NBER WORKING PAPER SERIES

EXIT, TWEETS AND LOYALTY

Joshua S. Gans
Avi Goldfarb
Mara Lederman

Working Paper 23046
http://www.nber.org/papers/w23046

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2017

We gratefully acknowledge financial support from SSHRC (grant # 493140). The paper benefited from helpful comments from Severin Borenstein, Judy Chevalier, Isaac Dinner, Francine Lafontaine, Dina Mayzlin, Amalia Miller and seminar participants at the University of Toronto, UC-Berkeley, the University of Minnesota, the University of North Carolina, Ebay, Facebook, the 2016 ASSA meetings, ZEW at Mannheim, the Searle Annual Antitrust Conference, the University of British Columbia, Harvard University, the NBER Summer Institute, the NBER Organizational Economics Working Group, Stanford University and Carnegie Mellon University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Exit, Tweets and Loyalty
Joshua S. Gans, Avi Goldfarb, and Mara Lederman
NBER Working Paper No. 23046
January 2017
JEL No. L13,L14,L93

ABSTRACT

Hirschman's Exit, Voice, and Loyalty highlights the role of "voice" in disciplining firms for low quality. We develop a formal model of voice as a relational contact between firms and consumers and show that voice is more likely to emerge in concentrated markets. We test this model using data on tweets to major U.S. airlines. We find that tweet volume increases when quality – measured by on-time performance – deteriorates, especially when the airline operates a large share of the flights in a market. We also find that airlines are more likely to respond to tweets from consumers in such markets.

Joshua S. Gans
Rotman School of Management
University of Toronto
105 St. George Street
Toronto ON M5S 3E6
CANADA
and NBER
joshua.gans@gmail.com

Avi Goldfarb
Rotman School of Management
University of Toronto
105 St. George Street
Toronto, ON M5S 3E6
CANADA
and NBER
agoldfarb@rotman.utoronto.ca

Mara Lederman
Joseph L. Rotman School of Management
University of Toronto
105 St. George Street
Toronto, Ontario M5S 3E6 Canada
Mara.Lederman@Rotman.Utoronto.Ca

1 Introduction

At the heart of economics is the belief that markets act to discipline firms for poor performance. While the role of markets in influencing firm behavior has been extensively studied, an alternative mechanism has received considerably less attention from economists. In his famous work, *Exit, Voice and Loyalty*, Albert Hirschman distinguishes two actions consumers might take when they perceive quality to have deteriorated: **exit** (withdrawing demand from a firm) and **voice** (supplying information to the firm). Hirschman defines voice as "Any attempt at all to change, rather than escape from, an objectionable state of affairs whether through individual or collective petition to the management directly in charge, through appeal to a higher authority with the intention of forcing a change in management or through various types of actions and protests, including those that are meant to mobilize public opinion." (p. 30) Hirschman offers many examples of the choice between exit and voice, including the case of school quality: parents who are unhappy with their child's school can either switch schools (exit) or complain to the principal and school board (voice). Exit may be particularly costly in this situation as it could involve moving, and so, Hirschman argues, many people may choose voice. While there is evidence that consumers exercise voice via complaints,[1] there has been little empirical work on the fundamental idea proposed by Hirschman: that exit and voice are, in fact, alternative ways to achieve the same thing, with each emerging under different market conditions.

In this paper, we begin to fill this void. We theoretically model and empirically study the relationship between voice and market structure. Hirschman himself points out that this relationship is not straightforward. On the one hand, the use of voice might grow as market concentration increases because the opportunities for exit decrease. On the other hand, since voice is more likely to be effective if backed by the threat of exit, the use of voice might decrease as market concentration increases because of the threat of exit becomes less credible. In the extreme case of monopoly, he argues that voice is the only available option but also unlikely to have much

[1] Richins (1983) examines why people complain and emphasizes what she calls "vigilantism." Gatignon and Robertson (1986) examine positive and negative word of mouth, with an emphasis on cognitive dissonance for negative and altruism and reciprocity for positive. Forbes (2008) shows that complaints are impacted by customer expectations. Beard, Macher, and Mayo (2015) explore exit and voice more directly in the context of complaints to the FCC about local telephone exchanges, and we discuss their work in further detail below.

impact. Thus, the equilibrium relationship between market structure and the use of voice is ambiguous.

To resolve this ambiguity, we model the interactions between consumers and a firm as a relational contract in which consumers use voice to alert the firm to quality deteriorations in exchange for a "concession." A key insight of our model is that, as competition decreases, the value to the firm of retaining a customer increases because the margins earned from the customer are higher. We show that there are conditions under which a relational contract with voice is an equilibrium of a repeated game and that, as competition in a market becomes stronger, those conditions become less likely to hold. Thus, our model predicts that voice is more likely to be observed when firms have a dominant position in a market.

We then turn to measuring the relationship between quality, market structure, and voice. Empirically studying this relationship is challenging. First, voice has historically been difficult to observe in a systematic way. As Beard, Macher, and Mayo (2015, p. 719) note in their study of voice in telecommunications, "[f]irms are simply not inclined to publicize their shortcomings. Consequently, the ability of researchers to directly observe and study data on complaints is limited." Second, voice is influenced by both quality and market structure but quality itself may be a function of market structure. As a result, unless quality is carefully controlled for, it may confound the estimated relationship between market structure and voice. For example, if market power incentivizes firms to degrade quality, then an analysis of the relationship between market structure and complaints might find more voice in concentrated markets even if there is little direct impact of market structure on voice.

We develop an empirical strategy that allows us to overcome both of these challenges. Our setting is the U.S. airline industry and we measure voice using the millions of comments, complaints, and compliments that consumers make to or about airlines via the social network Twitter. Whereas most traditional channels for complaints are private and observed only by firms, Twitter's public nature (the unit of communication – the 'tweet' – is public by default) provides us with a way of collecting systematic data on voice, albeit only voice exercised via this particular medium. While Twitter serves this role in many industries, several features of the airline industry (and the data available for this industry) allow us to develop an empirical strategy that overcomes the endogeneity issue described above. Specifically, the airline industry is comprised of a large number of local markets each with its own market structure. While market structure may influence

quality in this industry, one of the most important dimensions of quality – on-time performance – varies within markets and can be precisely measured. We exploit daily variation in an airline's on-time performance *within* a given market to estimate the relationship between quality and voice (as measured by daily tweet volume), while controlling for the underlying relationship between market structure and quality. We then exploit variation in market structure *across* cities to estimate how the relationship between quality and voice varies with market structure. Thus, rather than estimate the relationship between market structure and voice across markets, we estimate the relationship between quality deterioration and voice within a market and then how this relationship varies across markets with different market structures.

Our analysis combines three types of data. The first – and most novel – is a dataset that includes all tweets made between August 1, 2012 and July 31, 2014 that mention or are directed to one of the seven major U.S. airlines. This dataset includes several million tweets. For many of these tweets, we can identify the geographic location of the tweeter at the time of posting the tweet as well as the tweeter's home city, thus allowing us to link tweets to both a specific airline and a specific market. We use the tweet-level data to create a measure of the amount of voice directed at a given airline on a given day from consumers in a given market. We then combine this with data from the U.S. Department of Transportation (DOT) on the on-time performance of every domestic flight and data on airlines' flight schedules which allow us to construct measures of airport or city market structure.

Our empirical analysis delivers several interesting findings and supports the predictions of our model. First, we find that consumers do indeed respond to quality reductions via voice. In both simple descriptive analyses and across a variety of regression specifications, we find that the number of tweets that an airline receives on a given day from individuals in a given market increases as its on-time performance in that market deteriorates. This result is robust to alternative ways of matching tweets to locations and alternative ways of measuring on-time performance. In addition, when we consider the content of the tweets, we find that this relationship is strongest for tweets with a negative sentiment and tweets that include words related to on-time performance. We believe that our analysis is the first to provide systematic and large-scale evidence that consumers do respond to quality deterioration via voice.

Second, we find that the relationship between quality deterioration and tweet volume is stronger when the offending airline dominates an airport. It is well established that airport

dominance translates into route-level market power and higher fares (Borenstein, 1989 and 1991). Our finding that the relationship between quality deterioration and voice is stronger for dominant airlines is therefore consistent with the main prediction of our relational contracting model – that voice is more likely to emerge in concentrated markets where margins are higher and customers more valuable. Thus, our model and empirical findings serve to resolve the ambiguity in Hirschman about the relationship between market structure and voice.

Finally, the results of our analysis of airline responses are also consistent with the relational contracting model we propose. When we examine data on a sample of airline responses to tweets, we find that airlines are most likely to respond to tweets from their most valuable customers, defined as customers who are from a market where the airline is dominant or customers who mention the airline's frequent flier program in their tweet. This result is more speculative because we only have data on public responses by the airline through Twitter and hence do not observe all ways in which airlines can respond to complaints (for example, direct messaging, quality improvements, and email). Nevertheless, over 20% of tweets receive responses and these responses display a pattern that is consistent with a key prediction of our model – that airlines' incentives to respond to voice are higher when customers are more valuable to them. Furthermore, we find that twitter users are more likely to tweet again to an airline if the airline has responded to their first tweet (that we observe).

Hirschman's *Exit, Voice, and Loyalty* received a great deal of attention after its release, with glowing reviews in top journals in political science and economics (Adelman 2013) and a debate about the breadth of its applicability in the 1976 *American Economic Review Papers & Proceedings* (Hirschman 1976; Nelson 1976; Williamson 1976; Freeman 1976; Young 1976). Despite this attention, formal modeling and modern empirical work have been limited. Fornell and Wernerfelt (1987, 1988) develop formal models of the ideas in *Exit, Voice, and Loyalty* and emphasize that – when product or service failures are difficult for a firm to observe – firms will want to facilitate complaints in order to learn about their own quality. Abrahams et al (2012) shows that firms can discover product deterioration via voice, by studying evidence of vehicle defects that arises through social media. Other work has explored incentives to contribute to social media platforms (Trusov, Bucklin, and Pauwels 2009; Berger and Schwartz 2011; Miller and Tucker 2013; Wei and Xiao 2015) and the motivations to provide, and the consequences of, online reviews (e.g. Mayzlin (2006), Godes and Mayzlin (2004, 2006), Chevalier and Mayzlin (2006), Mayzlin,

Dover, and Chevalier (2014)). Nosko and Tadelis (2015) are able to link data on seller quality and transactions at the buyer level and show that buyers who have a more negative experience on eBay are more likely to exit (i.e.: less likely to transact again on the platform).

The most closely related research to our work is Beard, Macher, and Mayo (2015). They also study customer complaints using the lens of *Exit, Voice and Loyalty*. They examine complaints to the U.S. Federal Communications Commission about telecommunications companies. They estimate the relationship between complaints and market structure, while controlling for consumer perceptions of quality, and find that markets that are more competitive were associated with fewer complaints. Our empirical strategy is different in that we estimate the relationship between quality deterioration and voice within a market, and how this relationship varies with market structure. More importantly for exploring Hirschman's predictions, our data come from consumer complaints aimed at firms rather than from consumer complaints to a government regulator.

Overall, we believe this paper makes several contributions. First, we provide the first systematic evidence that consumers do indeed exercise voice in response to quality deterioration and that Twitter serves as a platform for such voice. Second, we present a formal model of the relationship between quality, voice, and market structure that offers a way to resolve the ambiguity in this relationship as presented by Hirschman. While Hirschman focused on how consumers' incentives to exercise voice vary with market structure, we also consider how firms' incentives to respond to voice vary with market structure. Accounting for the firm's incentives is what allows us to develop an equilibrium model of voice and comparative statics with the number of firms in the market. This relational contracting framework offers a conceptualization of voice as a mechanism for preserving valuable long-term relationships between customers and firms. We believe that this can be a useful way to model the role of voice in many markets. Third, we show that, in our setting, the responsiveness of voice to quality deterioration is greater in concentrated markets, consistent with the relational contracting model. Finally, the empirical strategy we develop, which exploits high-frequency within-market changes in quality, may offer a fruitful way of exploring these relationships in other settings.

The remainder of this paper is organized as follows. In the next section, we lay out the theoretical considerations. In Section 3, we highlight how Twitter serves as an instrument for

voice. Section 4 describes our sources of data and sample construction, and Section 5 discusses our empirical approach. Section 6 presents our results. A final section concludes.

2 Theoretical Considerations

In his treatise, Hirschman saw exit and voice as two actions that consumers might take to discipline a firm after they had noted a decline in quality. As the introduction of voice was, at that time, novel in economics, Hirschman argued that it was unclear whether voice was an alternative to exit or something that might be used in conjunction with it. Specifically, when he considered what consumers might do if their supplier was a pure monopoly, he saw voice as the only option and (extrapolating somewhat) as a residual that is exercised whenever opportunities for exit are removed. Nonetheless, Hirschman noted that, from the perspective of the firm, voice can complement exit in signalling issues within the firm that should be addressed. Moreover, to the extent that voice can prevent exit, voice gives the firm the opportunity to improve performance without suffering irreparable harm. However, Hirschman then questioned whether consumers would go to the trouble of exercising voice in the absence of a credible exit option to back them up. Thus, Hirschman realized that the use of voice might occur more often when exit opportunities (i.e., competition) were readily available.[2] As Hirschman wrote, "[t]he relationship between voice and exit has now become more complex. So far it has been shown how easy availability of the exit option makes the recourse to voice less likely. Now it appears that the effectiveness of the voice mechanism is strengthened by the possibility of exit. The willingness to develop and use the voice mechanism is reduced by exit, but the ability to use it with effect is increased by it." (p.83)

While Hirschman made numerous conjectures and arguments about the relationship between a consumer's choices between exit and voice and competition, to date there exists no formal model of that relationship; specifically, for variation in concentration among oligopolists. Here, we blend the third important aspect of Hirschman's work – *loyalty* – to provide that model. In an analogous way to a principal using an incentive contract to ensure that the quality of an agent's work is high, we consider a contract between the consumer (akin to the principal) and the

[2] Hirschman appears to reach no precise statement regarding the relationship between voice and competition but eventually becomes more interested in the notion that a monopoly, because it could possibly receive more voice than a competitive firm, might end up performing better than competitive firms. We note that this conjecture hinges on the proposition that voice is more likely to arise, and to generate a response, in a market with a monopolist rather than a market with competition.

firm (here the agent) to ensure that if the latter supplies lower than expected product quality, they will compensate the former. The special difficulty is that product quality is non-contractible (i.e., it is observable to both firm and consumer but is not verifiable by a third party). Thus, having already consumed a product and paid for it, a consumer must rely upon a firm fulfilling a promise for recompense that is not contained in a formal contract. The consideration of loyalty comes into play because we assume that what allows that promise to be credible is the expectation of repeated transactions between the consumer and the firm. This is an often-used game-theoretic notion of loyalty – in this case, the consumer's loyalty to the firm. In the absence of such loyalty, for instance, if consumers more randomly chose firms each period, there is no scope for a firm's promise to be made credible and, as we will show, no reason for the consumer to exercise voice. Here we provide a simple model based on a relational contract between a firm and each of its customers. While this model is straightforward, we believe it highlights the first order trade-offs involved and provides the sharp statement missing from the prior informal literature.

2.1 Formal Model

There is a continuum of consumers and $n \geq 2$ symmetric firms in a market with constant marginal supply costs of c per unit. Consider a consumer and their current supplier. The consumer demands one unit at each unit of time and the firms' products are perfect substitutes except that a consumer has an *infinitesimal preference* to stay with the firm it chose in the previous period. The firm and consumer have a common discount factor of δ.

The stage game of our model is as follows:

1. (Pricing) Firms announce prices to the consumer and the consumer selects a firm to purchase from.
2. (Quality Shock) With probability s, the consumer receives an unexpected quality drop on a product they have already purchased. This results in an immediate loss in consumer surplus of Δ which is the same for any consumer suffering the loss.
3. (Voice) The consumer can, at a one-time cost of C, communicate their dissatisfaction to the firm.
4. (Mitigation) If the consumer has complained, the firm can offer the consumer a concession of B (where B is a choice variable on the real line).

5. (Exit) The consumer chooses whether to stay with the firm or exit. Exit means committing to a different supplier next period.

Based on the stage game alone, the firm will offer the consumer no concession ($B = 0$) and the consumer will not exercise voice. This is because a concession will not alter the exit decision of the consumer and hence, cannot be credibly promised. Thus, the possibility of a concession and an observation of voice depends on the impact on future sales to the consumer - i.e., a consumer's expected loyalty.

