$	0.827	0.0781	0.907	0.932	0.997	0.924	1.054	1.036	0.890	1.001	0.922
MAE	0.321	0.364	0.301	0.305	0.308	0.307	0.306	0.316	0.334	0.313	0.333
MSE	0.158	0.204	0.141	0.143	0.145	0.145	0.143	0.153	0.170	0.149	0.169
RMSE	0.397	0.451	0.375	0.379	0.381	0.381	0.378	0.391	0.412	0.387	0.412

+ $p < 0.2$; ++ $p < 0.1$; * $p < 0.05$.

Note: Models estimated using Arellano & Bond (1991) generalized method of moments (GMM) with standard errors clustered at the show level (in parentheses). The lagged dependent variable is instrumented by all prior levels until period $t-2$. Each social media measure is differenced and is instrumented by all prior levels of the covariate until period $t-1$. Presented are the Arellano & Bond (1991) test for autocorrelation, where the null hypothesis is no autocorrelation (model specification is supported when the first order test is statistically significant, while the second order is not). The dependent variable is the log of Nielsen's rating. Each regression includes a dummy variable, *skipped week*, that equals one for weeks when the show's air date was not consecutive.

Table A.12: Robustness Test: *viewers* as DV

DV: *viewers*	(1)	(2)	(3)	(4)	(5)
Genre Subsample:				Niche	Not Niche
Community actions					
H1. followers	0.030	0.027	0.016	0.240***	0.004
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
H2. followers x matched network		0.131*			
		(0.06)			
H3. followers x niche genre			0.093**		
			(0.03)		
H4. followers x initial followers					
second quartile				-0.296**	-0.006
				(0.09)	(0.02)
third quartile				0.015	0.000
				(0.10)	(0.03)
top quartile				-0.111	0.013
				(0.06)	(0.04)
community replies	-0.017	-0.022	-0.019	-0.067	-0.018
	(0.01)	(0.01)	(0.01)	(0.04)	(0.02)
community tweets	0.002	0.005	0.003	0.057*	0.000
	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)
Show actions					
friends	-0.035	-0.036	-0.037	-0.005	-0.049
	(0.02)	(0.02)	(0.02)	(0.01)	(0.03)
show replies	0.012	0.013	0.004	-0.022	0.005
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
show tweets	0.026	0.026	0.032	-0.070***	0.043
	(0.03)	(0.03)	(0.03)	(0.02)	(0.04)
google trend(t-1)	0.192***	0.195***	0.201***	0.062	0.176**
	(0.05)	(0.06)	(0.06)	(0.10)	(0.06)
viewers(t-1)	-0.113**	-0.109**	-0.118**	0.052	-0.124*

Continued on next page

Table A.12 – *Continued from previous page*

DV: *viewers*	(1)	(2)	(3)	(4)	(5)
Genre Subsample:				Niche	Not Niche
	(0.04)	(0.04)	(0.04)	(0.10)	(0.05)
Constant	1.052***	0.942***	0.840**	-0.529	1.381***
	(0.28)	(0.29)	(0.26)	(0.43)	(0.21)
Calendar Week FE	Yes	Yes	Yes	Yes	Yes
Show-weeks	186	186	186	47	139
Shows	26	26	26	7	19
Instruments	181	185	185	48	140
Wald Chi-Square	1012.7	15832.3	5899.5	288.5	474.3
Arellano Bond $\mathcal{Z}_1$	-1.295	-1.318	-1.306	-2.209	-1.272
Arellano Bond $\mathcal{Z}_2$	-1.254	-1.179	-1.178	-2.111	-1.189

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Note: Models estimated using OLS with standard errors clustered at the show level (in parentheses). The dependent variable is the log of number of viewers, in millions. F-test represents joint test of significance for all covariates, excluding any time fixed effects. Each regression includes a dummy variable, *skipped week*, that equals one for weeks when the show's air date was not consecutive.

Table A.13: Robustness Test: Additional Social Media Independent Variables

DV: *rating*	(1)	(2)	(3)	(4)	(5)
Genre Subsample:				Niche	Not Niche
Community actions					
H1. followers	0.050*	0.050*	0.040	0.449***	0.039*
	(0.02)	(0.02)	(0.02)	(0.07)	(0.02)
H2. followers x matched network		0.122***			
		(0.03)			
H3. followers x niche genre			0.079**		
			(0.03)		
H4. followers x initial followers					

