Exit, Tweets, and Loyalty

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

July 2016

At the heart of economics is the belief that markets discipline firms for poor performance. However, in his famous book *Exit, Voice, and Loyalty*, Albert Hirschman highlights an alternative mechanism that has received considerably less attention from economists: voice. Hirschman argues that, rather than withdrawing demand from a firm, consumers may instead choose to communicate their dissatisfaction to the firm. In this paper, we develop a formal model of voice and empirically investigate the prevalence as voice as a response to quality deterioration and its relationship with market structure. Our model conceptualizes voice as the equilibrium of a relational contract between firms and consumers and predicts that voice is more likely to emerge in concentrated markets, thus resolving a key source of ambiguity in Hirschman's original formulation. Empirically, we study voice by combining data on tweets about major U.S. airlines with data on airlines' daily on-time performance and local airline market structure. Our analysis shows that the quantity of tweets increases in response to a deterioration in on-time performance and that this relationship is stronger when an airline has a greater share of flights in a given market. We also find that airlines are more likely to respond to tweets in these markets. Our findings indicate that voice is indeed an important mechanism that consumers use to respond to quality deterioration and that its use varies with opportunities for exit.

Keywords: exit voice and loyalty, complaints, airlines, Twitter, social media

JEL Classification: L1, D4, L86

---

* Rotman School of Management, University of Toronto and NBER (Gans and Goldfarb). Xinlong Li, Dan Haid, and Trevor Snider provided excellent research assistance. We gratefully acknowledge financial support from SSHRC (grant # 493140). The paper benefited from helpful comments from Judy Chevalier, Isaac Dinner, and seminar participants at the University of Toronto, UC-Berkeley, the University of Minnesota, the University of North Carolina, Ebay, the 2016 ASSA meetings, the University of British Columbia, Harvard University, and Stanford University.

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 exists but 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). Since exit may be particularly costly in this situation (involving moving or a switch to private school), Hirschman argues that many parents may choose voice. While there exists 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 effective if backed by the threat of exit, the use of voice might fall 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 impact. Thus, the equilibrium relationship between market structure and the use of voice is ambiguous.

---

[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.

To resolve this ambiguity, we model the interactions between consumers and the firm as a relational contract in which consumers use voice to alert firms to quality deteriorations in exchange for a "concession." As in Hirschman's original formulation, in our model, greater competition both makes exit more attractive and voice more effective because it is backed by the threat of exit. However, 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 the 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, it is difficult to isolate the relationship between market structure and voice because market structure may influence quality, which, in turn, influences 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 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 fact that 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 – also varies within markets and often for reasons outside an airline's control. Moreover, on-time performance can be precisely measured. We exploit daily variation in an airline's on-time performance in a given market to estimate the underlying relationship between quality and voice (as measured by daily tweet volume) as well as how this relationship varies with market structure. Thus, our empirical strategy allows us to control for the direct relationship between market structure and quality using airline-airport fixed effects and estimate the relationship between quality, voice and market structure by exploiting *within* market variation in quality across days.

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 over four million tweets to or about a major U.S. airline. For many of these tweets, we are able to 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 measures of the amount of voice directed at a given airline on a given day from consumers in a given market. We then combine these measures with data from the U.S. Department of Transportation (DOT) on the on-time performance of every domestic flight and data from the Official Airlines Guide (OAG) 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 the 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 the airline's on-time performance in that market deteriorates. This result is robust to the inclusion of airline-airport, airline-day and city-day fixed effects, to alternative ways of matching tweets to locations and to 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 poor quality via voice.

Second, we find that the relationship between quality deterioration and tweet volume is strongest when the offending airline dominates an airport. In our setting, a consumer's decision to exit after experiencing poor quality is effectively a decision to choose a different airline for future flights. While it is not possible for us to know when and to where a consumer's future travel will be or identify which airlines serve those routes, we can capture the likelihood of a consumer being able to use an alternative airline in the future using the airport or city share of the airline the consumer tweeted about.[2] Thus, our empirical specification assumes that it is more difficult and/or less desirable for a consumer to exit from an airline that operates a greater fraction of flights in their home market. Our regression results indicate that, controlling for the absolute size of the airline at the airport, the same deterioration in quality generates at least 50% more voice when an airline is the dominant carrier in the market than when it is not. This suggests that, while market dominance may undermine the threat of exit, voice is nevertheless more likely to emerge in equilibrium in concentrated markets, consistent with the prediction from our model.