Suppose that both the firm and consumer play a repeated game. Following Levin (2002) we consider the consumer as forming a relational contract with the firm where the firm promises the consumer a concession of *B* if the consumer alerts the firm to a quality drop. We assume that the quality drop is ex post verifiable by the firm.[3] Formally:

Definition. *A* (symmetric) relational contracting equilibrium with voice *exists if (i) a consumer exercises voice if and only if they observe a quality shock; (ii) all firms offer a concession, B, if the consumer has exercised voice; and (iii) a consumer exits their firm in the period following the exercise of voice if no concession is given.*

Clearly, the final element of the consumer's strategy in this definition involves a threat to exit which is not exercised on the equilibrium path.

What level of concession (*B*) will allow this relational contract to be an equilibrium of the proposed repeated game? First, consider the cost to a firm of losing a consumer. As each consumer prefers to stay, marginally, with its current firm, if a firm loses a consumer, it cannot attract another. Thus, it loses:

$$\frac{\delta}{1-\delta}(p(n,B) - c - sB).$$

Equilibrium price, $p(n,B)$, is written as a function of both the number of firms, *n*, and the symmetric concession offered by firms, *B*. As is common, *p* is assumed to be decreasing in *n*. Note that $p(n,B)$ is increasing in *B*. To see this, observe that, if $p(n,B) = m(n,B)(c + sB)$ (where *m* is a firm's mark-up and $c + sB$ is a firm's full marginal cost), each component is increasing in *B*.

Importantly, the *cost to the firm of a consumer choosing exit is increasing in market concentration* (i.e., with a fall in *n*). The intuition is that, when market concentration is high, the firm earns high margins from each consumer and faces larger costs should the consumer exit. Thus,

[3] This eliminates the notion of a false complaint by the consumer. However, it is not observable by third parties ruling out a formal contractual commitment. This is an interesting issue that we leave for future research.

absent other considerations, firms with greater degrees of market power face incentives to find ways to convince consumers to exercise voice and credibly promise recompense rather than lose those consumers in the face of a quality shock.

Second, a necessary condition for a consumer to exercise voice is that $B \geq C$. If this condition did not hold, then even if the consumer expects a concession, they would not file a complaint as the costs of voice would outweigh the benefit they would receive.

Third, what happens if a consumer exits? As there is a continuum of consumers, there will be no impact on the price in the market.[4] Similarly, if a relational contracting equilibrium with voice otherwise exists, the consumer can expect to receive additional utility of $s(B - C)$ by switching to another firm for which the relational contract is expected to hold. The consumer will lose the infinitesimal advantage to their present supplier, however, as this arises for whomever the consumer's supplier is in the next period, that shortfall will be temporary. Moreover, for this reason, the firm will not be able to replace, in the subgame following exit, the consumer with another.

Given the above discussion, we can now consider whether a relational contracting equilibrium with voice exists. Specifically, is there a *B* that the firm will offer to prevent exit and the consumer will accept to keep from exiting? That *B* must satisfy:

$$\frac{\delta}{1-\delta}(p(n,B) - c - sB) \geq B \Longrightarrow \frac{\delta}{1-\delta(1-s)}(p(n,B) - c) \geq B$$

$$B \geq C$$

The first incentive constraint is for the firm and says that the expected future value of a consumer is greater than the cost of providing a concession today. The second incentive constraint is for the consumer and says that the concession must induce the consumer to incur the costs of voice and not exit the firm.

Putting the two constraints together, we can see that a sufficient condition for a relational contracting equilibrium to exist is that:[5]

---

[4] One can imagine situations where there will be an impact on the price a consumer faces if they exit and commit not to consider their current supplier in the future. We explore this situation in the online appendix. For instance, price may be determined in a search model in which case the consumer may end up facing higher prices when removing a firm from its consideration list. Nonetheless, ultimately, we demonstrate that, accounting for potentially higher prices or other costs of exit, does not change the qualitative prediction of our model as the first order effects we identify here can still dominate.

[5] Here we substitute *C* for *B* in the pricing function as price is non-decreasing in *B*; making this a sufficient condition. A necessary condition would be there exists $B > C$ such that (*) for *B* in the pricing function.

$$\frac{\delta}{1-\delta(1-s)}(p(n,C)-c) \geq C \quad (*)$$

The following proposition summarizes the properties of this equilibrium:

Proposition 1. *A relational contracting equilibrium with voice exists for sufficiently high δ and low C. A relational contracting equilibrium does not exist for n sufficiently large.*

The first part of the proposition follows from the usual assumptions for the folk theorem in repeated games. The second part follows because the LHS of (*) is decreasing in *n* and converges to 0 whereas the RHS does not change in *n* and is positive.

The model confirms Hirschman's intuition that market power plays an important role in the efficacy of voice. However, it shows also that the future value of a customer to the firm plays a critical role in determining whether a consumer believes that exercising voice will be consequential. Hence, the higher is δ, the more the firm values its future margins from the customer and the more likely we are to observe voice in equilibrium.

The model highlights why Hirschman's informal intuition caused confusion as the impact of market concentration on voice does not operate in the same way at the extremes of pure monopoly and perfect competition. On the monopoly side, what happens if $n = 1$? In that case, should a consumer exit, the consumer has no other option and so loses all of the consumer surplus associated with the relationship. Importantly, this may render a relational contract with voice non-existent because exit is never credible as a consumer who complains but does not obtain a response comes 'crawling back.' When there is some competition, a consumer's threat to exit the firm forever can become credible as, in the relational contracting equilibrium, the consumer believes (a) that its current firm will not honor future promises and (b) that it only faces an infinitesimal cost for a single period if it exits the firm and chooses another. In other words, it will not come 'crawling back.' While (a) is also true for a pure monopoly situation, (b) is not and the consumer faces large costs if it does not return to the firm. Thus, for a monopoly situation, the firm may not offer a sufficient recompense to induce the consumer to exercise the costs associated with voice.

In the case of perfect competition (as n goes to infinity), then $p(n,C) \to c + sC$. Importantly, the firm no longer earns a positive margin from a consumer. In this situation, as demonstrated in Proposition 1, there will be no level of *B* that it would pay to retain a consumer regardless of other parameters. Thus, in this case, voice would not be exercised because the consumer would not expect the firm to respond to it. The key idea here is that an equilibrium with

voice is more likely as concentration falls; however, this result is potentially undermined at the extremes of pure monopoly and perfect competition but for distinct reasons.[6]

Our model presents the relational contract between a consumer and a firm as a grim trigger strategy whereby exit occurs if the consumer receives a quality decline without a concession. While this concession could encompass an actual payment or gift to the consumer, our model is consistent with a more general interpretation. For instance, a consumer who lodges a complaint may not expect an actual response but instead expect an improvement in the future (for instance, a reduced rate of quality decline). If the issues continued, then the consumer could engage in exit in the future without exercising additional voice. For this reason, the model is a predictor of consumer exercise of voice more than it is a predictor of the cause of the voice or the nature of the response. Thus, a consumer might complain for issues outside of the firm's control (say, a weather interruption) but not expect an explicit response unless other issues arose (such as the inability of the firm to reallocate resources in response to the adverse event). The key factor in predicting voice is that the consumer considers the likelihood that a firm will care to retain them rather than let them exit and this is what drives the decision to delay exit in favour of voice.

Of course, voice might arise for other reasons as well. Some people may gain utility from exercising voice (i.e., $C < 0$ for them) or, alternatively, exercise voice for pro-social reasons to signal issues with the firm to others. The relationship implied by Proposition 1, however, requires that there exist consumers for whom $C > 0$ and who receive no significant benefits from voice other than a firm response. Finally, while our model has focussed on the industrial organization drivers of voice, it is also possible that firms will encourage voice to learn about and respond to quality reductions. For instance, firms may want to use consumers to monitor employee performance and therefore encourage complaints or ratings of employees or agents. Of course, monitoring can also be achieved by exit and so it is possible to imagine that the firm's incentives to invest in organizational structures that are more responsive to voice may be related to the same considerations that drive the relational contract examined here (see Fornell and Wernerfelt (1987, 1988) for a formal analysis of complaints as monitoring).

[6] We explored variants of the model presented here. For instance, in the online appendix, we consider the full equilibrium outcome in a Cournot model that endogenized $p(n, B)$ in order to determine whether symmetric firms would choose to adhere to the proposed relational contract when others did so; confirming this is a full equilibrium outcome.

2.2 Implications for Empirical Analysis

Our model predicts that voice is more likely to be an equilibrium when market concentration is higher. Estimating the relationship between voice, quality deterioration, and market concentration is therefore the primary focus of our empirical analysis. Furthermore, in our model, the reason voice is more likely to emerge in concentrated markets is because firms are more likely to respond if they risk losing a valuable consumer. This suggests several other relationships that we can explore empirically. First, using data on airline responses to tweets, we can investigate whether airlines disproportionately respond to tweets from customers who are more valuable. Second, since our model predicts that the goal of voice is to elicit a response or concession from the firm, we will explore how quality deterioration impacts tweets to an airline relative to tweets that are simply about the airline. Third, since our model suggests that voice and a concession serve to maintain a future relationship between the customer and firm, we will investigate whether customers who receive a response to their tweet are more likely to tweet again.

3 Twitter as a Mechanism for Voice

Twitter provides a technology for observing and measuring voice. We are not the first to make the connection between tweets and voice. For example, Ma, Sun, and Kekre (2015) examine the reasons for voice by 700 Twitter users who tweet to a telecommunications company. They model optimal responses by the company and emphasize the service interventions improve the relationship with the customer. Bakshy et al (2011) show how ideas flow through Twitter. They emphasize that the idea of a small number of "influencers" does not hold in the data and that messages can be amplified through the network.

As a type of social media, Twitter also lowers the cost of exercising voice. It is lower cost than writing a letter to an airline or the FAA. Hirschman (p. 43) emphasizes that the use of voice will depend on "the invention of such institutions and mechanisms as can communicate complaints cheaply and effectively." Twitter and other social media also make voice, and the response to voice, visible to others. This should increase the effectiveness of voice and its expected payoff. In this paper, we do not emphasize how Twitter has changed voice. We treat Twitter as a platform for exercising and measuring voice and use the data to understand the interaction between voice and market power.

Many companies appear to have recognized that customers are "talking" about them on Twitter. They have invested considerable resources in managing social media in general and social media complaints in marketing. For example, Wells Fargo invested in a social media "command center" to manage and respond to complaints on Twitter (Delo 2014). In addition, there are companies that offer enterprises social media dashboards and management tools (such as Conversocial and Hootsuite). Indeed, many airlines have employees dedicated to responding to customers through social media.[7] Twitter itself has recognized that it plays this role and has published studies regarding their role in customer service (Huang, 2016) and their intention to make this a core product in their service (Cairns, 2016).

4 Empirical Setting and Data

4.1 Empirical Setting

Our empirical setting is the U.S. airline industry. While it is likely that Twitter has facilitated voice in many industries, we chose the airline industry as our setting because it has several features that make it particularly well suited for a study of the relationship between voice and market structure. First, a key measure of quality in this industry – on-time performance – is easily measured and data on flight-level on-time performance is readily available. This allows us to link the volume of voice to variation in an objective measure of vertical product quality. Importantly, on-time performance is determined at the flight level and therefore varies within markets not just across markets. Second, all the major U.S. airlines had established Twitter handles by 2012. Thus, it was technologically feasible for consumers to exercise voice to airlines via Twitter. Third, the airline industry is comprised of many distinct local markets. Each airport (or city) has its own market structure and configuration of airlines. This means that the opportunities for exit and the margins earned from consumers will vary across markets. Finally, since many consumers fly on a regular or even frequent basis, this setting is one in which the potential for future transactions to impact current behavior (i.e.: the scope for a relational contract) is quite real.

[7] See, for example, http://www.cnbc.com/2016/09/27/frustrated-flyers-listen-up-airlines-hear-your-rant-on-twitter.html and http://airrating.com/ (accessed by authors on October 30, 2106).

4.2 Data

Our analysis combines three types of data. The first is data on tweets made to or about one of the major seven U.S. airlines. We purchased this data from Gnip, a division of Twitter. We combine this with data on airline on-time performance, from the Department of Transportation (DOT), and with data on airline flight schedules, purchased from the Official Airlines Guide (OAG).

i. Twitter Data

The raw data purchased from Gnip contains all tweets made between August 1, 2012 12:00AM and August 1 2014 12:00 AM that include any of the following strings: "@alaskaair", "#alaskaair", "alaska airlines", "alaskaairlines", "@americanair", "#americanair", "americanairlines", "american airlines", "@delta", "#delta", "delta airlines", "deltaairlines", "@jetblue", "#jetblue", "jetblue", "jet blue", "@southwestair", "#southwestair", "southwestairlines", "southwest airlines", "@united", "#united", "unitedairlines", "united airlines", "@usairways", "#usairways", "us airways", "usairways". These strings include the Twitter handles of the seven largest U.S. airlines (Alaska Airlines, American Airlines, Delta Airlines, JetBlue, Southwest Airlines, United Airlines, and US Airways) as well as the names of these airlines, on their own and with a hashtag.[8] Together, these seven airlines accounted for over 80% of passenger enplanements at the start of our sample period.[9] The level of observation in this data is the "tweet". The raw tweet-level dataset contains 11,367,462 observations.

This data contains all initial communications from consumers to the airlines on Twitter. While the structure of Twitter now allows for private communication (or direct messages) between Twitter members who do not follow one another, during our sample period this was not possible. Specifically, if a consumer followed an airline but the airline did not follow a consumer, the consumer could not send a private message to the airline. By contrast, it is possible, and probable,

[8] A Twitter "handle" is the unique identifier, starting with the "@" symbol, for each participant on Twitter. While each tweet is public in the sense that anyone can see it, Twitter users let users know about a message by tagging them using their handle. A tweet that mentions an airline's handle is therefore directed at the airline and meant for the airline to see it. 58% of the tweets in our data mention the airline's handle. A Twitter "hashtag" is a way for Twitter users to highlight a phrase that other Twitter users may search for or find interesting, starting with the "#" symbol. A tweet that mentions an airline hashtag tells the users' followers that the airline is a key part of the tweet.

[9] This number is based on the enplanement data in the Air Travel Consumer Report for August 2012. It likely is an understatement as it does not include passengers travelling on these airlines' regional partners.

that some airline responses to consumers are done privately (even if via Twitter) and will not appear in our data.

Many tweets that met our initial filter criteria but were not about airlines. To identify these tweets, we looked at all hashtags and handles that started with the same characters as our tweets but did not end with these characters. The most common of these were mentions of arenas and stadiums named after airlines such as American Airlines Arena, mentions of the soccer team Manchester United, mentions of the United States or United Kingdom, and some hashtags such as @deltaforce. After eliminating the tweets that were clearly not about airlines, 5,900,691 tweets remained.

The Twitter data includes many variables including the date and time of the tweet, the content of the tweet, some information about the profile of the Twitter user (including where they are from and their number of followers) and, for a fraction of the tweets, the location from which the tweet was made. From the content of the tweet, it is possible to determine which tweets are "retweets", indicating that someone was passing on a tweet originally written by someone else. It is also possible to distinguish tweets *to* the airline from tweets *about* the airline based on whether the tweet includes the airline's Twitter handle. We are also able to determine which tweets were made by the airlines themselves. We focus on tweets to or about an airline and therefore exclude the 14,382 tweets in the data which were made by the airlines themselves. This yields 5,886,309 total tweets. 32% of these tweets were "retweets." We drop the retweets from our analysis and focus on the 4,003,326 unique tweets made by Twitter users to or about the major U.S. airlines. Finally, we exclude all observations from two specific time periods: (1) the days around Super Storm Sandy (Oct. 27 to Nov. 1 2012), when delays and cancellations were widespread but few people were likely to be tweeting about airlines; (2) April 13 to 15, 2014, when twitter use related to airlines was unusually high because of a fake bomb threat made on twitter against American Airlines and a US Airlines customer service tweet containing a pornographic image. This leaves 3,860,528 tweets to or about the seven U.S. airlines.

To collect data on airline responses to tweets, we created a program that called up each of the 3,860,528 tweets in our data on the twitter website (through the Application Program Interface). The program examined all responses to the tweet to see if any of the responses were from the airline's handle. If so, then we code the airline as having responded. By May 2016, US Airways had discontinued its twitter handle after its 2015 merger with American Airlines.

Therefore, because we collected the response data in 2016, we do not observe any responses to tweets by US Airways and we drop the US Airways data from the response analysis.[10]

ii. On-Time Performance Data

We combine the Twitter data with data on the on-time performance of each of the airlines. Since September 1987, all airlines that account for at least one percent of domestic U.S. passenger revenues have been required to submit information about the on-time performance of their domestic flights to the DOT. These data are collected at the flight level and include information on the scheduled and actual departure and arrival times of each flight, allowing for the calculation of the precise departure and arrival delay experienced on each flight.[11] The data also contains information on canceled and diverted flights.