Continued on next page

Table A.13 – *Continued from previous page*

DV: *rating*	(1)	(2)	(3)	(4)	(5)
Genre Subsample:				Niche	Not Niche
second quartile				-0.491***	-0.004
				(0.10)	(0.02)
third quartile				-0.254	-0.036
				(0.14)	(0.03)
top quartile				-0.425**	0.038
				(0.15)	(0.02)
community replies	-0.010	-0.017	-0.015	-0.141***	-0.003
	(0.01)	(0.01)	(0.01)	(0.04)	(0.02)
community tweets	-0.007	-0.003	-0.003	0.085**	-0.064
	(0.01)	(0.01)	(0.01)	(0.03)	(0.04)
community hashtags	0.006	0.005	0.004	-0.084**	0.038*
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)
community account mentions	0.000	0.001	0.001	0.094**	0.020
	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)
Show actions					
friends	-0.006	-0.007	-0.008	-0.009	-0.019***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
show replies	-0.000	0.001	-0.003	0.024	-0.002
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)
show tweets	0.018	0.020	0.041	0.095	0.056
	(0.03)	(0.03)	(0.03)	(0.10)	(0.04)
show hashtags	0.031	0.029	0.024	-0.111	0.009
	(0.03)	(0.03)	(0.02)	(0.10)	(0.03)
show account mentions	-0.026	-0.026	-0.036*	-0.074**	-0.036*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
google trend(t-1)	0.104**	0.106**	0.120**	0.103	0.118***
	(0.04)	(0.04)	(0.04)	(0.21)	(0.03)
rating(t-1)	0.148**	0.145**	0.117*	0.340**	0.066
	(0.05)	(0.05)	(0.05)	(0.13)	(0.04)

Continued on next page

Table A.13 – *Continued from previous page*

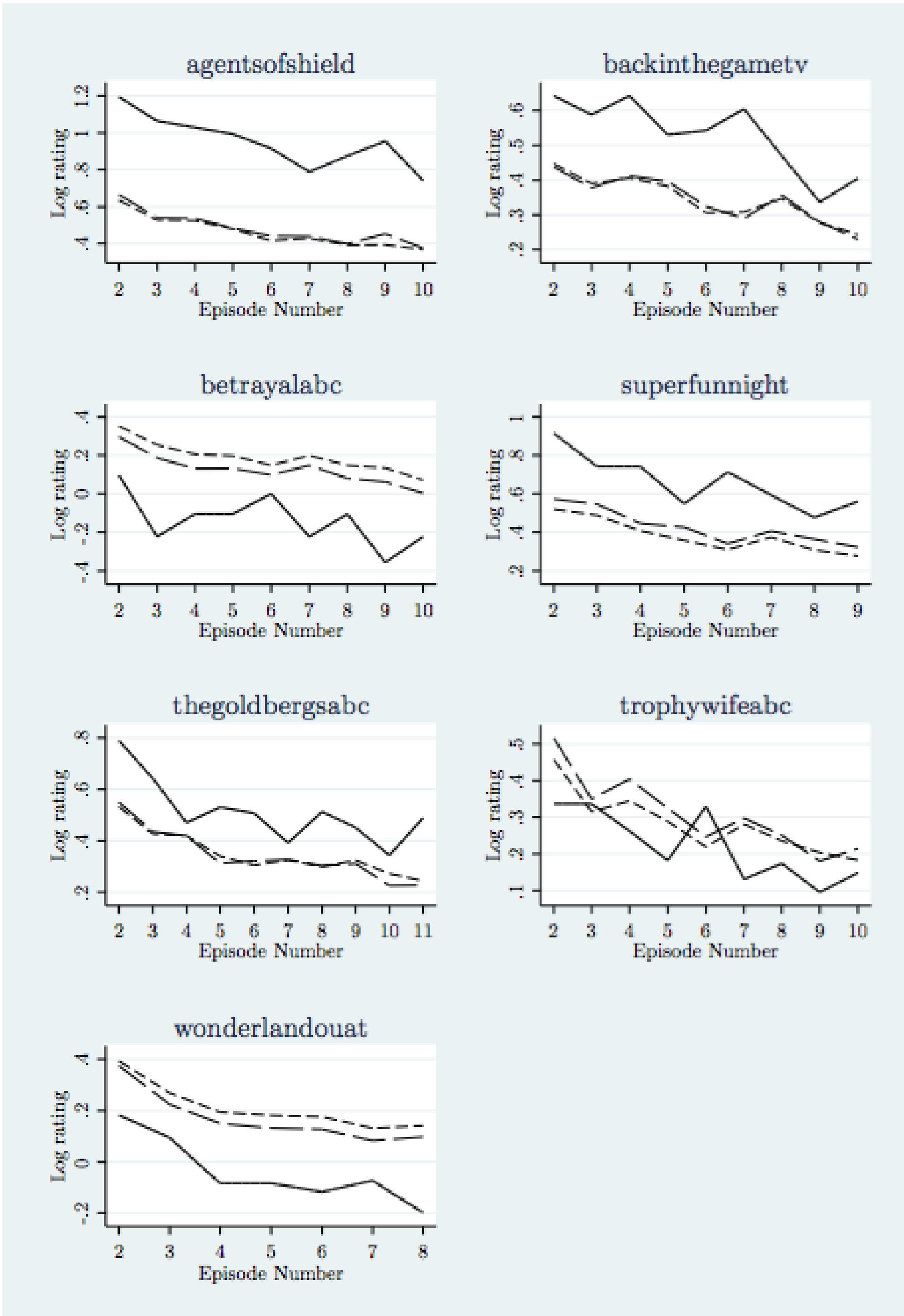
DV: *rating*	(1)	(2)	(3)	(4)	(5)
Genre Subsample:				Niche	Not Niche
Constant	-0.347	-0.452*	-0.519*	-1.412**	-0.141
	(0.19)	(0.22)	(0.23)	(0.53)	(0.12)
Calendar Week FE	Yes	Yes	Yes	Yes	Yes
Show-weeks	186	186	186	47	139
Shows	26	26	26	7	19
Instruments	187	187	187	48	140
Wald Chi-Square	29851.6	12193.1	4752085.2	180.7	1912.2
Arellano Bond $\mathcal{Z}_1$	-3.365	-3.381	-3.359	-2.210	-2.940
Arellano Bond $\mathcal{Z}_2$	-0.247	-0.111	-0.342	1.384	-1.433

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Note: Models estimated using OLS with standard errors clustered at the show level (in parentheses). The dependent variable is the log of Nielsen's rating. F-test represents joint test of significance for all covariates, excluding any time fixed effects. Each regression includes a dummy variable, *skipped week*, that equals one for weeks when the show's air date was not consecutive.