Finally, the results of our analysis of airline responses to tweets are consistent with the relational contracting model that 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 has a dominant share of local flights 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). Still, over 20% of tweets get responses in our data 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 individuals are more likely to tweet to an airline again if an airline responds to the first tweet 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).

[2] It is well-established in the airline literature that airport-level market power translates into route-level market power (see, for example, Borenstein (1989, 1991) for early evidence on this).

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) provide evidence that firms can discover product deterioration via voice, by studying evidence of vehicle defects that arises through social media.[3] The most closely related research to our work is Beard, Macher, and Mayo (2015). This paper also studies customer complaints using the lens of *Exit, Voice and Loyalty*. It examines complaints to the U.S. Federal Communications Commission about telecommunications companies, estimating the relationship between complaints and market structure, while controlling for consumer perceptions of quality. The authors find that markets that are more competitive are 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 important contributions. First, we provide systematic evidence that consumers do indeed exercise voice in response to quality deterioration and that voice is more likely to be used by consumers in concentrated markets. While we are not able to directly measure the impact of voice on quality, our findings suggest that voice may serve as a mechanism for consumers to discipline firms in concentrated markets. Second, our model of voice as the equilibrium of a relational contract between consumers and firms provides a novel conceptualization of voice and customer complaints. Relational contracts have used in a variety of settings to model behavior when there is a long-term value of maintaining a relationship between parties. Typically, these relationships operate within or between firms. We show that relationships between individual consumers and firms can be modelled as a relational contract and that voice (or complaints) can serve the important role of alerting firms when the level of quality or service provided falls below the expected level. Finally, the empirical strategy we develop - which exploits high-frequency changes in quality within a market rather than differences in quality or

[3] 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)).

market structure across markets - may offer a fruitful way for future researchers to study the use of voice 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.[4] 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 consumer choices of exit and voice and competition, to date there exists no formal model

[4] 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.

of that relationship, in particular, for variation in concentration among oligopolists. Here, we blend the third important aspect of Hirschman's work – loyalty – to provide that model. Specifically, 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 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, say, 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 point 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 is related to 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.[5] 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 involving a consumer threat to exit that 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

[5] 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.

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*.

Note, importantly, that the *cost of a consumer choosing exit for a firm is increasing in market concentration* (i.e., with a fall in *n*). The intuition is that, when market concentration is low, the firm earns high margins from each consumer and faces larger costs should the consumer exit. Thus, 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 is 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.[6] 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 turn to 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:

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

[6] One can imagine situations where there will be an impact on the price a consumer faces if they commit not to consider one supplier. The idea is that this pricing comes from some sort of search model so the consumer ends up facing higher prices when removing a firm from its consideration list. Of course, this is not an innocuous assumption in that after a single period the consumer has no incentive to return to the original firm but would have an incentive if that firm were the only option around. Thus, the exit option may only be exercised for a single period and, after that, the original firm may be in the consideration set. However, that firm would still involve lower consumer surplus as the relational contract would not hold and the consumer would not be compensated by the firm. If Δ were small, however, this firm may have a significant role still.

$$B \geq C$$

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:[7]

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

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

[7] 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.

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 $p(n, C) \rightarrow c + sC$. 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.

2.2 Discussion and Implications for Empirical Analysis

The key insight from our model is that consumers consider the likelihood that a firm will care to retain them, rather than let them exit, and this drives their decision to delay exit in favour of voice. This predicts that voice, and firm responses to voice, are more likely to be observed when the long-term value of a customer is greater. In our model, the long-term value of customers varies with market structure because, in more concentrated markets, margins are higher. This relationship forms the basis of our empirical strategy: we estimate whether a given quality decline generates more voice when the offending firm is the dominant provider in the market.