We use these data to construct daily measures of an airline's on-time performance in a given market (as well as a measure of the airline's total number of flights from a market, to use as a control variable). There are multiple ways to measure on-time performance – for example, the number or share of the airline's flights that are delayed, the average delay in minutes, or the number or share of flights delayed more than a certain amount of time. Cancellations can either be included with delays or considered on their own. In general, different measures of on-time performance are highly correlated with each other.

As our main measure of on-time performance, we calculate the number of an airline's flights from a given airport on a given day that depart more than 15 minutes late or are canceled. For multi-airport cities, we calculate the number of an airline's flights from any of the airports in the city that depart more than 15 minutes late or are canceled. We use the 15-minute threshold because the DOT has adopted the convention of considering a flight to be "on-time" if it arrives within 15 minutes of its scheduled arrival time. We focus on departure delays but could use arrival

[10] We encountered one other issue in collecting the response data. Tweets from accounts that had been closed or were private would not appear on the twitter when we searched for responses. We coded these tweets as not having received a response though it is possible that they did. A random sample of 200 of our tweets found nine such closed and private accounts. This will result in some noise in our response variable.

[11] Airlines' regional partners report the on-time performance of the flights they operate on behalf of a major under their own code, not the major's code. Since customers likely associate these flights with the major given that they are flown under the major's brand, we include flights operate by a major's regionals partners in our measures of the major airlines' on-time performance. To do this, we use information from the Official Airlines Guide (OAG) data to match regional flights in the BTS data to their affiliated major airline.

delays instead as – within an airline-airport-day – departure and arrival days are highly correlated with each other. Our results are robust to alternative measures of on-time performance.

iii. Flight Schedule Data

We use data from the Official Airlines Guide (OAG) to construct measures of airline's size and share of operations in a given market. The OAG data provide detailed flight schedule information for each airline operating in the U.S. Each observation in this data is a particular flight and contains information on the flight number, airline, origin airport, arrival airport, departure time, and arrival time. Our sample of OAG data includes the complete flight schedule for each airline for a representative week for each month (specifically, the third week of each month).

From the OAG data, we calculate each airline's total number of domestic flights from each airport during the representative week as well as the total number of domestic flights from the airport by any of the seven airlines. We then use this to construct each airline's share of flights from the airport. This gives us a measure of each airline's dominance at an airport each month. For our analysis, we want a time-invariant measure of an airline's dominance at an airport. We calculate each airline's average share of flights at each airport over our two-year sample period and, from these shares, we construct four categories of airport dominance: less than 15% of the flights from the airport, between 15% and 30% of flights from the airport, between 30% and 50% of the flights from the airport, 50% or more of the flights from the airport.[12] We construct analogous measures of dominance at the city level for multi-airport cities.

An airline's share of flights from a given airport (or city) captures how easy or difficult it would be for a consumer to avoid (i.e.: exit from) that airline on subsequent flights. As discussed earlier, however, the ease of exit makes voice less necessary but more effective since backed by a credible threat of exit. As our model highlights, the likelihood that a firm responds to voice and, in turn, the incentive for consumers to exercise voice depends on the future value of the consumer to the firm. Airlines with a dominant position at an airport charge higher fares and are particularly attractive to high willingness-to-pay travelers because their large network means they offer the

[12] There are several different ways to capture an airline's dominance at an airport. Previous work (for example, Lederman 2007) has also used an airline's share of departing flights. Borenstein (1989) uses an airline's share of originating passengers at an airport but reports that his results are robust to using an airlines' share of departing flights, departing seats, or departing seat miles. Some studies simply identify the airports that an airline uses as its hubs. These different measures are typically highly correlated with each other.

most attractive frequent-flier program to consumers in that market (see Borenstein (1989) and Lederman (2008)). As a result, the costs of losing a customer may be greater for dominant airlines.

4.3 Construction of the Estimation Samples

The central goal of our analysis is to explore the relationship between quality (measured by on-time performance) and voice (measured by the volume of tweets) and investigate how this relationship varies with market structure. Thus, our empirical strategy requires us to link tweets to the on-time performance of the tweeted-about airline and the market structure faced by the individual who made the tweet. While we are not able to match individual tweets to particular flights, we can match tweets to airports (or cities) and, in turn, to an airline's on-time performance in that airport (or city) on the day the tweet was made. Since market structure varies at the airport (or city) level, once we have matched tweets to airports, we can also integrate information on the market structure at the airport (or city).

We use three different methods for matching tweets to airports. First, many Twitter users identify a location in their Twitter profile. This location does not change from tweet to tweet and can be interpreted as "home", as identified by the Twitter user. Because we are focusing on how the relationship between quality deterioration and voice varies with market structure, we use the location given in the profile of the Twitter user as our primary measure of the tweeter's home market. Many Twitter users in our data leave this location blank, identify an international location, a non-specific location (such as "united states", "california"), or identify a humorous location (such as "Hogwarts" or "in a cookie jar"). We, of course, cannot identify a location in profile for these tweets. However, for 36% of the tweets in our data, the location is specific enough that we can match it to a U.S. city with a major airport. In our tables, we describe this source of location information as "Location given in profile". For cities with multiple airports, we create a code to capture the city rather than a specific airport. For example, we use the code "NYC" for a tweet from a profile that identifies New York City as home. Because of the multi-airport cities, when we use this location measure, we construct our airline on-time measures and market structure measures at the city – rather than airport – level.

Second, for some of the tweets in the data (approximately 7%), the Twitter user chose to use a feature of Twitter that identifies, through GPS, the location from which the tweet was posted. Specifically, the data indicates the latitude and longitude coordinates of the location from which the tweet was made. We combine this with data on the latitude and longitude of each U.S. airport

and identify the nearest airport. We refer to tweets with this location information as "geocode stamp on tweet".

The third way that we link tweets to airports is by exploiting information in the content of the tweet. Some tweets contain the code of a specific airport. For each tweet in the data, we determine whether the tweet contains the airport codes of any of the 193 largest airports in the U.S. We do this by determining whether the tweet includes the airport code in capital letters with a space on either side. For example, we code a tweet with "ORD" as having Chicago's O'Hare airport in the tweet. 4% of tweets have an airport mentioned in the tweet under this definition. We refer to these tweets as the "Airport mentioned in tweet" observations.

Overall, we have airport-level information for 427,536 tweets (based on the latter two measures of location) and city-level information for 1,394,070 tweets (based on all three measures of location).[13] As a check on the reliability of the different location measures, we examine the 195,945 tweets for which we have both city information (from the user's profile) and airport information (from either a geocode stamp or an airport mentioned in the tweet) information. For these 195,945 tweets, the city and airport locations match 47.0% of the time. As a benchmark, if the measures perfectly captured the correct city and airport, we might expect them to match slightly less than 50% of the time because of return trips and stopovers. We view this as suggesting validity to both the airport and city measures.

Having matched tweets to cities and/or airports, we are able to construct the airline-airport-day and airline-city-day datasets that we use for our regression analysis. We restrict the sample to airports/cities with at least 140 flights per week in the OAG data (i.e.: at least 20 flights per day). This produces 100 airports in the airline-airport-day sample and 82 cities in the airline-city-day sample. For each airline operating at each airport on each day (or in each city each day), we combine measures of the airline's on-time performance at the airport (or in the city) on the day with the total number of tweets to or about the airline that day from individuals associated with the airport (or city). Finally, we merge in the measures of the airline's dominance at the airport (or in the city). Our final airline-airport-day dataset contains 382,141 observations while the final airline-city-day dataset contains 318,077 observations.

[13] We exclude 63,090 tweets (4.4% of the tweets with city information) that mention more than one airline because we are not able to associate these tweets with one particular airline.

4.4 Descriptive Statistics

Table 1 provides descriptive statistics at the tweet-level. Panel A shows the share of tweets for which we have different types of location information. Panel B compares the distribution of tweets across airlines for the three sets of observations we use (all tweets, tweets with geocodes, and tweets with any location information). American Airlines is the most common airline mentioned in tweets, with 26% of all tweets relating to American Airlines. Alaska Airlines is the least common, with less than 3% of all tweets. As the table suggests, the composition of the three samples, in terms of the fraction of tweets to or about each airline, is very similar.

Figure 1a shows the average number of daily tweets by month over time for the subsample of our data with city information.[14] The figure shows that the average number of tweets about airlines increases from around 1,500 per day at the beginning of the sample to over 2,500 per day toward the end of the sample. Figure 1b shows that all airlines experienced an increase in tweet volume over time.

Table 2 contains descriptive statistics for the airline-city-day (in the top panel) and airline-airport-day datasets (in the bottom panel). Because cities with multiple airports are aggregated across airports, the city-airline-day data has fewer observations. Also, both because of aggregation and because we have many more tweets with city-level information than airport-level information, the number of tweets per day is much higher at the city level (on average, 4.26 tweets per airline-city-day compared to 0.59 tweets per airline-airport-day). In addition to the number of tweets, the table presents summary statistics for the on-time performance and airline dominance measures. The table indicates that, for 48% of airline-city combinations, the airline operates less than 15% of flights from the city. For about 35% of the combinations, the airline operates between 15% and 30% of flights at the city, for about 12%, the airline operates 30%-50% of the flights from the city, and for about 5% of observations, the airline operates more than 50% of the domestic flights from the city. The numbers for the airline-airport level dataset are similar though not identical.[15] In both

[14] We focus on this subset of our data because we use it for most of the analysis that follows. The patterns look similar when we use all tweets, but the numbers are larger as Figure 1 uses only 36% of all tweets.

[15] In both datasets, the observations in which an airline operates more than 50% of domestic flights are primarily airlines at large hubs (for example, Delta Air Lines in Atlanta, United Airlines in Cleveland, American Airlines in Dallas-Fort Worth, and Southwest Airlines in Las Vegas). There is a larger number of observations in which an airline operates between 30% and 50% of domestic flights. These include both airlines at their own (less dominated) hubs

datasets, about 20% of an airline's flights at an airport or in a city are delayed more than 15 minutes or canceled on a given day.[16]

For the majority of our empirical analysis, we define an airline's level of dominance using the city-level measures, even when we match tweets at the airport level. We do this because there is likely substitution across the different airports in a given city and therefore we want our measure of a consumer's ability to exit from an airline to include alternatives at other airports. Brueckner, Lee, and Singer (2014), for example, argue and provide evidence that city-pairs rather than airport-pairs should be the relevant unit of analysis in studies of airline markets.

We also construct a number of variables to capture the content and sentiment of the tweets received. From these tweet-level characteristics, we construct airline-city-day level counts of the number of tweets with these characteristics. These variables serve as more nuanced and detailed measures of voice. First, we construct a variable ("# of tweets to handle") that measures the number of tweets to the airline's handle. Tweets to the airline's handle are directed through Twitter to the airline whereas tweets about the airline are not. On average, an airline receives 2.96 tweets to its handle, on a given day from consumers associated with a given city. Second, we measure the number of tweets that mention on-time performance, which has a mean of 0.77.[17] Third, we construct a variable that captures whether the content of the tweet is positive or negative. This measure of "sentiment" is a standard measure from computer science and provides a probability that a particular tweet is negative. The idea of the algorithm is to look for the symbols ":)" for positive sentiment and ":(" for negative sentiment.[18] The algorithm then identifies the probability

---

and, mostly, airlines at smaller airports where they have a significant share of flights but the airport is not a hub to them or to any carrier.

[16] For a subset of the flights, we have a measure (reported by the airline) of whether the airline is at fault in the delay. The average number at fault is close to the average number delayed because we disproportionately observe larger airports for this data.

[17] We define a tweet being about on-time performance if it contains one of seven strings related to on-time performance: "wait", "delay", "cancel", "time", "late", "miss", or "tarmac". We define a tweet being about frequent flier programs if it contains one of the following strings: "aadvantage", "mileage" (includes "mileageplus"), "miles" (includes "dividend miles"), "trueblue", "skymile", "lounge", "rewards" (includes "rapidrewards"), "admiral", "club" (includes "united club"), "gold", "diamond", "silver", "elite", "frequent", "status", "premier", "100k", "50k", or "25k". While these words may appear in our contexts, in our sample of airline tweets they almost always refer to frequent flier programs.

[18] Read (2005) developed the idea of using emoticons to measure sentiment. It appears in reviews on sentiment analysis such as Pang and Lee (2008) and has been shown to be particularly useful for Twitter data (e.g. Agarwal et al 2011, Pak and Paroubek 2010). The algorithm we use builds on code from a June 16, 2010 post at

the :) or :(symbol appears, given the appearance of the various word pairs ("bi-grams"). For example, the word pair ("again", "cancel") appears disproportionately often with ":(" and the word pair ("great", "service") appears more often with ":)". Then, for the full tweet-level data set, we predict the probability that a particular tweet has negative sentiment based on the word pairs contained in the tweet. Table 3 provides sample tweets for different levels of sentiment.

It is difficult to algorithmically assess sentiment with the 140 characters in a tweet, and so this measure is noisy, with little obvious difference between a tweet given a score of 0.4 and a tweet given a core of 0.6. Furthermore, the average score variable is missing for airline-location-days without tweets. The algorithm does a better job with tweets that score very positive (below 0.1) or very negative (above 0.9).[19] Therefore, we identify very positive and very negative tweets, in addition to the average score. On average, across airline-city-days, airlines receive 1.90 very positive tweets and 0.98 very negative tweets.

5 Empirical Approach

We proceed with our analysis in four stages. After some motivating descriptive analysis, we first investigate the relationship between the volume of tweets received and on-time performance to determine whether, in this setting, consumers use voice to respond to quality deterioration. Second, we examine whether market dominance increases or decreases the strength of this relationship, the core empirical question underlying Hirschman's *Exit, Voice, and Loyalty*. Third, we carry out some analyses that exploit the content and sentiment of tweets to provide evidence that our main results are consistent with voice being a response to quality deterioration. Fourth, we carry out a number of supplementary analyses that specifically explore aspects of the relational contracting model that we propose.

In most of the analysis that follows, our empirical approach focuses on the relationship between tweets and on-time performance. We view this correlation as measuring a response

http://streamhacker.com/2010/06/16/text-classification-sentiment-analysis-eliminate-low-information-features/ (accessed May 14, 2015). The code is modified to remove user names and add "stemming" of words (so that "cancel", "cancels", and "canceled" are all coded as the same word). For a training data set, we combine all the tweets in our data with happy or sad emoticons with the tweet training data set available at http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip.

[19] The algorithm, however, does not do a very good job of recognizing sarcasm in tweets, as exemplified by the first tweet with probability negative of 0.10 in Table 3. As a result, sarcastic tweets, intended to be negative, are sometimes mistakenly classified as positive.

elasticity to service failures. Fundamental to Hirschman's framework, and to our formalization, is that voice is a response to quality deterioration. A key advantage to our setting is that airline delays and cancellations provide a measure of quality deterioration that changes frequently, even for a given airline in a given market. This enables us to measure how the elasticity of voice to quality deterioration changes with market structure while controlling for the average relationship between market structure and quality. While our model does not explicitly distinguish between quality deterioration that the consumer believes is or is not the airline's fault, we present some analysis in the results section that attempts to separate these two types of reductions in quality.

Our main empirical specification regresses the number of tweets about an airline on a given day by consumers associated with a given location on the on-time performance of that airline at that location on that day. To analyze whether and how this relationship varies with an airline's dominance of a market, we interact an airline's on-time performance with a measure of its dominance of the airport or city. Our models control for airline-location, which will control for factors that influence the average amount of voice that an airline receives from consumers in a particular market. Importantly, these controls will capture the overall scale of an airline's operations at an airport or in a city, which likely impacts the amount of voice received as larger operations imply more passengers carried. These controls will also capture any impact that an airline's level of dominance in a market has on the amount of voice it receives. Note that an airline's scale of operations and level of dominance are not necessarily related. Airlines will have both many flights and a large share of flights at their hub airports. However, airlines may also dominate small airports at which they do not operate very many flight, in absolute terms. In addition, at large airports that are not a hub to any carrier (such as Boston's Logan Airport and New York's La Guardia Airport), several airlines operate a significant number of flights but none dominates the airport. Our specifications also include location-day fixed effects, which capture both location-level causes of delay (such as weather) and the diffusion of Twitter during our sample period in a very flexible way, allowing the diffusion rate to differ across locations.