Figure A.1 ***(following page)*****:** Predicted Values from Social Media and Baseline Prediction Models (ABC Shows). Each panel shows how both the baseline prediction model and the social media prediction models compare to actual ratings for one show. The solid line represents actual *ratings*. The short dashed line is the baseline model. The long dashed line is the social media model.

Figure A.1: (continued)



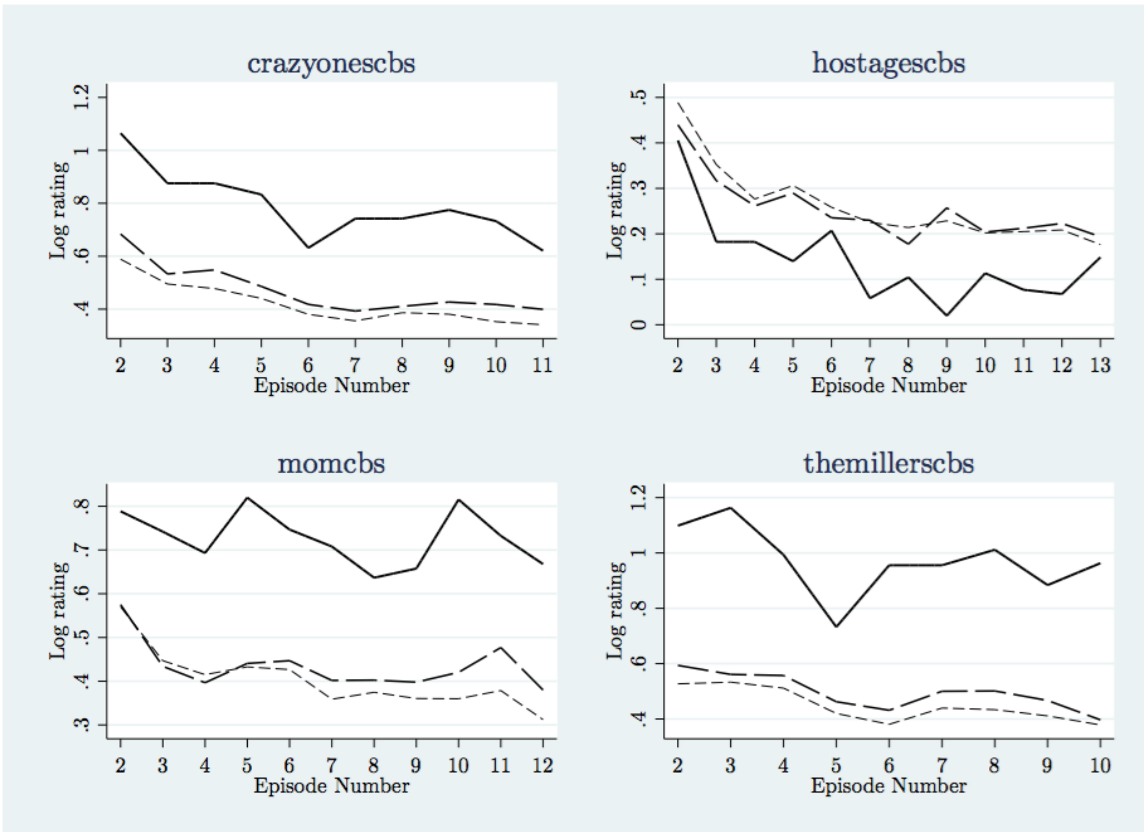


Figure A.2: Predicted Values from Social Media and Baseline Prediction Models (CBS Shows). Each panel shows how both the baseline prediction model and the social media prediction models compare to actual ratings for one show. The solid line represents actual *ratings*. The short dashed line is the baseline model. The long dashed line is the social media model.