One challenge we encounter is setting up our empirical analysis is choosing the appropriate functional form for our dependent variable as well as our measure of on-time performance. Both the number of tweets an airline receives on a given day in a given market as well as the number of its flights that are delayed or canceled have a large mass at zero and a very long right tail. In particular, at locations in which airlines have a larger scale of operations, they can receive more

tweets and have a greater number of delayed or canceled flights. As a result, for these variables, both the mean and standard deviations vary substantially across airline-locations. For example, in our data, we observe Delta, at its hub in Atlanta, have an average of 157.6 delayed or canceled flights per day, with a standard deviation of 113.8. On the other hand, US Airways in Atlanta has 1.9 flights delayed or canceled, on average, per day, with a standard deviation of 2.1. For US Airlines, at its hub in Charlotte, the mean and standard deviation are 49.5 and 33.6, while Delta's values are 3.5 and 3.4. A similar comparison holds for the number of tweets received per day.

To create a measure of on-time performance deterioration that is comparable across airline-locations, we standardize the number of flights delayed more than 15 minutes or canceled variable by subtracting the airline-location mean and dividing by its standard deviation and we use this as our main measure of on-time performance.[20] Because the mean and variance of the number of tweets variables are similarly impacted by an airline's scale of operations at an airport, we standardize them in a similar fashion. For robustness, in the online appendix, we carry out all of the analyses using the logarithm of each of these variables (plus one).

Using this standardized data, our core regression specification for airline *a* in location *l* on day *t* is:

$$StdTweets_{alt} = \alpha StdDelays_{alt} + \beta StdDelays_{alt} \times AirlineDominance_{al} + \gamma StdFlights_{alt} + \mu_{lt} + \varepsilon_{alt}$$

Because of the standardization, airline-location fixed effects are not appropriate. Instead, they are, in effect, already differenced out. Because of this, the main effect of $AirlineDominance_{al}$ is not included. In the robustness analysis that uses a non-standardized logged specification, the airline-location fixed effects are included. Standard errors are clustered at the location level.

6 Results

a. Motivating Analysis

Before turning to the regression analysis, in Table 4 we illustrate the variation in our data that we exploit in our regression analysis. Using the location provided in a consumer's twitter profile as the location definition, each cell shows the correlation coefficient between poor on-time performance and the average number of tweets by airline-location-day, both normalized by

[20] This approach has been used in other settings to adjust outcome measures that have different means and variances. See, for example, Chetty, Friedman and Rockoff (2014) and Bloom, Liang, Roberts and Ying (2014).

location-airline mean and standard deviation, using the method described above. The table shows a positive correlation between delays and tweets, which gets larger as an airline's market dominance increases.

b. Tweets and On-Time Performance

Table 5 estimates the relationship between tweets and on-time performance without interactions with market structure. The first row contains the coefficient of interest: the (normalized) number of the airline's flights in a location delayed at least 15 minutes or canceled. If, as hypothesized, tweets are a response to quality deterioration, we would expect the coefficient to be positive. In most of our analysis, our main dependent variable is the normalized number of tweets to or about an airline on a day by individuals associated with a given city, based on the location information in the individual's Twitter profile. We focus on this measure because it captures the Twitter users' home city and is therefore most likely to capture the market structure they typically face. In Tables 5 and 6, we also show robustness to the alternative ways of matching tweets to locations.

Table 5 shows a robust statistical relationship between on-time performance and tweet volume. Across four different specifications, the point estimate is always positive, statistically significant, and large in magnitude. Column 1 includes controls for the number of flights that the airline has at that airport, and location-city fixed effects. As expected, having more flights from a location increases the number of tweets received from consumers in that city. This serves as our main empirical specification for the remainder of the paper. Note that the variable capturing the (standardized) number of flights the airline operates is only identified off of differences in the scale of an airline's operations across days and the coefficient on this variable is, not surprisingly, insignificant and small in magnitude.

The coefficient estimate in column 1 suggests that an increase in the share of delayed or canceled flights of one standard deviation is associated with 0.078 standard deviations more tweets. Column 2 shows robustness to associating tweets to locations using any of the three sources of location information. Column 3 changes the dependent variable to log(tweets with location given in profile+1), demonstrating that the sign of the correlation is robust though the coefficient should not be interpreted as an elasticity. The R-squared here is much larger than in the other columns, suggesting that the standardization differences out much of the explainable variation. In Column 4, tweets are matched to the airport (rather than the city) closest to the user at the time the

tweet was made and then aggregated to the airline-airport-day level. The airport-level analysis also shows a positive and statistically significant relationship between delays and tweet volume. Overall, we view Table 5 as clearly revealing that there is a robust statistical relationship between tweets and quality deterioration, which emerges across various location measures, fixed effect specifications, and functional forms.

c. Tweets, On-Time Performance, and Market Structure

To assess how market dominance affects the relationship between tweets and on-time performance, we add interactions between our measures of on-time performance and an airline's level of dominance in a city or at an airport. The first column in Table 6 re-estimates column 1 of Table 5 with the added interactions as specified in the regression equation above. The first row shows the main effect of delays and cancellations, which captures the relationship between tweets and on-time performance when an airline operates less than 15% of the flights in a market. The following rows show the interactions with the three higher categories of airport dominance.

Column 1 shows that the relationship between on-time performance deterioration and tweets is stronger when an airline has a dominant position at an airport. In particular, a one standard deviation deterioration in on-time performance generates about 85% more voice when an airline operates between 30% and 50% of flights in the market and more than double the amount of voice when an airline has more than 50% of the flights in the market. When an airline operates between 15% and 30% of flights in a city, the impact of a deterioration in on-time performance is only marginally statistically (and economically) different from the impact when an airline has less than 15% of flights. Therefore, for the remainder of specifications, we combine the two lower categories and use that as the excluded category. We show this in column 2. In columns 3 to 5, we show that the pattern of interaction effects is robust to using any of the three sources of location information, to using log(tweets with location given in profile+1) as the dependent variable, and to using the airport (rather than the city) closest to the user at the time the tweet was made.

Across all specifications, the coefficients on the interactions between quality and airline dominance (measured by 30-50% share of flights or over 50% of flights from the city) are positive and statistically significant. Furthermore, the coefficient when airlines have over 50% of flights is larger than the coefficient when airlines have 30-50% of flights. Thus, our results indicate that - when airlines are dominant in a market - the relationship between on-time performance and tweets is stronger. Interpreted through the lens of *Exit, Voice, and Loyalty,* and as predicted by our

relational contracting model, we find that voice is more likely to emerge as a response to quality deterioration when an airline is the dominant firm in a market

d. Evidence that the Results are driven by Comments about Quality Deterioration

In this section, we include additional analyses that investigate whether the increase in voice that we measure is likely to be a response to an unexpected deterioration in quality.[21] In particular, we show that tweets specifically about on-time performance rise when on-time performance deteriorates and that tweets become more negative in sentiment when on-time performance deteriorates. We also show delays that are the airline's fault generate a larger increase in tweets (in general and specifically for dominant airlines) than delays that are not the airline's fault. Together, we view these results as suggesting that the increase in tweets that we are capturing is indeed a response to unexpected quality deterioration and not the result of some other factor (such as, a mechanical increase in tweeting because people have time to use Twitter while waiting at the airport or simple complaining about factors outside the airline's control, such as adverse weather).

Table 7 re-estimates the main specification from Tables 5 and 6 using two alternative dependent variables: the number of tweets that mention on-time performance and the number of tweets that do not. The results in the first two columns show that, when delays and cancellations increase, tweets that mention on-time performance increase twice as much as tweets that do not mention on-timer performance. Columns 3 and 4 show that, as dominance grows, the increase in the number of tweets about on-time performance is larger than the increase in the number of tweets not about on-time performance.

Table 8 explores tweet sentiment. Recall that for each tweet, the algorithm predicts the likelihood that the sentiment of the tweet is negative. The dependent variable in columns 1 and 2 is the average predicted sentiment of the tweets received by an airline in market on a day. The value is missing when there are no tweets on a day. These columns investigate whether on-time performance impacts the average sentiment of tweets received. We find that the average negative sentiment of the tweets received is higher when delays and cancellations increase and that the same deterioration in on-time performance generates more negative sentiment when an airline is

[21] From this point on, we only present standardized results at the city level. However, in the online appendix, we present all of these specifications estimated with non-standardized logged variables and estimated with standardized variables at the airport level.

dominant. In columns 3 and 4, we explore whether a deterioration in on-time performance impacts the number of very negative or very positive tweets received. We find that both very negative and very positive tweets increase when on-time performance is worse, but the impact on very negative tweets is much larger.[22] Columns 5 and 6 include the interactions with market share and, again, show that the increase in very negative tweets is much larger than the increase in very positive tweets and that the impact of market dominance on the relationship between on-time performance and tweets is larger for very negative tweets.

A feature of our setting is that quality may deteriorate for reasons outside the airline's control, such as bad weather. Consumers know that this is possible when they purchase their tickets and therefore may not voice in response to this type of quality deterioration. If this were the case, we would expect our results to be strongest for deteriorations in quality that are – or are perceived to be – within the airline's control. We investigate this in Table 9, first by explicitly including variables measuring daily weather and then by distinguishing between delays that are and are not the airline's fault. Before turning to these results, it is worth pointing out that all of our specifications include city-day (or airport-day) fixed effects. Thus, we are already controlling for the weather in a city (or at an airport) on a day and cannot directly include measures of the weather experienced on that day. Moreover, this implies that the coefficients on the delay variables in our regressions are only identified off differences in on-time performance across airlines at an airport on a day, after accounting for the average impact of that day's weather on delays and cancellations. However, because it is possible that adverse weather may impact dominant airlines differently than non-dominant airlines (and this could, in turn, confound the interaction terms in our regressions), we estimate specifications where we interact weather variables with the dominance variables.[23] The results are presented in columns 1 and 2 of Table 9. The first column includes a single weather

[22] The finding that very positive tweets increase when on-time performance deteriorates may seem surprising but can be explaining by two factors. First, a deterioration in on-time performance gives airlines an opportunity to remedy problems and a successful remedy can lead to a very positive tweet. Second, as mentioned above, the algorithm often misclassifies sarcastic tweets, which are intended to be negative but sound positive. These types of tweets are likely to increase when on-time performance gets worse.

[23] The weather data are from the National Oceanic and Atmospheric Administration (NOAA) Quality Controlled Local Climatological Data. These data provide daily information on a large number of weather variables captured by weather stations. Stations exist at every airport. We collected the data for every airport in our dataset. For our city-level analysis, when a city had multiple airports, we randomly chose one of the airports in the city and used that airport's readings for all airports in the city. The weather data can be found at https://www.ncdc.noaa.gov/qclcd/QCLCD?prior=N (last accessed December 20, 2016).

variable that equals one if there was rain, snow, or fog in the departure city on a given day. The second column instead uses a continuous measure of the total amount of precipitation (rain and snow) in the city on the day. Both specifications show that inclusion of the additional interaction terms has little impact on the coefficients on the on-time performance variables and their interactions. The coefficients on the weather interactions are statistically not significant.

In the third column of Table 9, we directly investigate whether the response to quality deterioration via voice is greater when the quality deterioration is likely to be the airline's fault. To do this, we take advantage of the fact that the DOT on-time performance data also contain information on the (self-reported) causes of delay for each flight that is more than 15 minutes late.[24] We then construct a variable measuring the number of flights delayed more than 15 minutes that are the airline's fault and the number delayed more than 15 minutes that are not the airline's fault. We include both of these and their interactions with the dominance variables, in the regression. The results appear are shown in column 3 of Table 9. While correlation between fault and not-at-fault delays reduces power, we find that the interaction effects are larger in magnitude for delays that are the airline's fault.

Overall, we view the results in this section as indicating that the relationships we have uncovered are indeed evidence of consumers using voice when they experience unexpectedly poor quality.

e. Support for the Relational Contracting Model

In this section, we carry out a number of analyses that investigate specific predictions of the relational contracting conceptualization of voice proposed above.

i. The model emphasizes the value of customers

The model emphasizes that firms have a larger incentive to respond to voice exercised by more valuable (or profitable) customers. We therefore examine whether the airlines are more likely to respond to tweets from customers that are more valuable. We capture the expected profitability

[24] This information is part of the on-time performance data collected by the DOT. For each delayed flight, airlines are required to indicate the cause(s) of the delay: air carrier delay, weather delay, National Aviation System delay, security delay and late arriving aircraft delay. Airlines can attribute delay the multiple categories, indicating the number of minutes by each cause. We consider delays categorized as carrier delays or late aircraft delays to be the airline's fault. Because this variable is self-reported by airlines, it is reasonable to see it as a lower bound on the fraction of delays that are the fault of the airline.

of consumers in two ways. First, as in the analysis above, by whether the customer lives in a city where the airline has a large share of flights. Second, by whether the tweet mentions that the customer is in a frequent flier program. Customers who are entrenched in an airline's frequent flier program (FFP) are more valuable for several reasons. First, they are more likely to be business travelers. Business travelers have a higher willingness-to-pay, which airlines exploit through price discrimination. Second, if they are already invested in the airline's FFP, the marginal value of additional frequent flier points will be higher for them (due to the non-linearity of most FFP reward structures). This, in turn, will further raise their willingness-to-pay (Lederman 2007, 2008). Third, they are more likely to fly frequently which increases the value of preserving a long-term relationship with them.

As mentioned above, we collected data on whether each tweet received a response for the airline (excluding US Airways). Overall, 21.4% of tweets receive responses. Of tweets that mention the airline's handle, 34.7% receive responses. Figure 2a shows that the fraction of tweets that receive responses grew rapidly until June 2013, and then leveled off. Figure 2b shows that there is considerable variation in response rates by airline, with American being most responsive during this period and Southwest being the least responsive.

Before proceeding with the airline response analysis, it is important to recognize that there are other ways airlines could respond to tweets, including email, direct messages, and future quality improvements. We are unable to observe these, yet they would be consistent with the "concession" we describe in our model. Nevertheless, we view the relatively high response rate as consistent with our theoretical framework and indicative that Twitter is an important channel of communication with customers in this industry.

In Table 10, we estimate whether airlines are more likely to respond to tweets from customers who are more valuable. For this analysis, the level of observation is the tweet and the dependent variable is an indicator variable for whether the tweet received a response from the airline. We estimate a logit model. We control for other factors that might elicit an airline response including whether the tweet contains the airline's handle, whether the tweet contains a customer service keyword,[25] and whether the tweet contains an on-time performance keyword. We also

[25] We define customer service strings as "food", "water", "desk", "agent", "attendant", "attendent", "counter", "queue", "manning", "crew", "rude", "nasty", "service", "staff", "awful", "drink", "svc", and "handling".

control for the airline, the tweeter's number of followers, the tweet sentiment, and a linear time trend.

The estimate in the first row of Column 1 shows that airlines are more likely to respond to tweets from customers associated with markets in which the airline operates more than a 30% of flights. The remaining rows show the impact of the other variables: airlines respond more often to tweets with a negative sentiment, to tweets to their handle, to tweets with customer service keywords, and to tweets with on-time performance keywords. We see no consistent correlation between the number of followers and response rates, a result we revisit below.[26]

Column 2 switches the definition of most valuable customers from location to whether the tweet contains a word that suggests that the tweet comes from a frequent flier. In many ways, we believe that this is a better measure because airline social media managers will have easy access to the tweet content while the location information may be harder to find. The result suggests that airlines respond more to tweets with frequent flier keywords. Column 3 includes both frequent flier keyword and location information and shows that the positive coefficients are robust. Overall, we interpret Table 10 as suggesting that airlines are more likely to respond to tweets from their more profitable customers.

ii. The model emphasizes direct communication

In our relational contracting model, customers use voice to complain directly to the firm, rather than to "vent" or punish the firm by telling others about their bad experiences. Of course, one difference between Twitter and other channels for voice is its public nature. This raises the possibility that venting or inflicting demand losses on the airline in other markets may be part of the reason people tweet in response to delays and cancellations. Here, we provide evidence that suggests venting is unlikely to be the primary motivation for the relationships we observe.

If a tweet is directed to an airline's handle, it suggests that the customer wants the airline to see that tweet (rather than simply complain about the airline to friends and followers). In particular, tweets to a handle will show up in the airline's notification center automatically. Thus, a tweet to an airline's handle is a (public) message directed to the airline rather than a public

[26] The relationship between number of followers and responses is non-linear. To communicate the non-linearity, we split the data into 0-25th percentile, 25th to 50th percentile, 50th to 75th percentile, 75th to 99th percentile, and (to account for the few twitter users with a very large number of followers) over 99th percentile.

message *about* the airline directed to the sender's Twitter followers. Table 11 compares the impact of on-time performance deterioration on tweets made to an airline's handle and tweets not directed at the handle. Columns 1 and 2 show that when delays and cancellation increase both tweets to the handle and tweets not to the handle rise, though tweets to the handle increase slightly more. Thus, while there seems to be some public complaining in response quality deterioration, much of the additional voice is directed at the airlines. Furthermore, columns 3 and 4 show that dominance has a larger impact on the responsiveness of tweets to the handle to poor on-time performance than tweets not to the handle. In addition, returning to Table 10, which estimated the airline response models, we find there that airlines are more likely to respond to tweets to their handle than tweets that simply mention them. Overall, we see this collection of results as suggesting that the customers use Twitter to communicate with the airline rather than simply complain publicly about the airline.

Table 11, however, does not address the fact that even a tweet to an airline's handle is public and that the public nature of the tweet might nevertheless be driving the consumer's decision to exercise voice. We explore this in two ways, each using the number of followers as a signal of the important of the public nature of the tweet. First, we replace number of tweets with average number of followers as the dependent variable. Table 12 displays the results. Column 1 shows that the average number of followers for people who tweet on days with delays and cancellations is very slightly higher—a one standard deviation increase in delays is correlated with a 0.005 standard deviation increase in average number of followers. Furthermore, and perhaps more importantly, column 2 shows that this relationship is unrelated to market dominance. Thus, the number of followers does not appear to be substantially different for tweets that are about delays or cancellations in places where an airline is dominant.

Second, returning to Table 10, which estimated the airline response models, we find no consistent relationship between a tweeter's number of followers and the likelihood of receiving an airline response though we do find that airlines are much more likely to respond to tweeters in the 99^{th} percentile of the follower's distribution. We interpret Table 11, Table 12, and the followers results in Table 10 as together suggesting that tweets about airlines during periods of poor performance are often communications to the airline.

iii. The model implies responses should lead to future tweets

Finally, in Table 13, we look at whether twitter users who receive a response from an airline are more likely to tweet again to the same airline in the future. Many of the twitter users in our data tweet multiple times to an airline. The 3,860,528 tweets in the data are made by 1,457,945 different users, Of these, 520,807 tweet more than once. The median number of tweets is 1, the 75^{th} percentile is 2, the 99^{th} percentile is 26 and the maximum is 6635. Excluding those whose first tweet was to US Airways, we can analysis 1,375,416 different users.

Table 13 explores whether users are more likely to tweet again to an airline if their first tweet received a response. In this way, the results explore whether responses (suggesting a successful use of the relational contract) lead to repeated use of the relational contract. Columns (1) and (2) look at the first tweet by each user. The dependent variable is whether the user tweeted again to the same airline during our sample period. The main covariate is whether an airline responded to the first tweet. Column (1) shows a logit regression of tweeting again on responses without additional controls. There is a positive correlation between airlines responding to an individual's first tweet and that individual tweeting again in the following years. Column (2) adds controls for sentiment, number of followers, whether the tweet was to the handle, customer service keywords in the tweet, on time performance keywords in the tweet, whether the original tweet contained a frequent flier keyword, the share of flights for the airline in the location of the tweeter, airline fixed effects, and a linear time trend. The coefficient on airline response is still positive. The controls generally suggest, unsurprisingly, that more active and experienced twitter users are more likely to tweet again.

Two potential concerns with this analysis are that the later tweets are part of the same conversation as the initial tweet and that tweeters who show up early in the sample have more opportunities to tweet again. Therefore, columns (3) and (4) look only at users whose first tweet in our data was in 2012. The dependent variable is whether we observe another tweet to an airline by these users in the later part of the data set, in 2013 or 2014. Again, the results show that users who received a response are more likely to tweet again.

Overall, we view our collection of results as consistent with a relational contracting model of voice. While the evidence here does not reject the possibility that other motivations for voice may also operate, it suggests that voice elicits an airline response when it comes from the highest value customers, rather than by the customers that have the greatest ability to damage the airline's

reputation by communicating a complaint to a large number of followers. Furthermore, when the airline responds (as expected in the relational contracting model), the twitter users are more likely to tweet again to an airline.

7 Conclusion

Based on the original ideas in Hirschman's *Exit, Voice and Loyalty*, we have developed a formal model of voice as the equilibrium of relational contract between a firm and its customer. Our model resolves a key ambiguity in Hirschman's formulation – namely, how market structure influences the choice between exit and voice. Our model predicts that voice is more likely to emerge in concentrated markets because the value to firms of retaining consumers is higher. Empirically, we have developed a strategy for estimating the relationship between quality deterioration, voice and market structure. Our analysis uses Twitter data, which provides us with a systematic way of measuring voice. Our empirical strategy takes advantage of the fact that, in the airline industry, a key dimension of quality – on-time performance – varies at very high frequency and therefore we can exploit daily variation in the quality an airline provides in a given market. This allows us to control for the underlying relationship between market structure and quality while tracing out the relationship between market structure and voice.

Our empirical results show that consumers do indeed use voice to express disappointment when quality deteriorates. We believe that this is the first large-scale study to document this fact. With respect to the relationship with market structure, our results indicate that consumer are more likely to use voice when the quality deterioration is by a firm that is dominant in the consumer's home market. These relationships are more pronounced for tweets that mention on-time performance and tweets that are negative in sentiment. The relationships are also more pronounced when delays are the airline's own fault. Consistent with a relational contracting model, we find that firms are more likely to respond to their most valuable customers and that users whose tweets receive a response from the airline are more likely to tweet to that airline again in the future.

New communication technologies such as social media are both lowering the costs of voice and making voice observable to researchers. These new technologies may increase the use of voice in markets and generate renewed research interest on the topic of voice. We view this paper as a first step in establishing such a research agenda. By exploiting the public nature of tweets and taking advantage of detailed data on quality that is available in the airline industry, we have

investigated the relationship between quality, voice, and market structure. Going forward, there are several interesting follow-up research questions worthy of future investigation. One such question is how the advent of new communication technologies affects the volume and nature of voice. This requires an empirical setting in which one could study complaints before-and-after the introduction of a new communication channel for voice. A second promising line of research could study the choice between voice and exit, at the customer level. Our setting does not allow us to observe exit directly, rather we infer exit options based on market structure. However, other settings, might allow researchers to observe this choice at the individual level. Finally, a fundamental question in this literature is whether lowering the costs of voice impacts equilibrium quality and welfare.

8 References

Abrahams, Alan, S., Jian Jiao, G. Alan Wang, and Weiguo Fan. 2012. Vehicle Defect Discovery from Social Media. *Decision Support Systems.* 54, 87-97.

Adelman, Jeremy. 2013. *Worldly Philosopher: The Odyssey of Albert O. Hirschman*. Princeton University Press, Princeton NJ.

Agarwal, Apoorv and Xie, Boyi and Vovsha, Ilia and Rambow, Owen and Passonneau, Rebecca. 2011. Sentiment Analysis of Twitter Data. *Proceedings of the Workshop on Languages in Social Media*. 30-38.

Bakshy, E., J. M. Hofman, W. A. Mason, and D. J. Watts 2011. Everyone's an influencer: quantifying influence on Twitter. In Proceedings of the fourth ACM international conference on Web search and data mining, WSDM '11, New York, NY, USA, pp. 65-74.ACM.

Beard, T.R., J.T. Macher and J.M. Mayo. 2015. Can You Hear Me Now? Exit, Voice and Loyalty Under Increasing Competition. *Journal of Law and Economics* 58(3), 717-745.

Berger, Jonah, Eric Schwartz. 2011. What drives immediate and ongoing word of mouth? *Journal of Marketing Research* 48(5), 869-880.

Borenstein, Severin. 1989. Hubs and high fares: Dominance and market power in the U.S. airline industry. *RAND Journal of Economics* 20(3), 344-365.

Borenstein, Severin. 1991. The Dominant-Firm Advantage in Multiproduct Industries: Evidence from the U.S. Airlines. *Quarterly Journal of Economics* 106(4), 1237-1266.

Breuckner, Jan, Darin Lee, and Ethan Singer. 2014. City-Pairs vs Airport-Pairs: A Market-Definition Methodology for the Airline Industry. *Review of Industrial Organization* 44, 1-25.

Cairns, Ian. 2016. Making customer service even better on Twitter. *mimeo.,* Twitter.

Chetty, Raj, John N. Friedman and Jonah E. Rockoff. 2014. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9): 2593-2632.

Chevalier, Judith and Dina Mayzlin 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-354.

Delo, Cotton. 2014. Wells Fargo Opens Command Center to Handle Surge of Social Content. *Advertising Age*, Published online April 8, 2014. http://adage.com/article/cmo-strategy/risk-averse-wells-fargo-opens-social-media-command-center/292476/ Accessed May 11, 2015.

Forbes, Silke 2008. The Effect of Service Quality and Expectations on Customer Complaints. *Journal of Industrial Economics* 56(1), pp. 190-213.

Fornell, Claes, and Birger Wernerfelt. 1987. Defensive Marketing Strategy by Customer Complaint Management: A Theoretical Analysis. *Journal of Marketing Research* 24(4), 337-346.

Fornell, Claes, and Birger Wernerfelt. 1988. A Model for Customer Complaint Management. *Marketing Science* 7(3), 287-298.

Freeman, Richard B. 1976. Individual Mobility and Union Voice in the Labor Market. *American Economic Review* 66(2), 361–68.

Godes, David and Dina Mayzlin. 2004. Using Online Conversations to Study Word of Mouth Communication. *Marketing Science*, 23(4), 545-560

Gatignon, Hubert and Thomas S. Robertson. 1986. An Exchange Theory Model of Interpersonal Communication. *Advances in Consumer Research*, 13, 534-38.

Godes, David and Dina Mayzlin. 2009. Firm-Created Word-of-Mouth Communication: Evidence from a Field Study. *Marketing Science*, 28 (4), 721-739.

Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty*. Harvard University Press, Cambridge MA.

Hirschman, Albert O. 1976. Discussion. *American Economic Review* 66(2), 386–391.

Horrace, William C. and Ronald L. Oaxaca. 2006. Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters* 90, 321-327.

Huang, Wayne 2016. New research: consumers willing to spend more after a positive customer service interaction on Twitter. *mimeo*., Twitter.

King, Gary, and Langche Zeng. 2001. Logistic Regression in Rare Events Data. *Political Analysis* 9, 137–163.

Lederman, Mara. 2007. Do Enhancements to Loyalty Programs Affect Demand? The Impact of Frequent Flyer Partnerships on Domestic Airline Demand. *RAND Journal of Economics* 38(4), 1134-1158.

Lederman, Mara. 2008. Are Frequent Flyer Programs a Cause of the Hub Premium? *Journal of Economics and Management Strategy* 17(1), 35-66.

Levin, Jonathan. 2002. Multilateral Contracting and the Employment Relationship. *Quarterly Journal of Economics*, 117(3), 1075-1103.

Ma, Liye, Baohong Sun, and Sunder Kekre. 2015. The Squeaky Wheel Gets the Grease—An empirical analysis of customer voice and firm intervention on Twitter. *Marketing Science* 34(5), 627-645.

Mayzlin, Dina. 2006. Promotional Chat on the Internet. *Marketing Science*, 25 (2), 155-163

Mayzlin, Dina, Yaniv Dover, and Judy Chevalier. 2014. Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *American Economic Review* 104(8), 2421-2455.

Miller, Amalia, and Catherine Tucker. 2013. Active Social Media Management: The Case of Health Care. *Information Systems Research* 24(1), 52-70

Nelson, Richard R. 1976. Discussion. *American Economic Review* 66(2), 386–391.

Nosko, Chirs and Steve Tadelis. 2015. The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment. Working paper, University of California, Berkeley.

Pak, Alexander, and Patrick Paroubek, 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proceedings of the International Conference on Language Resources and Evaluation*, 1320-1326.

Pang, Bo, and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval* 2(1-2), 1-135.

Read, Jonathon. 2005. Using Emoticons to reduce Dependency in Machine Learning Techniques of Sentiment Classification. *Proceedings of the ACL student Research Workshop*, p. 43-48.

Richins, Marsha L. 1983. Negative WOM by Dissatisfied Consumers: A Pilot Study. *Journal of Marketing*, 47(1), 68-78.

Trusov, M., R.E. Bucklin, and K. Pauwels. 2009. Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing* 73(5), 90-102.

Wei, Zaiyan, and Mo Xiao. 2015. For Whom to Tweet? A Study of a Large-Scale Social Network. Working paper, University of Arizona.

Williamson, Oliver E. 1976. The Economics of Internal Organization: Exit and Voice in Relation to Markets and Hierarchies. *American Economic Review* 66 (2), 369–77.

Young, Dennis R. 1976. Consolidation or Diversity: Choices in the Structure of Urban Governance. *American Economic Review* 66 (2), 378–85.

Table 1: Tweet-level Descriptive Statistics

Panel A: GEOGRAPHIC INFORMATION IN FULL SAMPLE					
Variable	Obs.	Mean	Std. Dev	Min	Max
Location given in profile	3,860,528	0.3611	0.4803	0	1
Airport mentioned in tweet	3,860,528	0.0434	0.2037	0	1
Geocode stamp on tweet	3,860,528	0.0727	0.2597	0	1
Any location information	3,860,528	0.4199	0.4935	0	1
Airport in tweet or geocode	3,860,528	0.1107	0.3138	0	1

Panel B: FRACTION OF TWEETS BY AIRLINE			
	FULL SAMPLE	**SAMPLE WITH AIRPORT INFORMATION (GEOCODE OR IN TWEET)**	**SAMPLE WITH CITY INFORMATION (GEOCODE, IN TWEET, OR CITY IN PROFILE)**
American Airlines	0.2560	0.2451	0.2584
Alaska Airlines	0.0292	0.0265	0.0343
JetBlue	0.1203	0.1269	0.1389
Delta Air Lines	0.1291	0.1499	0.1349
United Airlines	0.2495	0.2380	0.2082
US Airways	0.0993	0.0999	0.0959
Southwest Airlines	0.1167	0.1136	0.1293

Table 2: Location-Airline-Day Descriptive Statistics

CITY LEVEL DATA					
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Min**	**Max**
# tweets (location given in profile)	318,077	4.2575	12.3134	0	1179
# tweets (any location definition)	318,077	4.6184	13.0956	0	1212
Airline-airport flights	318,077	33.7528	78.2108	1	948
Airline share of flights in city					
Under 15%	318,077	0.4830	0.4997	0	1
15-30%	318,077	0.3477	0.4762	0	1
30-50%	318,077	0.1219	0.3272	0	1
Over 50%	318,077	0.0474	0.2125	0	1
Number delayed					
Dep. delay > 15 min. or canceled	318,077	7.1733	22.0518	0	813
Delays that are airline's fault	221,957	6.8400	17.8120	0	466.8
Delays that are not airline's fault	221,957	2.7761	8.6877	0	622.3
Tweet content (for location in profile tweets)					
# tweets to handle	318,077	2.9590	8.9405	0	768
# tweets not to handle	318,077	1.2986	4.4583	0	492
Average sentiment	177,703	0.3580	0.2915	0	1
# tweets mention on time performance	318,077	0.7735	2.8192	0	450
# very positive tweets	318,077	1.8997	5.6670	0	457
# very negative tweets	318,077	0.9768	3.5939	0	587
Average # followers	177,703	3737.8	23844.7	0	2,917,676
AIRPORT LEVEL DATA					
Variable					
# tweets (geocode stamp)	382,141	0.5900	1.8693	0	97
Airline-airport flights/week	382,141	28.2882	69.5922	1	948
Airline share of flights at airport					
Under 15%	382,141	0.5045	0.5000	0	1
15-30%	382,141	0.3230	0.4676	0	1
30-50%	382,141	0.1288	0.3350	0	1
Over 50%	382,141	0.0413	0.1990	0	1
Number delayed	382,141	0.5045	0.5000	0	1
Dep. delay > 15 min. or canceled	382,141	6.0115	19.6152	0	813

Table 3: Sample tweets by sentiment

Tweets with probability negative less than 0.01
thanks @united for the upgrade to an exit row seat; just arrived at dulles. #goodservice
@united @boeingairplanes incredible plane design! really like the gold streak across the front of the plane as well!
@americanair you're welcome american airlines. i love your planes, they are very bigs.
thanks @unitedairlines for another great flight to nyc!
Tweets with probability negative of 0.10
love the @united premieraccess telephone number. no waiting & no change fee.
congrats @southwestair you are 5 for 5 in being late on flights. i'm at 300 hours of list time for the year!
@southwestair this is a nice aircraft with the slick blue over head lighting and better design air vents... #greatcompany
Tweets with probability negative of 0.30
@united will your b787 ever fly to @heathrowairport
is it just me or has @united gotten better... two upgrades in one travel.
@jetblue not much info. looks like they are taking us back to the gate now.
Tweets with probability negative of 0.50
knock knock @united anybody home ??
i can't. i am done. standing applause for southwest airlines, no encore, i can't do it
@united - i gave many of years to ual for which i'm grateful.
judge approves american airlines' bankruptcy plan - yahoo finance http://t.co/z701ojfrnv via @yahoofinance
Tweets with probability negative of 0.70
@united why in the world did you guys do away with infant preboarding?
@americanair about to but flight is oversold. thoughts?
crazy traffic, on my way to #jfk #delta
Tweets with probability negative of 0.90
@united embarrassing to fly with you tonight. multiple points of failure.
11 hours later i've arrived in austin, cheers @americanair #awful
@americanair classless, no help flt attendants. airline industry is just so sad.
Tweets with probability negative more than 0.99
@united you have terrible customer service. how do you run a business with such uneducated employees
delayed 12 hrs @united customer service packed with complaints #typical #embarrassingairline
@jetblue even more disappointing that you're making seem like she accidentally hung up on me #jetbluetakesnoblame
@americanair just ignore me if you want, but don't patronize me. your service sucks. if you cared you would do something.