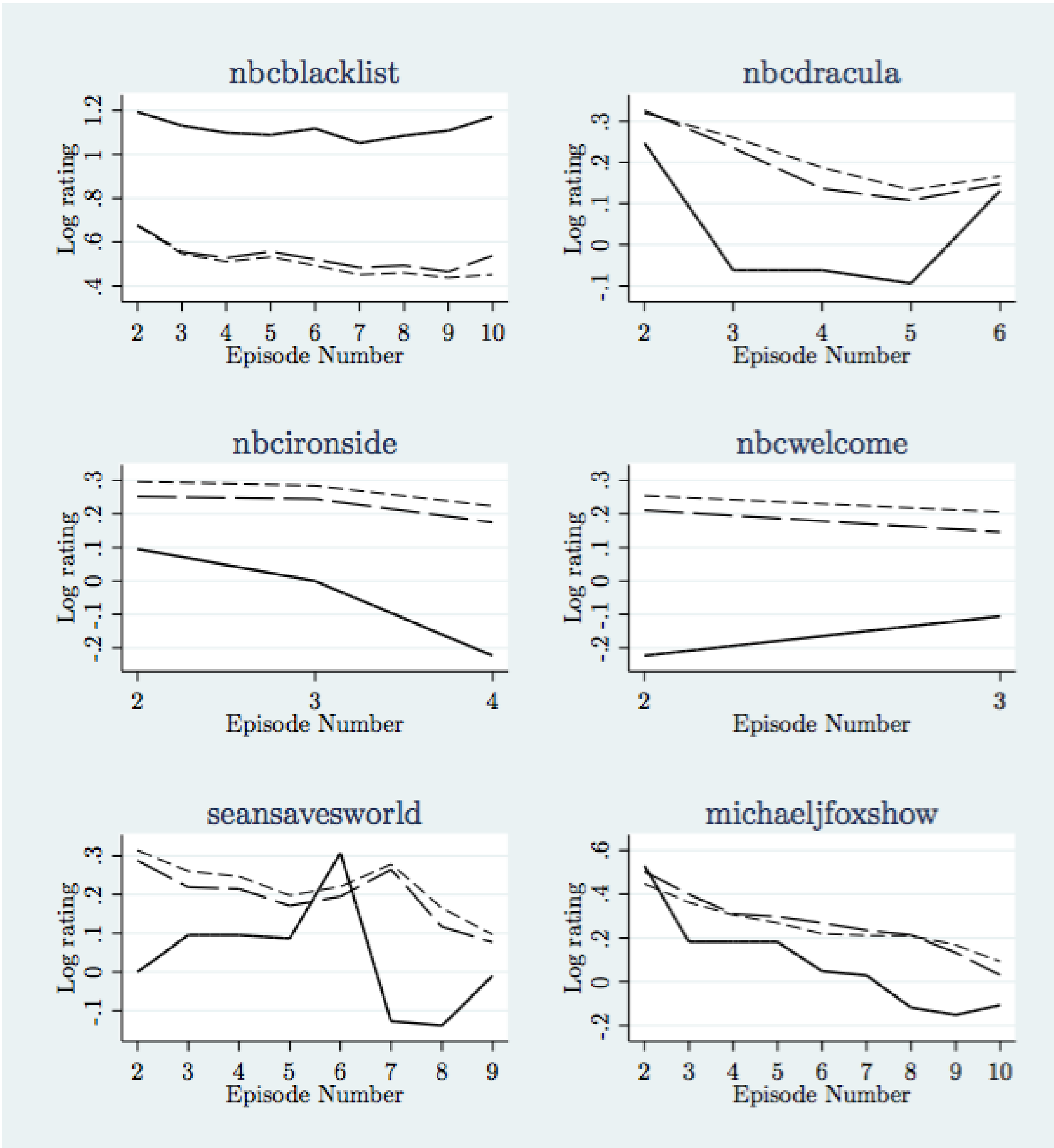


Figure A.3: Predicted Values from Social Media and Baseline Prediction Models (NBC Shows). Each panel shows how both the baseline prediction model and the social media prediction models compare to actual ratings for one show. The solid line represents actual *ratings*. The short dashed line is the baseline model. The long dashed line is the social media model.