Table 4: Correlation between On-Time Performance and Number of Tweets, by Dominance

Airline share of flights at airports in the city	Correlation coefficient
Under 15%	0.112
15-30%	0.125
30-50%	0.181
Over 50%	0.204

Unit of observation is airline-location-day. Location identified as location in profile. Correlation coefficients shown.

Table 5
Relationship between On-Time Performance and Tweet Volume

	(1)	(2)	(3)	(4)
Dependent Variable	Standardized # Tweets	Standardized # Tweets	Log(# Tweets+1)	Standardized # Tweets
Location Measure	Location in profile (city)	Any Location Information (city)	Location in profile (city)	Geocoded Tweets (airport)
# flights delayed>15 min or canceled	0.078***	0.081***	0.069***	0.052***
	(0.005)	(0.005)	(0.004)	(0.004)
# airline flights departing that location	0.001	0.0004	0.001	-0.0001
	(0.004)	(0.004)	(0.009)	(0.003)
Fixed effects	Day-location	Day-location	Day-location, Airline-location	Day-location
N	318,077	328,692	338,754	382,141
R-sq	0.005	0.005	0.451	0.002

Dependent variable identified in column headers. In columns 1, 2, and 4, all variables are normalized using airline-location mean and standard deviation (and so airline-location fixed effects are not included). In column 3, variables are logged. Unit of observation is the location-airline-day. In columns 1-3, location is defined by city. In column 4, location is defined by airport. Robust standard errors clustered by airport in parentheses. Airline-location fixed effects are estimated directly in column 3. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 6
Relationship between On-Time Performance, Tweet Volume and Market Dominance

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Standardized # Tweets	Standardized # Tweets	Standardized # Tweets	Log(# Tweets+1)	Standardized # Tweets
Location Measure	Location in profile (city)	Location in profile (city)	Any Location Information (city)	Location in profile (city)	Geocoded Tweets (airport)
# flights delayed >15 min or canceled	0.063***	0.069***	0.071***	0.063***	0.044***
	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)
# flights delayed >15 min or canceled × 15-30% share	0.013+				
	(0.008)				
# flights delayed >15 min or canceled × 30-50% share	0.054***	0.048***	0.051***	0.023**	0.040***
	(0.012)	(0.012)	(0.011)	(0.007)	(0.009)
# flights delayed >15 min or canceled × >50% share	0.091***	0.087***	0.094***	0.061***	0.097***
	(0.020)	(0.020)	(0.021)	(0.017)	(0.019)
# airline flights departing that airport	-0.0001	0.000008	-0.0002	0.001	-0.001
	(0.004)	(0.004)	(0.004)	(0.009)	(0.003)
Fixed effects	Day-location	Day-location	Day-location	Day-location, Airline-location	Day-location
N	318,077	318,077	328,692	338,754	382,141
R-sq	0.005	0.005	0.006	0.451	0.003

Dependent variable identified in column headers. In columns 1, 2, 3, and 5, all variables are normalized using airline-location mean and standard deviation. In column 4, variables are logged. Unit of observation is the location-airline-day. In columns 1-4, location is defined by city. In column 5, location is defined by airport. Robust standard errors clustered by airport in parentheses. Airline-location fixed effects are estimated directly in column 4. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 7
Relationship between On-Time Performance, Tweet Volume and Market Dominance, by On-Time Performance Mentioned in Tweet

	(1)	(2)	(3)	(4)
Dependent Variable	Standardized # tweets about on-time performance	Standardized # tweets not about on-time performance	Standardized # tweets about on-time performance	Standardized # tweets not about on-time performance
# flights delayed >15 min or canceled	0.112***	0.052***	0.103***	0.045***
	(0.008)	(0.004)	(0.007)	(0.004)
# flights delayed >15 min or canceled × 30-50% share			0.041**	0.042***
			(0.015)	(0.010)
# flights delayed >15 min or canceled × >50% share			0.119***	0.066***
			(0.025)	(0.016)
# airline flights departing that airport	-0.012***	0.005	-0.013***	0.005
	(0.003)	(0.004)	(0.003)	(0.004)
Fixed effects	Day-location	Day-location	Day-location	Day-location
N	318,077	318,077	318,077	318,077
R-sq	0.009	0.002	0.010	0.003

Dependent variable identified in column headers. All variables are normalized using airline-location mean and standard deviation. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 8
Relationship between On-Time Performance, Tweet Volume and Market Dominance, by Tweet Sentiment

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Standardized Average negative sentiment of tweets	Standardized Average negative sentiment of tweets	Standardized # very negative tweets	Standardized # very positive tweets	Standardized # very negative tweets	Standardized # very positive tweets
# flights delayed or canceled	0.080***	0.072***	0.097***	0.026***	0.088***	0.020***
	(0.007)	(0.006)	(0.007)	(0.003)	(0.007)	(0.003)
# flights delayed >15 min or canceled × 30-50% share		0.047**			0.044**	0.033***
		(0.015)			(0.013)	(0.009)
# flights delayed >15 min or canceled × >50% share		0.044*			0.106***	0.056***
		(0.020)			(0.025)	(0.011)
# airline flights departing that airport	-0.012*	-0.012*	-0.010*	0.011**	-0.010**	0.011**
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Fixed effects	Day-location	Day-location	Day-location	Day-location	Day-location	Day-location
N	177,703	177,703	317,325	318,077	317,325	318,077
R-sq	0.004	0.004	0.007	0.001	0.007	0.001

Dependent variable identified in column headers. All variables are normalized using airline-location mean and standard deviation. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 9
Weather, Delay Cause, and the Relationship between On-Time Performance, Tweet Volume and Market Dominance

	(1)	(2)	(2)
Dependent Variable	Standardized # tweets	Standardized # tweets	Standardized # tweets
# flights delayed or canceled	0.070***	0.071***	
	(0.006)	(0.006)	
# flights delayed >15 min or canceled × 30-50% share	0.049***	0.049***	
	(0.012)	(0.012)	
# flights delayed >15 min or canceled × >50% share	0.075***	0.073***	
	(0.020)	(0.019)	
Rain, Snow, or Fog Dummy × 30-50% share	-0.0002		
	(0.006)		
Rain, Snow, or Fog Dummy × >50% share	-0.011		
	(0.008)		
Quantity of Precipitation × 30-50% share		-0.006	
		(0.005)	
Quantity of Precipitation × >50% share		-0.003	
		(0.010)	
# flights delayed > 15 min that are airline's fault			0.063***
			(0.005)
# flights delayed > 15 min that are airline's fault × 30-50% share			0.032*
			(0.012)
# flights delayed >15 min that are airline's fault × >50% share			0.058**
			(0.018)
# flights delayed > 15 min that are *not airline's fault*			0.038***
			(0.004)
# flights delayed > 15 min that are *not* airline's fault × 30-50% share			0.017
			(0.010)
# flights delayed >15 min that are *not* airline's fault × >50% share			0.016
			(0.024)
# airline flights departing that airport	-0.001	-0.001	0.002
	(0.004)	(0.004)	(0.005)
Fixed effects	Day-location	Day-location	Day-location
N	292,295	289,439	221,957
R-sq	0.005	0.005	0.006

Dependent variable is city-level tweets with the location in profile known. Airline fault is defined by the airline in regulatory filings. All variables are normalized using airline-location mean and standard deviation. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 10
Response Rates

	(1)	(2)	(3)
30-50% share	0.241***		0.238***
	(0.008)		(0.008)
>50% share	0.176***		0.173***
	(0.013)		(0.013)
Frequent flier keyword		0.262***	0.258***
		(0.027)	(0.027)
Probability sentiment is negative	0.048**	0.060***	0.062***
	(0.017)	(0.017)	(0.017)
# followers, 25th -50th percentile	0.042***	0.057***	0.043***
	(0.009)	(0.009)	(0.009)
# followers, 50th -75th percentile	-0.054***	-0.034**	-0.052***
	(0.011)	(0.011)	(0.011)
# followers, 75th -99th percentile	-0.119***	-0.096***	-0.118***
	(0.013)	(0.013)	(0.013)
# followers, over 99th percentile	0.135***	0.153***	0.136***
	(0.024)	(0.024)	(0.024)
Handle	3.125***	3.134***	3.120***
	(0.034)	(0.034)	(0.034)
Customer service keyword	0.392***	0.399***	0.398***
	(0.010)	(0.010)	(0.010)
On time performance keyword	0.482***	0.490***	0.486***
	(0.010)	(0.010)	(0.010)
American Airlines	4.024***	3.998***	4.017***
	(0.071)	(0.071)	(0.071)
Alaska Airlines	2.630***	2.639***	2.628***
	(0.077)	(0.077)	(0.077)
JetBlue	3.356***	3.339***	3.359***
	(0.074)	(0.074)	(0.074)
Delta Air Lines	1.397***	1.385***	1.382***
	(0.071)	(0.071)	(0.071)
United Airlines	2.819***	2.818***	2.803***
	(0.071)	(0.071)	(0.071)
Date	0.001***	0.001***	0.001***
	(0.0001)	(0.0001)	(0.0001)
N	3,477,105	3,477,105	3,477,105
Log Likelihood	-1,231,187	-1,230,723	-1,229,926

Logit regression. Dependent variable is whether the airline responded to the tweet. Unit of observation is the tweet. Southwest airlines is the base for the airline dummy variables. No response data for US Airways. Regressions include 11 month-of-the-year dummy variables. +p<.10, *p<0.05, **p<0.01, ***p<0.001

Table 11
Relationship between On-Time Performance, Tweet Volume and Market Dominance, Tweets to Handle and Not to Handle

	(1)	(2)	(3)	(4)
Dependent Variable	Standardized # tweets to handle	Standardized # tweets not to handle	Standardized # tweets to handle	Standardized # tweets not to handle
# flights delayed or canceled	0.069***	0.048***	0.059***	0.045***
	(0.005)	(0.004)	(0.005)	(0.004)
# flights delayed >15 min or canceled × 30-50% share			0.050***	0.015+
			(0.010)	(0.009)
# flights delayed >15 min or canceled × >50% share			0.092***	0.049**
			(0.021)	(0.015)
# airline flights departing that airport	0.002	0.001	0.001	0.001
	(0.004)	(0.004)	(0.004)	(0.004)
Fixed effects	Day-location	Day-location	Day-location	Day-location
N	318,077	317,844	318,077	317,844
R-sq	0.004	0.002	0.004	0.002

Dependent variable is in column headers with city-level tweets with the location in profile known. All variables are normalized using airline-location mean and standard deviation. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 12
Relationship between On-Time Performance, Market Dominance, and Average Number of Followers

	(1)	(2)
Dependent Variable	Standardized Average # of followers	Standardized Average # of followers
# flights delayed or canceled	0.0054+	0.0056
	(0.0032)	(0.0034)
# flights delayed >15 min or canceled × 30-50% share		-0.0018
		(0.0081)
# flights delayed >15 min or canceled × >50% share		0.0001
		(0.0099)
# airline flights departing that airport	-0.0008	-0.0008
	(0.0035)	(0.0035)
Fixed effects	Day-location	Day-location
N	177,671	177,671
R-sq	0.0001	0.0001

Dependent variable is in column headers with city-level tweets with the location in profile known. All variables are normalized using airline-location mean and standard deviation. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 13
Relationship between Receiving a Response to a Tweet and Tweeting Again

Dependent Variable	=1 if Tweet again, after first tweet		=1 if Tweet in 2013 or 2014, given first tweet in 2012	
	(1)	(2)	(3)	(4)
Airline responded to first tweet	0.897***	0.772***	0.517***	0.341***
	(0.005)	(0.006)	(0.015)	(0.017)
Frequent flier keyword		0.221***		0.503***
		(0.009)		(0.021)
30-50% share		0.234***		0.401***
		(0.007)		(0.015)
>50% share		0.435***		0.798***
		(0.011)		(0.024)
Probability sentiment is negative		0.056***		-0.204***
		(0.005)		(0.012)
# followers, 25th -50th percentile		0.044***		0.292***
		(0.005)		(0.012)
# followers, 50th -75th percentile		0.169***		0.466***
		(0.005)		(0.013)
# followers, 75th -99th percentile		0.446***		0.804***
		(0.006)		(0.014)
# followers, over 99th percentile		0.703***		1.217***
		(0.026)		(0.060)
Handle		0.490***		0.503***
		(0.004)		(0.010)
Customer service keyword		0.081***		0.017
		(0.006)		(0.015)
On time performance keyword		0.123***		0.045***
		(0.006)		(0.013)
American Airlines		0.059***		-0.066***
		(0.007)		(0.016)
Alaska Airlines		-0.098***		-0.313***
		(0.012)		(0.032)
JetBlue		0.103***		-0.113***
		(0.007)		(0.018)
Delta Air Lines		-0.195***		-0.166***
		(0.007)		(0.017)
United Airlines		0.096***		0.130***
		(0.006)		(0.016)
Date		-0.002***		-0.002***
		(0.0001)		(0.0001)
Constant	-1.035***	33.811***	-1.064***	44.824***
	(0.002)	(0.182)	(0.005)	(1.908)
N	1,375,416	1,375,416	259,299	259,299
Log Likelihood	-812,567	-780,934	-149,639	-144,111

Dependent variable in columns 1 and 2 is whether tweeted again to the same airline. Dependent variable in columns 3 and 4 is whether tweeted again to the same airline in 2013 or 2014. Sample in columns 1 and 2 is first tweet. Sample in columns 3 and 4 is first tweet by tweeter in 2012. Unit of observation is the tweeter. Logit regression. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Figure 1a: Average Daily Tweets by Month (Data with city information)

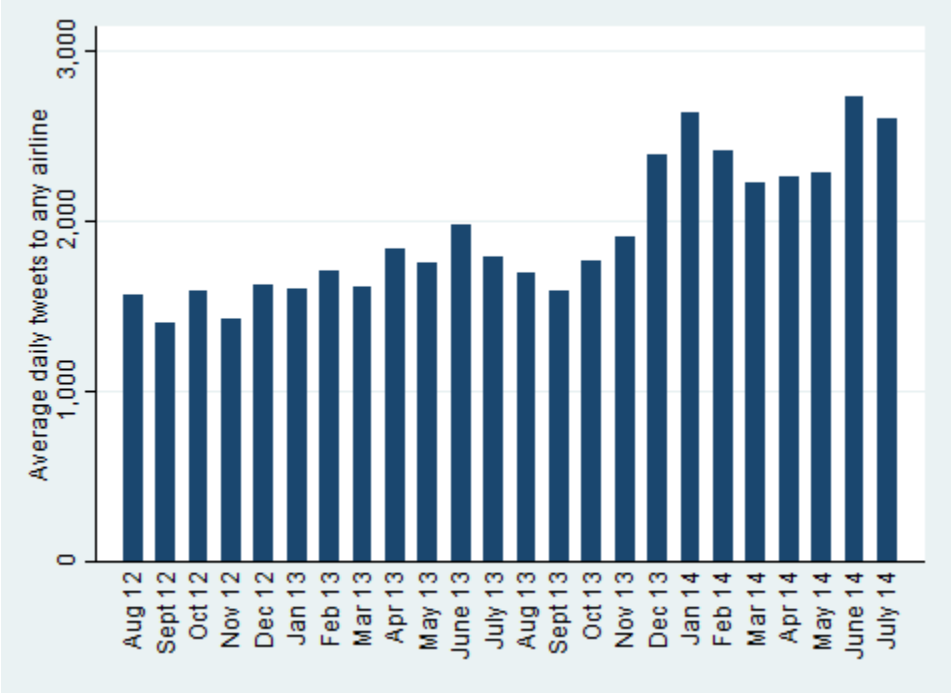


Figure 1b: Average Daily Tweets by Month by Airline (Data with city information)

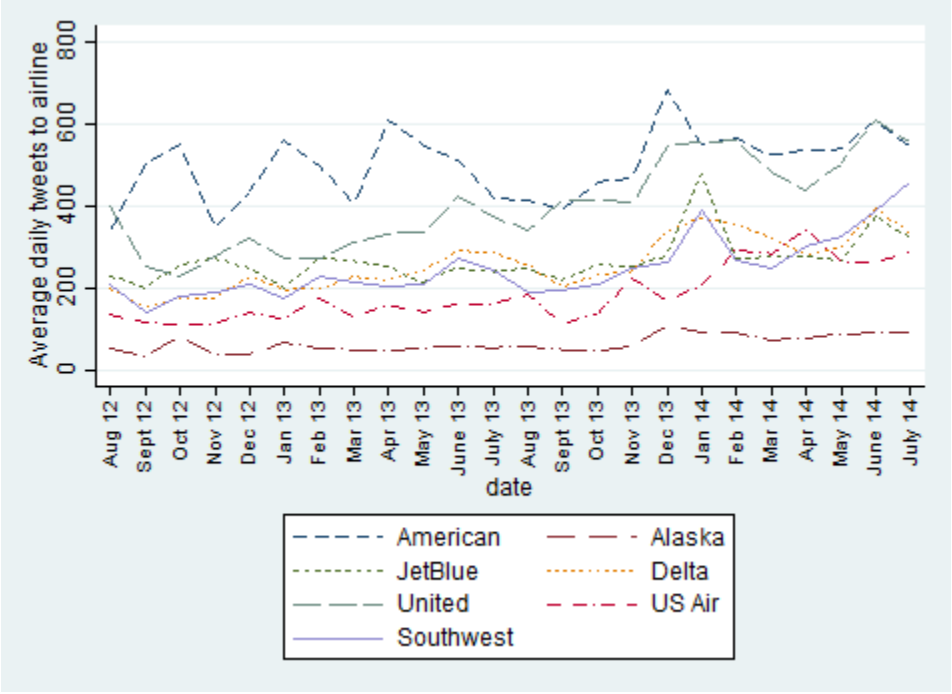


Figure 2a: Response rates, over time

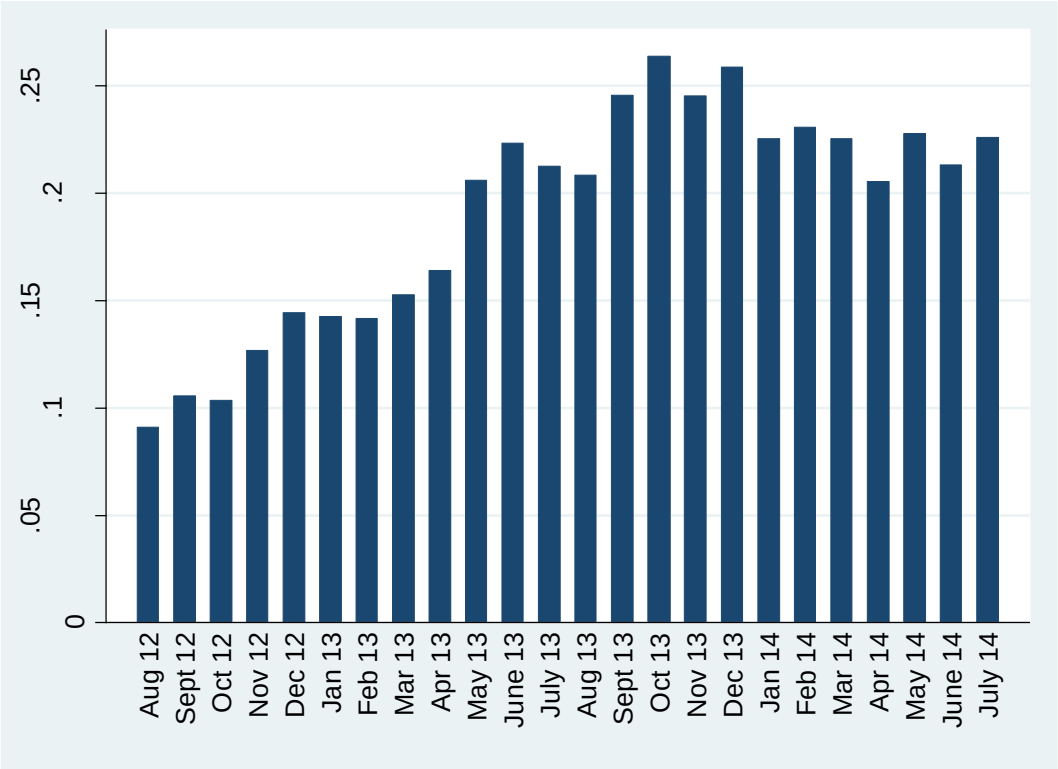
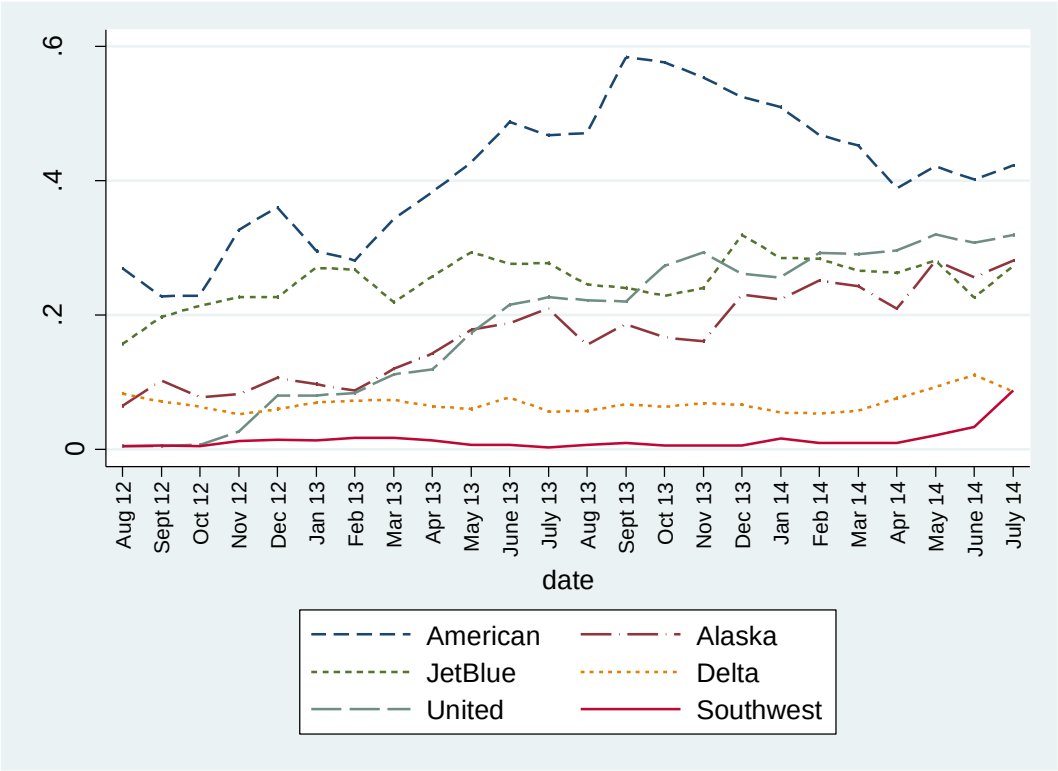


Figure 2b: Response rates by airline, over time



ONLINE APPENDIX to

Exit, Tweets, and Loyalty

By Joshua S. Gans, Avi Goldfarb, and Mara Lederman

Not for publication

Table of Contents

APPENDIX A: TABLES

Table A1: Descriptive Statistics for Response Data

Variable	Obs.	Mean	Std. Dev.	Min	Max
Airline replied	3,477,105	0.2143	0.4103	0	1
Airline replied if tweet to airline handle	2,040,504	0.3612	0.3956	0	1
Frequent flier keyword	3,477,105	0.0539	0.2258	0	1
Airline 30-50% share city	3,477,105	0.0958	0.2943	0	1
Airline >50% share city	3,477,105	0.0367	0.1880	0	1
Probability sentiment is negative	3,477,105	0.3612	0.3956	0	1
Number of followers, 25th -50th percentile	3,477,105	0.2521	0.4342	0	1
Number of followers, 50th -75th percentile	3,477,105	0.2503	0.4332	0	1
Number of followers, 75th -99th percentile	3,477,105	0.2379	0.4258	0	1
Number of followers, over 99th percentile	3,477,105	0.0099	0.0989	0	1
Handle	3,477,105	0.5868	0.4924	0	1
Customer service keyword	3,477,105	0.1041	0.3054	0	1
On time performance keyword	3,477,105	0.1588	0.3655	0	1
American Airlines	3,477,105	0.2842	0.4510	0	1
Alaska Airlines	3,477,105	0.0324	0.1770	0	1
JetBlue	3,477,105	0.1336	0.3402	0	1
Delta Air Lines	3,477,105	0.1433	0.3504	0	1
United Airlines	3,477,105	0.2770	0.4475	0	1

US Airways tweets are omitted as there is no response data to those tweets.

Table A2: Descriptive Statistics for Repeat Tweeter Analysis

Variable	Obs.	Mean	Std. Dev.	Min	Max
All tweeters, first tweet about airline					
Tweeted again to same airline	1,457,945	0.3296	0.4701	0	1
Airline replied	1,457,945	0.1475	0.3547	0	1
Frequent flier keyword	1,457,945	0.0436	0.2041	0	1
Airline 30-50% share city	1,457,945	0.0823	0.2748	0	1
Airline >50% share city	1,457,945	0.0270	0.1622	0	1
Probability sentiment is negative	1,457,945	0.3539	0.3932	0	1
Number of followers, 25th -50th percentile	1,457,945	0.2844	0.4512	0	1
Number of followers, 50th -75th percentile	1,457,945	0.2379	0.4258	0	1
Number of followers, 75th -99th percentile	1,457,945	0.1656	0.3717	0	1
Number of followers, over 99th percentile	1,457,945	0.0050	0.0703	0	1
Handle	1,457,945	0.5038	0.5000	0	1
Customer service keyword	1,457,945	0.1011	0.3015	0	1
On time performance keyword	1,457,945	0.1559	0.3628	0	1
American Airlines	1,457,945	1,457,945	0.2369	0.4252	0
Alaska Airlines	1,457,945	1,457,945	0.0343	0.1821	0
JetBlue	1,457,945	1,457,945	0.1343	0.3410	0
Delta Air Lines	1,457,945	1,457,945	0.1748	0.3798	0
United Airlines	1,457,945	1,457,945	0.2650	0.4413	0
First tweet for 2012 tweets					
Tweeted to same airline in 2013 or 2014	259,299	0.3933	0.4885	0	1
Airline replied	259,299	0.0809	0.2728	0	1
Frequent flier keyword	259,299	0.0409	0.1981	0	1
Airline 30-50% share city	259,299	0.0887	0.2843	0	1
Airline >50% share city	259,299	0.0316	0.1748	0	1
Probability sentiment is negative	259,299	0.3521	0.3919	0	1
Number of followers, 25th -50th percentile	259,299	0.2991	0.4579	0	1
Number of followers, 50th -75th percentile	259,299	0.2360	0.4246	0	1
Number of followers, 75th -99th percentile	259,299	0.1665	0.3726	0	1
Number of followers, over 99th percentile	259,299	0.0046	0.0679	0	1
Handle	259,299	0.3929	0.4884	0	1
Customer service keyword	259,299	0.0935	0.2912	0	1
On time performance keyword	259,299	0.1503	0.3573	0	1
American Airlines	259,299	0.2580	0.4376	0	1
Alaska Airlines	259,299	0.0275	0.1635	0	1
JetBlue	259,299	0.1583	0.3651	0	1
Delta Air Lines	259,299	0.1546	0.3615	0	1
United Airlines	259,299	0.2781	0.4481	0	1

ROBUSTNESS TO LOGGED SPECIFICATION

Table A3
Robustness of Table 5: Tweets and On-Time Performance

	(1)	(2)
	City-level location in profile only	City-level all three location measures
Flights delayed or canceled	0.069***	0.073***
	(0.004)	(0.004)
Airline flights departing that location	0.001	0.001
	(0.009)	(0.009)
Fixed effects	Day-location, Airline-location	Day-location, Airline-location
N	338,754	338,754
R-sq	0.451	0.468

Dependent variable is number of tweets as identified in column headers. Number of tweets and delays use log(variable+1). Airline flights is logged. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Airline-location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A4
Robustness of Table 6: Tweets, On-Time Performance, and Market Dominance

	(1)	(2)	(3)
	City-level location in profile only	City-level location in profile only	City-level all three location measures
Flights delayed or canceled	0.061***	0.063***	0.066***
	(0.005)	(0.005)	(0.005)
Flights delayed or canceled x Airline 15-30% share city	0.004		
	(0.007)		
Flights delayed or canceled x Airline 30-50% share city	0.025**	0.023**	0.026***
	(0.007)	(0.007)	(0.007)
Flights delayed or canceled x Airline >50% share city	0.062***	0.061***	0.068***
	(0.017)	(0.017)	(0.017)
Airline flights departing that airport	0.001	0.001	0.001
	(0.009)	(0.009)	(0.009)
Fixed effects	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location
N	338,754	338,754	338,754
R-sq	0.451	0.451	0.468

Dependent variable is number of tweets as identified in column headers. Number of tweets and delays use log(variable+1). Airline flights is logged. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Airline-location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A5
Robustness of Table 7: On-Time Performance Mentioned in Tweet

	(1)	(2)	(3)	(4)
	Number tweets about on-time performance	Number tweets not about on-time performance	Number tweets about on-time performance	Number tweets not about on-time performance
Flights delayed or canceled	0.071***	0.047***	0.059***	0.042***
	(0.007)	(0.004)	(0.006)	(0.004)
Flights delayed or canceled x Airline 30-50% share city			0.044**	0.019**
			(0.014)	(0.006)
Flights delayed or canceled x Airline >50% share city			0.112**	0.053***
			(0.033)	(0.013)
Airline flights departing that airport	-0.020***	0.007	-0.021***	0.006
	(0.004)	(0.008)	(0.004)	(0.008)
Fixed effects	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location
N	338,754	338,754	338,754	338,754
R-sq	0.357	0.442	0.359	0.443

Dependent variable type identified in column headers. Number of tweets and delays use log(variable+1). Airline flights is logged. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Airline-location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A6
Robustness of Table 8: Sentiment

	(1)	(2)	(3)	(4)	(5)	(6)
	Average negative sentiment	Average negative sentiment	Number of very negative tweets	Number of very positive tweets	Number of very negative tweets	Number of very positive tweets
Flights delayed or canceled	0.026***	0.027***	0.070***	0.024***	0.060***	0.019***
	(0.002)	(0.002)	(0.007)	(0.002)	(0.006)	(0.003)
Flights delayed or canceled x Airline 30-50% share city		-0.001			0.038**	0.019**
		(0.003)			(0.012)	(0.006)
Flights delayed or canceled x Airline >50% share city		-0.007+			0.100**	0.054***
		(0.004)			(0.030)	(0.011)
Airline flights departing that airport	-0.010*	-0.010*	-0.017***	0.011	-0.018***	0.011
	(0.005)	(0.005)	(0.005)	(0.007)	(0.005)	(0.007)
Fixed effects	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location
N	178,908	178,908	338,754	338,754	338,754	338,754
R-sq	0.079	0.079	0.371	0.419	0.372	0.420

Dependent variable type identified in column headers. Number of tweets and delays use log(variable+1). Airline flights is logged. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Airline-location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A7
Robustness of Table 9
Weather, Delay Cause, and the Relationship between On-Time Performance, Tweet Volume and Market Dominance