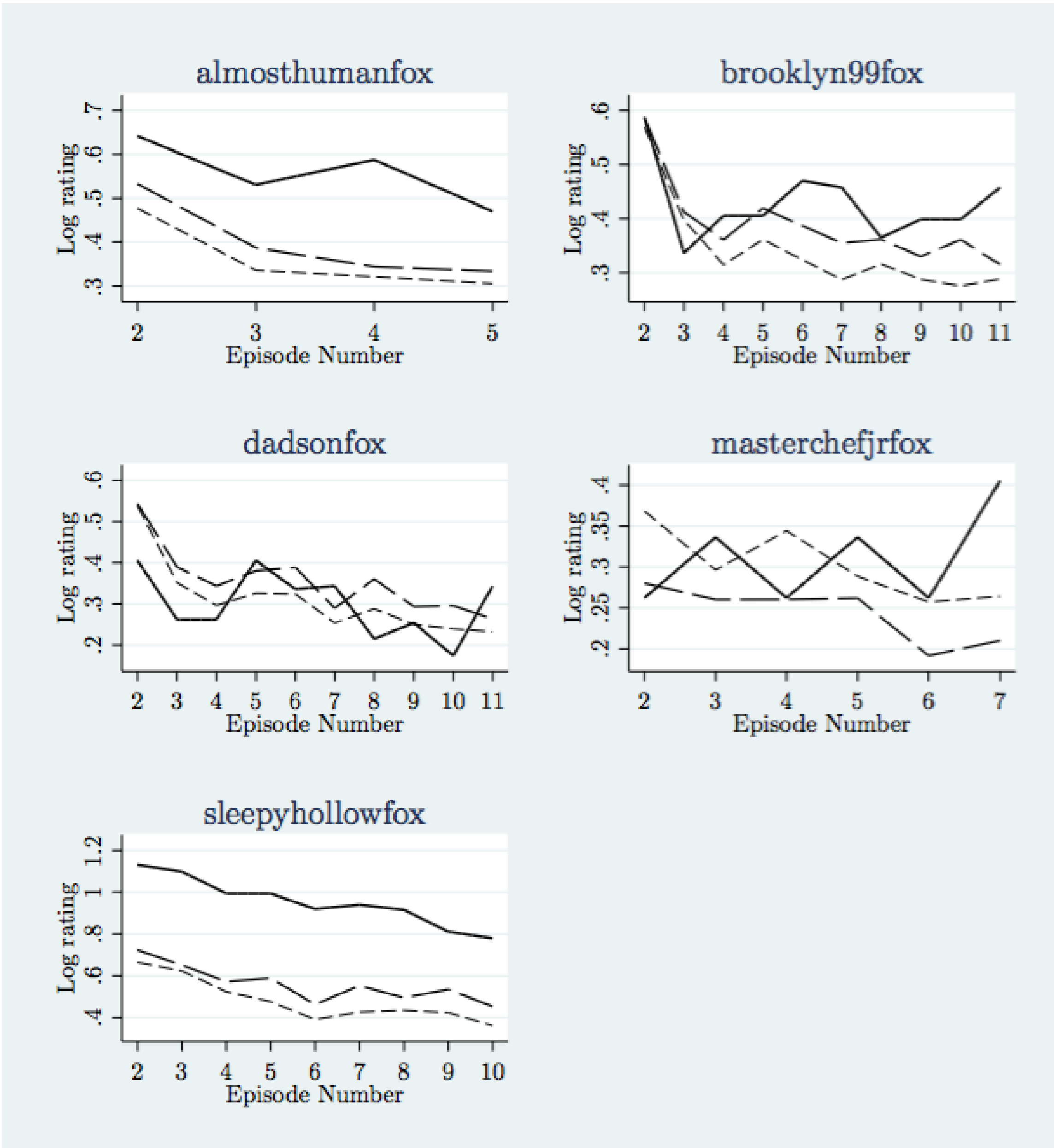


Figure A.4: Predicted Values from Social Media and Baseline Prediction Models (FOX Shows). Each panel shows how both the baseline prediction model and the social media prediction models compare to actual ratings for one show. The solid line represents actual *ratings*. The short dashed line is the baseline model. The long dashed line is the social media model.

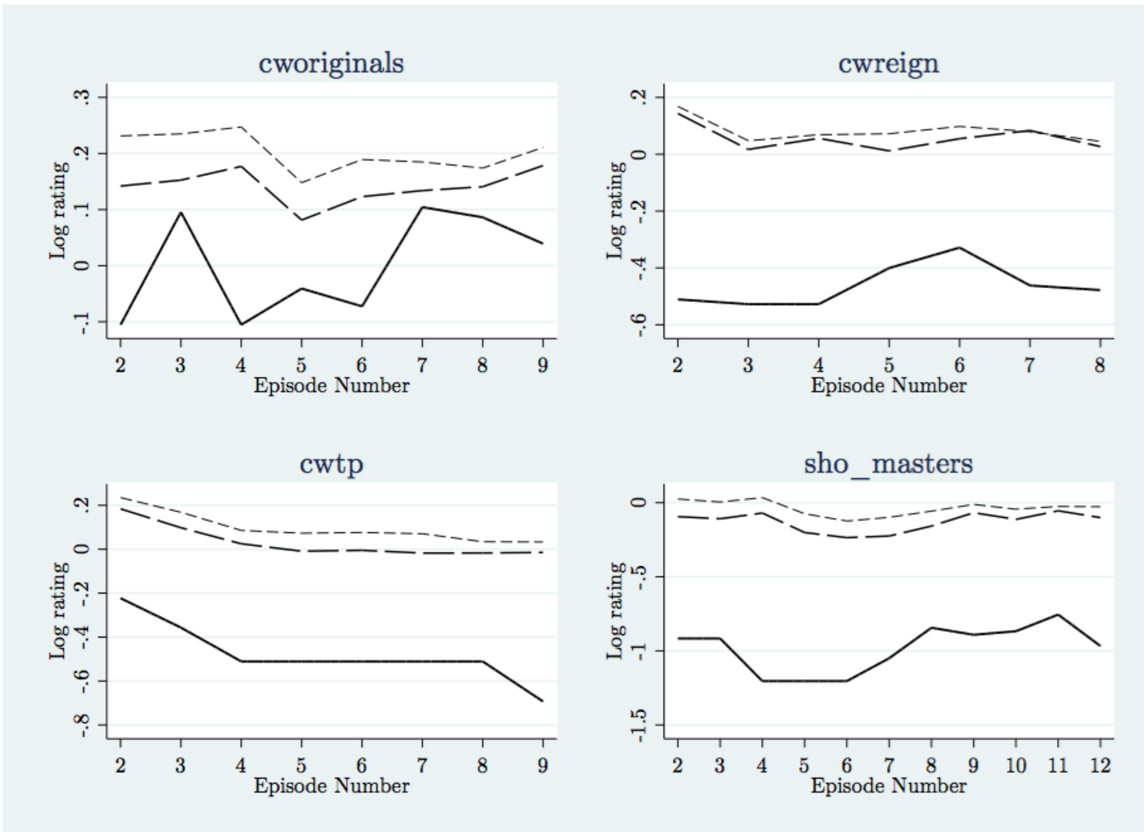


Figure A.5: Predicted Values from Social Media and Baseline Prediction Models (CW and Showtime Shows). Each panel shows how both the baseline prediction model and the social media prediction models compare to actual ratings for one show. The solid line represents actual *ratings*. The short dashed line is the baseline model. The long dashed line is the social media model.