	(1)	(2)	(2)
Dependent Variable	# tweets	# tweets	# tweets
# flights delayed or canceled	0.063***	0.063***	
	(0.005)	(0.005)	
# flights delayed >15 min or canceled × 30-50% share	0.020**	0.022**	
	(0.008)	(0.008)	
# flights delayed >15 min or canceled × >50% share	0.056**	0.055**	
	(0.018)	(0.018)	
Rain, Snow, or Fog Dummy × 30-50% share	0.013		
	(0.008)		
Rain, Snow, or Fog Dummy × >50% share	0.004		
	(0.012)		
Quantity of Precipitation × 30-50% share		-0.012	
		(0.007)	
Quantity of Precipitation × >50% share		0.020	
		(0.019)	
# flights delayed > 15 min that are airline's fault			0.058***
			(0.004)
# flights delayed > 15 min that are airline's fault × 30-50% share			0.014
			(0.009)
# flights delayed >15 min that are airline's fault × >50% share			0.036*
			(0.018)
# flights delayed > 15 min that are *not airline's fault*			0.047***
			(0.005)
# flights delayed > 15 min that are *not* airline's fault × 30-50% share			-0.005
			(0.008)
# flights delayed >15 min that are *not* airline's fault × >50% share			0.001
			(0.015)
# airline flights departing that airport	0.0001	0.0001	0.010
	(0.009)	(0.009)	(0.012)
Fixed effects	Day-location	Day-location	Day-location
N	312,011	308,715	229,984
R-sq	0.435	0.435	0.429

Dependent variable is number of tweets with city-location known. Number of tweets and delays use log(variable+1). Airline flights is logged. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Airline-location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A8
Robustness of Table 11: Handles

	(1)	(2)	(3)	(4)
	Number tweets to handle	Number tweets not to handle	Number tweets to handle	Number tweets not to handle
Flights delayed or canceled	0.065***	0.030***	0.057***	0.027***
	(0.005)	(0.003)	(0.005)	(0.003)
Flights delayed or canceled x Airline 30-50% share city			0.028***	0.006
			(0.007)	(0.006)
Flights delayed or canceled x Airline >50% share city			0.077***	0.038**
			(0.019)	(0.014)
Airline flights departing that airport	-0.004	0.002	-0.004	0.001
	(0.008)	(0.007)	(0.008)	(0.007)
Fixed effects	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location
N	338,754	338,754	338,754	338,754
R-sq	0.452	0.320	0.452	0.320

Dependent variable type identified in column headers. Number of tweets and delays use log(variable+1). Airline flights is logged. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Airline-location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A9
Robustness of Table 12
Relationship between On-Time Performance, Market Dominance, and Average Number of Followers

	(1)	(2)
Dependent Variable	Average # of followers	Average # of followers
# flights delayed or canceled	0.023*	0.019*
	(0.009)	(0.009)
# flights delayed >15 min or canceled × 30-50% share		0.022
		(0.024)
# flights delayed >15 min or canceled × >50% share		0.029
		(0.028)
# airline flights departing that airport	-0.021	-0.021
	(0.038)	(0.039)
Fixed effects	Day-location, Airline-location	Day-location, Airline-location
N	178,810	178,810
R-sq	0.059	0.059

Dependent variable is in column headers with city-level tweets with the location in profile known. Number of delays and average number of followers use log(variable+1). Airline flights is logged. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by airport in parentheses. Airline-location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

ROBUSTNESS TO AIRPORT LEVEL SPECIFICATION
Table A10
Robustness of Table 5: Tweets and On-Time Performance

	(1)	(2)	(3)	(4)
Dependent Variable	Standardized # Tweets	Standardized # Tweets	Standardized # Tweets	Standardized # Tweets
Location Measure	Closest airport	Airport in tweet	Both Airport-level location measures	Within two miles of airport
Flights delayed or canceled	0.052***	0.066***	0.073***	0.046***
	(0.004)	(0.004)	(0.004)	(0.004)
Airline flights departing that location	-0.0001	0.002	0.002	0.006+
	(0.003)	(0.003)	(0.003)	(0.003)
Fixed effects	Day-location	Day-location	Day-location	Day-location
N	382,141	382,141	382,141	371,883
R-sq	0.002	0.003	0.004	0.002

Dependent variable is number of tweets as identified in column headers. All variables are normalized using airline-airport mean and standard deviation. Location is defined by airport. Robust standard errors clustered by airport in parentheses. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A11
Robustness of Table 6: Tweets, On-Time Performance, and Market Dominance

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Standardized # Tweets	Standardized # Tweets	Standardized # Tweets	Standardized # Tweets	Standardized # Tweets
Location Measure	Closest airport	Closest airport	Airport in tweet	Both Airport-level location measures	Within two miles of airport
Flights delayed or canceled	0.039***	0.044***	0.055***	0.062***	0.039***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Flights delayed or canceled x Airline 15-30% share city	0.012+				
	(0.006)				
Flights delayed or canceled x Airline 30-50% share city	0.045***	0.040***	0.049***	0.051***	0.030**
	(0.009)	(0.009)	(0.011)	(0.009)	(0.010)
Flights delayed or canceled x Airline >50% share city	0.100***	0.097***	0.135***	0.136***	0.091***
	(0.018)	(0.019)	(0.023)	(0.023)	(0.019)
Airline flights departing that airport	-0.001	-0.001	0.001	0.001	0.005+
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Fixed effects	Day-location	Day-location	Day-location	Day-location	Day-location
N	382,141	382,141	382,141	382,141	371,883
R-sq	0.003	0.003	0.004	0.005	0.002

Dependent variable is number of tweets as identified in column headers. All variables are normalized using airline-airport mean and standard deviation. Location is defined by airport. Robust standard errors clustered by airport in parentheses. Day-location fixed effects are differenced out using stata's xtreg, fe command.
+p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A12
Robustness of Table 7: On-Time Performance Mentioned in Tweet

	(1)	(2)	(3)	(4)
	Standardized Number tweets about on-time performance	Standardized Number tweets not about on-time performance	Standardized Number tweets about on-time performance	Standardized Number tweets not about on-time performance
Flights delayed or canceled	0.064***	0.034***	0.056***	0.028***
	(0.005)	(0.003)	(0.004)	(0.003)
Flights delayed or canceled x Airline 30-50% share city			0.040***	0.031***
			(0.011)	(0.008)
Flights delayed or canceled x Airline >50% share city			0.108***	0.077***
			(0.020)	(0.018)
Airline flights departing that airport	-0.004	0.001	-0.005+	0.001
	(0.002)	(0.003)	(0.002)	(0.003)
Fixed effects	Day-location	Day-location	Day-location	Day-location
N	380,637	382,141	380,637	382,141
R-sq	0.003	0.001	0.003	0.001

Dependent variable is number of tweets as identified in column headers. All variables are normalized using airline-airport mean and standard deviation. Location is defined by airport. Robust standard errors clustered by airport in parentheses. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A13
Robustness of Table 8: Sentiment

	(1)	(2)	(3)	(4)	(5)	(6)
	Standardized Average negative sentiment	Standardized Average negative sentiment	Standardized Number of very negative tweets	Standardized Number of very positive tweets	Standardized Number of very negative tweets	Standardized Number of very positive tweets
Flights delayed or canceled	0.085***	0.079***	0.063***	0.014***	0.055***	0.009***
	(0.007)	(0.007)	(0.004)	(0.003)	(0.004)	(0.003)
Flights delayed or canceled x Airline 30-50% share city		0.036*			0.037**	0.021**
		(0.015)			(0.011)	(0.007)
Flights delayed or canceled x Airline >50% share city		0.005			0.104***	0.053***
		(0.014)			(0.019)	(0.015)
Airline flights departing that airport	-0.027***	-0.028***	-0.008**	0.006*	-0.008***	0.006*
	(0.007)	(0.007)	(0.002)	(0.003)	(0.002)	(0.003)
Fixed effects	Day-location	Day-location	Day-location	Day-location	Day-location	Day-location
N	88,807	88,807	379,590	382,141	379,590	382,141
R-sq	0.004	0.004	0.003	0.0001	0.003	0.0001

Dependent variable is number of tweets as identified in column headers. All variables are normalized using airline-airport mean and standard deviation. Location is defined by airport. Robust standard errors clustered by airport in parentheses. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A14
Robustness of Table 9: Weather, Delay Cause, and the Relationship between On-Time Performance, Tweet Volume and Market Dominance

	(1)	(2)	(2)
Dependent Variable	Standardized # tweets	Standardized # tweets	Standardized # tweets
# flights delayed or canceled	0.045***	0.044***	
	(0.004)	(0.004)	
# flights delayed >15 min or canceled × 30-50% share	0.041***	0.041***	
	(0.009)	(0.009)	
# flights delayed >15 min or canceled × >50% share	0.090***	0.094***	
	(0.020)	(0.020)	
Rain, Snow, or Fog Dummy × 30-50% share	0.0005		
	(0.005)		
Rain, Snow, or Fog Dummy × >50% share	0.004		
	(0.009)		
Quantity of Precipitation × 30-50% share		-0.005	
		(0.005)	
Quantity of Precipitation × >50% share		-0.013	
		(0.009)	
# flights delayed > 15 min that are airline's fault			0.040***
			(0.004)
# flights delayed > 15 min that are airline's fault × 30-50% share			0.017*
			(0.008)
# flights delayed >15 min that are airline's fault × >50% share			0.069***
			(0.016)
# flights delayed > 15 min that are *not airline's fault*			0.018***
			(0.004)
# flights delayed > 15 min that are *not* airline's fault × 30-50% share			0.017*
			(0.008)
# flights delayed >15 min that are *not* airline's fault × >50% share			0.026*
			(0.011)
# airline flights departing that airport	-0.001	-0.001	0.001
	(0.003)	(0.003)	(0.004)
Fixed effects	Day-location	Day-location	Day-location
N	352,415	348,881	260,442
R-sq	0.003	0.003	0.003

Dependent variable is airport-level tweets with goecode. Airline fault is defined by the airline in regulatory filings. All variables are normalized using airline-location mean and standard deviation. Unit of observation is the location-airline-day. Location is defined by airport. Robust standard errors clustered by airport in parentheses. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A15
Robustness of Table 11: Handles

	(1)	(2)	(3)	(4)
	Standardized Number tweets to handle	Standardized Number tweets not to handle	Standardized Number tweets to handle	Standardized Number tweets not to handle
Flights delayed or canceled	0.050***	0.024***	0.042***	0.020***
	(0.004)	(0.003)	(0.003)	(0.003)
Flights delayed or canceled x Airline 30-50% share city			0.043***	0.016+
			(0.009)	(0.008)
Flights delayed or canceled x Airline >50% share city			0.089***	0.068***
			(0.018)	(0.016)
Airline flights departing that airport	-0.002	0.005+	-0.003	0.005+
	(0.002)	(0.003)	(0.002)	(0.003)
Fixed effects	Day-location	Day-location	Day-location	Day-location
N	382,141	382,141	382,141	382,141
R-sq	0.002	0.0001	0.002	0.001

Dependent variable is number of tweets as identified in column headers. All variables are normalized using airline-location mean and standard deviation. Location is defined by airport. Robust standard errors clustered by airport in parentheses. Airline-location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table A6
Robustness of Table 12: Relationship between On-Time Performance, Market Dominance, and Average Number of Followers

	(1)	(2)
Dependent Variable	Standardized Average # of followers	Standardized Average # of followers
# flights delayed or canceled	0.001	-0.001
	(0.008)	(0.008)
# flights delayed >15 min or canceled × 30-50% share		0.011
		(0.009)
# flights delayed >15 min or canceled × >50% share		0.007
		(0.015)
# airline flights departing that airport	-0.005	-0.005
	(0.005)	(0.005)
Fixed effects	Day-location	Day-location
N	88,807	88,807
R-sq	0.0001	0.0001

Dependent variable is airport-level tweets with goecode. All variables are normalized using airline-location mean and standard deviation. Unit of observation is the location-airline-day. Location is defined by airport. Robust standard errors clustered by airport in parentheses. Day-location fixed effects are differenced out using stata's xtreg, fe command. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

APPENDIX B: EXTENSIONS TO THE THEORY

The model in the paper is geared towards a mass market industry like airlines. When a consumer leaves a firm, they do not expect the prices they face at other firms to change. In addition, the model does not fully model equilibrium outcomes. Here we amend the baseline model on both of these dimensions to demonstrate the robustness of the model's conclusions.

When exit raises price

Here we consider a situation where, if a consumer exits, it has one less firm it can deal with. Suppose that all firms understand this and so they can set prices to this consumer differently. Then, an exiting consumer will face a higher price from other firms.

To illustrate this, consider a Salop circle model with per unit transportation cost of *t*. Suppose that if a consumer exits a firm, they face a smaller circle with $n - 1$ firms evenly spaced. Thus, the price they face changes from $c + sB + \frac{t}{n}$ to $c + sB + \frac{t}{n-1}$.[27]

With this, should it exit, the total costs for the consumer if they stay with the firm (LHS of B1 below) is that they face the same price (because the other consumers with that firm are presumably still under the assumption that there is a relational contract) while if they exit (RHS of B1 below), the price goes up but the consumer believes that if there is a complaint, they will get mitigation. The costs of staying are higher than those of exit if:

$$c + sB + \frac{t}{n} \geq c + sB + \frac{t}{n-1} - s(B - C) \qquad \text{(B1)}$$

$$\Longrightarrow B \geq C + \frac{t}{sn(n-1)} \qquad \text{(B2)}$$

As *n* gets large, B2 will hold. For $n = 2$, $B \geq C + \frac{t}{2s}$. Notice that this constraint varies with *n* in the opposite direction as the firm incentive constraint. This captures the other part of the Hirschman intuition.

Given this, the full condition (that is, (*) in the paper's model) becomes:

$$\frac{\delta}{1-\delta}\frac{t}{n} \geq C + \frac{t}{sn(n-1)}$$

or

$$\frac{\delta st(n-1) - t(1-\delta)}{sn(n-1)(1-\delta)} \geq C$$

[27] In a more realistic setting, the fact that a consumer is at a particular firm means that it will have a further distance from other firms if it exits and, if other firms understood that, the pricing outcome may be different. This is an interesting possibility but involves unnecessary complications for our purposes here, so we make the simplifying 'shrinkage' assumption instead

It can be shown that this constraint is more likely to hold as *n* falls so the comparative static we have emphasized still holds. That is despite consumers facing increased prices if they exit, the relative costs of exit rise faster for firms.

Full Equilibrium Model

We now provide a model based on our baseline model (without exit-associated price changes) that will give us $p(n, B)$ as an equilibrium outcome. Suppose that the *n* firms compete a la Cournot with homogeneous goods. Suppose that the relational contract is in place (where each firm agrees to pay *B* should a consumer communicate a quality decline). Then $p = a - s\Delta + s(B - C) - bQ$ where Q is industry output. In this case, under a Cournot equilibrium it can be shown that:

$$p(n, B) = \frac{a + cn + s(B(n + 1) - C - \Delta)}{n + 1}$$

Note that as *B* is a pure transfer, it does not impact on prices nor profits. Given (*) in the paper, this means that a relational contract equilibrium will hold if:

$$\delta\left(\frac{a + c - s\Delta}{(n + 1)(1 - \delta) - \delta s}\right) \geq C$$

As expected as *n* increases, the left hand side of this inequality falls.

Note that allowing voice strictly reduces welfare. Prices are higher but consumer surplus is lower as are profits. The reason is that mitigation (*B*) is a transfer while communication cost (*C*) is a cost. For there to be welfare improvements, the firm must actually be induced to do something of value.

Given this, what we need to check is whether the firm has a long-term incentive to keep to the relational contract. That is, this analysis checks if the firm is willing to honor a relational contract with individual customers but is it willing to honor it with all of them? Suppose that a firm decides to deviate and publicly promise not to respond to complaints while the remaining *n* – 1 firms continue to uphold the relational contract. In this case, the deviating firm ends up with an equilibrium price of:

$$\tilde{p} = \frac{a + cn + s(C(n - 1) - \Delta)}{n + 1}$$

while the remaining firms end up with an equilibrium price of:

$$p(n, B) = \frac{a + cn + s(B(1 + n) - 2C - \Delta)}{n + 1}$$

Note that $\tilde{p} < p(n, B) \Longrightarrow C < B$. Let's now compare the profits of the deviating firm to its profits in the full relational contracting equilibrium. The profits from deviating will be lower if:

$$B \geq \frac{n}{n + 1} C\left(2 - \frac{nsC}{a - c + s((n - 1)C - \Delta)}\right)$$

For *n* large, this becomes $C \leq B$ while for $n = 2$, this becomes:

$$B \geq \frac{4}{3}\left(\frac{a-c-s\Delta}{a-c+s(C-\Delta)}\right)C$$

If $B = C$, then this becomes:

$$\frac{a-c-s\Delta+s(n-1)C}{2(a-c-s\Delta)+s(n-2)C} \geq \frac{n}{n+1}$$

For n low, this holds but for n very large it does not hold.

In summary, for n low, it is not worthwhile for a firm to deviate and refuse to acknowledge voice.