Bibliography

Abrahamson, E. & Rosenkopf, L. (1997). Social network effects on the extent of innovation diffusion: A computer simulation. *Organization Science*, 8(3), 289–309.

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H. . H., & Liu, B. (2011). Predicting flu trends using twitter data. In *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on* (pp. 702–707).: IEEE.

Agrawal, A. K., Catalini, C., & Goldfarb, A. (2011). *The geography of crowdfunding*. Technical report, National Bureau of Economic Research.

Agrawal, A. K., Catalini, C., & Goldfarb, A. (2014). Some simple economics of crowdfunding. *Innovation Policy and the Economy*, 14(1), 63–97.

Alcacer, J. & Chung, W. (2007). Location strategies and knolwedge spillovers. *Management Science*, 53(5), 760–776.

Altman, E. J. (2015). Joining an ecosystem: Dependency challenges and paradox management strategies. *Working Paper*.

Ambrus, A. & Argenziano, R. (2009). Asymmetric networks in two-sided markets. *American Economic Journal: Microeconomics*, 1(1), 17–52.

Aral, S., Dellarocas, C., & Godes, D. (2013). Introduction to the special issue social media and business transformation: A framework for research. *Information Systems Research*, 24(1), 3–13.

Aral, S. & Walker, D. (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9), 1623.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8), 1485–1509.

Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2), 277.

Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics*, 37(3), 668–691.

Asur, S. & Huberman, B. A. (2010). Predicting the future with social media. *Arxiv preprint arXiv:1003.5699*.

Augereau, A., Greenstein, S., & Rysman, M. (2006). Coordination versus differentiation in a standards war: 56k modems. *The RAND Journal of Economics*, 37(4), 887–909.

Azoulay, P., Graff Zivin, J. S., & Sampat, B. N. (2010a). The diffusion of scientific knowledge across time and space: Evidence from professional transitions for the superstars of medicine. *Proceedings from the Rate & Direction of Inventive Activity*.

Azoulay, P., Zivin, J. S. G., & Wang, J. L. (2010b). Superstar extinction. *Quarterly Journal of Economics*, 125(2), 549–589.

Balasubramanian, N. & Lee, J. (2008). Firm age and innovation. *Industrial and Corporate Change*, 17(5), 1019–1047.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science*, 15(5), 215–227.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2010). Crowdfunding: tapping the right crowd.

Bickart, B. & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of interactive marketing*, 15(3), 31–40.

Binken, J. L. & Stremersch, S. (2009). The effect of superstar software on hardware sales in system markets. *Journal of Marketing*, 73(2), 88–104.

Bollen, J., Mao, H., & Zeng, X.-J. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.

Boudreau, K. J. (2012). Let a thousand flowers bloom? an early look at large numbers of software app developers and patterns of innovation. *Organization Science*, 23(5), 1409–1427.

Boudreau, K. J. & Hagiu, A. (2009). *Platform Rules: Multi-Sided Platforms as Regulators*. Edward Elger: London.

Bresnahan, T. F. & Yin, P.-L. (2005). Economic and technical drivers of technology choice: Browsers. *Annals of Economics and Statistics / Annales d'Économie et de Statistique*, (79/80), 629–670.

Brown, J. (2011). Quitters never win: The (adverse) incentive effects of competing with superstars. *Journal of Political Economy*, 119(5), 982–1013.

Caillaud, B. & Jullien, B. (2003). Chicken & egg: Competition among intermediation service providers. *The RAND Journal of Economics*, 34(2), 309–328.

Cantillon, E. & Yin, P.-L. . L. (2011). Competition between exchanges: A research agenda. *International journal of industrial organization*, 29(3), 329–336.

Chevalier, J. A. & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345–354.

Chintagunta, P. K. . K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, 29(5), 944–957.

Choi, H. & Varian, H. A. L. (2012). Predicting the present with google trends. *Economic Record*, 88, 2–9.

Chung, W. & Alcacer, J. (2002). Knowledge seeking and location choice of foreign direct investment in the united states. *Management Science*, 48(12), 1534–1554.

Claussen, J., Kretschmer, T., & Mayrhofer, P. (2013). The effects of rewarding user engagement: The case of facebook apps. *Information Systems Research*, 24(1), 186–200.

Cohen, W. M. & Levinthal, D. A. (1990). Absorptive capacity: a new perspective on learning and innovation. *Administrative science quarterly*, 35(1).

Das, S. R. & Chen, M. Y. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9), 1375–1388.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10), 1407–1424.

Dou, Y., Niculescu, M. F., & Wu, D. J. (2013). Engineering optimal network effects via social media features and seeding in markets for digital goods and services. *Information Systems Research*, 24(1), 164–185.

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *Stata Journal*, 3(2), 168–177.

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter? an empirical investigation of panel data. *Decision Support Systems*, 45(4), 1007 – 1016.

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). Social media update 2014.

Eisenhardt, K. M. & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal*, 21(10/11), 1105–1121.

Ellison, G. & Fudenberg, D. (2003). Knife-edge or plateau: When do market models tip? *The Quarterly Journal of Economics*, 118(4), 1249–1278.

Evans, D. S. & Schmalensee, R. (2010). Failure to launch: Critical mass in platform businesses. *Review of Network Economics*, 9(4).

Fang, X., Hu, P. J.-H., Li, Z. L., & Tsai, W. (2013). Predicting adoption probabilities in social networks. *Information Systems Research*, 24(1), 128–145.

Fudenberg, D. & Tirole, J. (1984). The fat-cat effect, the puppy-dog ploy, and the lean and hungry look. *The American Economic Review*, 74(2), 361–366.

Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, 43(7), 1239–1249.

Geroski, P. A. (2000). Models of technology diffusion. *Research Policy*, 29(4-5), 603–625.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31(3), 493–520.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4), 665–676.

Gilbert, R. J. & Newbery, D. M. G. (1982). Preemptive patenting and the persistence of monopoly. *The American Economic Review*, 72(3), 514–526.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012–4.

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

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., Libai, B., Sen, S., Shi, M., & Verlegh, P. (2005). The firm's management of social interactions. *Marketing Letters*, 16(3), 415–428.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with web search. *Proc Natl Acad Sci U S A*, 107(41), 17486–90.

Goh, K.-Y., Heng, C.-S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user- and marketer-generated content. *Information Systems Research*, 24(1), 88–107.

Gort, M. & Klepper, S. (1982). Time paths in the diffusion of product innovations. *The Economic Journal*, 92(367), 630–653.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica, Journal of the Econometric Society*, 25(4), 501–522.

Hagiu, A. (2009). Two-sided platforms: Product variety and pricing structures. *Journal of Economics & Management Strategy*, 18(4), 1011–1043.

Helfat, C. E. & Lieberman, M. B. (2002). The birth of capabilities: market entry and the importance of pre-history. *Industrial and Corporate Change*, 11(4), 725–760.

Hendel, I., Nevo, A., & Ortalo-Magne, F. (2009). The relative performance of real estate marketing platforms: Mls versus fsbomadison. com,. *American Economic Review*, 99(5), 1878–1898.

Henderson, R. (1993). Underinvestment and incompetence as responses to radical innovation: evidence from the photolithographic alignment equipment industry. *RAND Journal of Economics*, 24(2), 248–270.

Henderson, R. M. & Clark, K. B. (1990). Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 35(1).

Hirschman, A. O. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Harvard University Press.

Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organization science*, 2(1), 88–115.

Jin, G. Z. & Rysman, M. (2013). *Platform pricing at sports card conventions*. Technical report, National Bureau of Economic Research.

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). Understanding the predictive power of social media. *Internet Research*, 23(5), 544–559.

Kappel, T. (2008). Ex ante crowdfunding and the recording industry: A model for the us. *Loy. LA Ent. L. Rev.*, 29, 375.

Katz, M. L. & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American economic review*, 75(3), 424–440.

Katz, M. L. & Shapiro, C. (1986). Technology adoption in the presence of network externalities. *The Journal of Political Economy*, 94(4).

Kickstarter (2012). Blockbuster effects.

Kickstarter (2013a). The truth about spike lee and kickstarter.

Kickstarter (2013b). Who is kickstarter for?

Kickstarter (2014a). Introducing launch now and simplified rules.

Kickstarter (2014b). Introducing two new categories: Journalism and crafts.

Kickstarter (2014c). A subcategory for everything.

Kickstarter (2014d). Thanks a billion.

King, A. A. & Tucci, C. L. (2002). Incumbent entry into new market niches: The role of experience and managerial choice in the creation of dynamic capabilities. *Management Science*, 48(2), 171–186.

Kuppuswamy, V. & Bayus, B. (2014). Crowdfunding creative ideas: The dynamics of project backers in kickstarter. *SSRN eLibrary*.

Lambert, T. & Schwienbacher, A. (2010). An empirical analysis of crowdfunding. In *Social Science Research Network, http://ssrn. com/abstract*, volume 1578175.

Lampos, V. & Cristianini, N. (2012). Nowcasting events from the social web with statistical learning. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(4), 72.

Lee, R. S. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review*, 103(7), 2960–3000.

Levinthal, D. A. (1991). Organizational adaptation and environmental selection-interrelated processes of change. *Organization Science*, 2(1), 140–145.

Levinthal, D. A. (1997). Adaptation on rugged landscapes. *Management science*, 43(7), 934–950.

Lieberman, M. B. & Montgomery, D. B. (1988). First-mover advantages. *Strategic management journal*, 9(S1), 41–58.

Liebowitz, S. J. & Margolis, S. E. (1994). Network externality: An uncommon tragedy. *The Journal of Economic Perspectives*, (pp. 133–150).

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization science*, 2(1), 71–87.

Meyer, J. W. & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340–363.

Miller, A. R. & Tucker, C. (2013). Active social media management: The case of health care. *Information Systems Research*, 24(1), 52–70.

Mitchell, W. C. & Munger, M. C. (1991). Economic models of interest groups: An introductory survey. *American Journal of Political Science*, 35(2), 512.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, 29(1), 1–16.

Mollick, E. & Nanda, R. (2015). Wisdom or madness? comparing crowds with expert evaluation in funding the arts. *Management Science*.

Nagle, F. (2015). Stock market prediction via social media: The importance of competitors. *Working paper*.

Nelson, R. R. (1995). Recent evolutionary theorizing about economic change. *Journal of economic literature*, 33(1), 48–90.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6), 1417.

Norton, J. A. & Bass, F. M. (1987). A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Management science*, 33(9), 1069–1086.

Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal*, 18(S1), 187–206.

O'Reilly, C. A. & Tushman, M. L. (2008). Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. *Research in Organizational Behavior*, 28, 185 – 206.

Parker, G. G. & Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management Science*, 51(10), 1494–1504.

Prahalad, C. K. & Bettis, R. A. (1986). The dominant logic: A new linkage between diversity and performance. *Strategic Management Journal*, 7(6), 485–501.

Reinganum, J. F. (1983). Uncertain innovation and the persistence of monopoly. *The American Economic Review*, 73(4), 741–748.

Rishika, R., Kumar, A., Janakiraman, R., & Bezawada, R. (2013). The effect of customers' social media participation on customer visit frequency and profitability: An empirical investigation. *Information Systems Research*, 24(1), 108–127.

Rivkin, J. W. & Siggelkow, N. (2003). Balancing search and stability: Interdependencies among elements organizational design. *Management Science*, 49(3), 290–311.

Rochet, J. C. & Tirole, J. (2006). Two-sided markets: a progress report. *The RAND Journal of Economics*, 37(3), 645–667.

Rochet, J.-C. . C. & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4), 990–1029.

Rogers, E. M. (1995). *Elements of Diffusion*, (pp. 1–37). The Free Press: New York.

Rumelt, R. P. (1982). Diversification strategy and profitability. *Strategic Management Journal*, 3(4), 359–369.

Rysman, M. (2004). Competition between networks: A study of the market for yellow pages. *The Review of Economic Studies*, 71(2), 483–512.

Rysman, M. (2009). The economics of two-sided markets. *The Journal of Economic Perspectives*, 23(3), 125–143.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69, 239–241.

Seamans, R. & Zhu, F. (2014). Responses to entry in multi-sided markets: The impact of craigslist on local newspapers. *Management Science*.

Shankar, V. & Bayus, B. L. (2003). Network effects and competition: an empirical analysis of the home video game industry. *Strat. Mgmt. J.*, 24(4), 375–384.

Silverman, B. S. (1999). Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Science*, 45(8), 1109–1124.

Sonnier, G. P., McAlister, L., & Rutz, O. J. (2011). A dynamic model of the effect of online communications on firm sales. *Marketing Science*, 30(4), 702–716.

Sundararajan, A. (2007). Local network effects and complex network structure. *The BE Journal of Theoretical Economics*, 7(1).

Sweeting, A. (2012). Dynamic pricing behavior in perishable goods markets: Evidence from secondary markets for major league baseball tickets. *Journal of Political Economy*, 120(6), 1133–1172.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533.

Tripsas, M. & Gavetti, G. (2000). Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic Management Journal*, 21(10/11), 1147–1161.

Tsukayama, H. (2013). Twitter turns 7: Users send over 400 million tweets per day.

Tucker, C. (2008). Identifying formal and informal influence in technology adoption with network externalities. *Management Science*, 54(12), 287–304.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10, 178–185.

Tushman, M. & Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative Science Quarterly*, 31(3), 439–465.

Twitter, I. (2015). Twitter reports fourth quarter and fiscal year 2014 results.

Von Hippel, E. (1986). Lead users: A source of novel product concepts. *Management science*, (pp. 791–805).

von Hippel, E. (2005). *Democratizing innovation*. Cambridge, Mass.: MIT Press, illustrated edition.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, 5(2), 171–180.

Wernerfelt, B. (1995). The resource-based view of the firm: Ten years after. *Strategic Management Journal*, 16(3), 171–174.

Weyl, E. G. (2010). A price theory of multi-sided platforms. *The American Economic Review*, 100(4), 1642–1672.

Woolridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.

Wu, J., Sun, H., & Tan, Y. (2013). Social media research: A review. *Journal of Systems Science and Systems Engineering*, 22(3), 257–282.

Wu, L. & Brynjolfsson, E. (2013). The future of prediction: How google searches foreshadow housing prices and sales. In *Economics of Digitization*. University of Chicago Press.

Zeng, X. & Wei, L. (2013). Social ties and user content generation: Evidence from flickr. *Information Systems Research*, 24(1), 71–87.

Zhu, F. & Iansiti, M. (2012). Entry into platform-based markets. *Strategic Management Journal*, 33(1), 88–106.

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