In our setting, there is an additional mechanism through which market structure may impact the long-term value of customers. It has long been recognized that frequent flier programs can be most effectively used by dominant airlines as they offer consumers the greatest opportunities for points accumulation and redemption (Borenstein, 1989 and 1991). Lederman (2007) and (2008) show that frequent flier programs can lead to increased demand and higher fares for dominant airlines. This “loyalty” increases the long-term value of customers in these markets and, in the context of our model, will increase the likelihood that airlines will want to retain these customers. Note that the observation that many airlines offer exclusive phone numbers for “elite” members of their frequent flier program is consistent with airlines trying to facilitate voice among their most valuable customers. While we do not have data on which tweets to an airline are from members of its frequent flier program, we explore whether this dimension of customer of value is relevant by estimating whether airlines are more likely to respond to tweets from customers who live in their hub cities and to tweets that reference its frequent flier program.

While our model has focused on one motivation for voice there are other factors that may influence voice. For example, a consumer may simply get utility from complaining in the face of adversity (i.e., *C* may be low or negative) or a consumer may leverage a more public complaint because she believes a public complaint may increase the firm's incentives to respond. In this situation, and consistent with our model, we would expect the publicness of a complaint to lead to a stronger response when the value of the consumers who might exit is greater. In addition, while our model has focussed on the industrial organization drivers of voice, it is also possible that firms will wish to encourage voice as a means of identifying quality declines. 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 were 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). While our empirical analysis will not be able to rule out that these other motivations for voice might also operate, we will carry out several additional analyses that suggest a strong role for the relational contracting model.

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 take Twitter as a platform for facilitating and measuring voice and use the data to try to understand the interaction between voice and market power generally.

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). Many airlines have employees dedicated to responding to customers through social media. In addition, Twitter itself has realized this 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).

Finally, while the structure of Twitter now allows for private communication (or direct messages) between Twitter members who do not follow one another, in the period that we analyze, 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. Thus, all communications from consumers to the airlines on Twitter are public and will be included in our data. By contrast, it is possible, and probable, that some airline responses to consumers are done privately (even if via Twitter) and will not appear in our data.

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 of 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 a large number of distinct local markets.

Each airport (or city) has its own market structure and configuration of airlines. This means that the opportunities for exit 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 real.

4.2 Data

Our analysis combines three types of data. The first is data on tweets that are made to or about an airline. This data was purchased from Gnip, a division of Twitter. We combine this with data on airline on-time performance, from the Department of Transportation (DOT), and 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.

[8] A Twitter “handle” is the unique identifier, starting with the “@” symbol, for each participant on Twitter. While 7each tweet is public in the sense that anyone can see it, Twitter users let particular 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.

A large number of these tweets 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 @united_religion or @deltaforce.[10] After eliminating the tweets that were clearly not about airlines, 5,900,691 tweets remained.

The Twitter data includes a large number of 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 so we 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.

To collect data on airline responses to tweets, we created a program that examined called up each of the 4,003,326 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 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.[11]

[10] Not surprisingly, tweets containing the term "united" were the most likely not to be about the airline.

[11] One other issue related to closed and private accounts emerged in the response analysis: Tweets from accounts that were closed or private were coded as not receiving a response. A random sample of 200 of our tweets found 9 such closed and private accounts.

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.[12] The data also contains information on cancelled and diverted flights.

We use these data to construct daily measures of airline's on-time performance in market. There are multiple different ways to measure on-time performance – for example, the number or share of the airline's flights that were 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 will be highly correlated with each other.

As our main measure of on-time performance, we calculate the share of an airline's flights from a given airport on a given day that depart more than 15 minutes late or are cancelled. For multi-airport cities, we calculate the share of an airline's flights from any of the airports in the city that depart more than 15 minutes late or are cancelled. 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 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

[12] 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.

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 three categories of airport dominance: having less than 30% of the flights from the airport, having between 30% and 50% of the flights from the airport, having 50% or more of the flights from the airport. We interpret an airline's share of flights as a measure of how easy or difficult it would be for a consumer to avoid (or exit from) that airline on subsequent flights. Note that a larger share of flights not only means that an airline is more likely to serve many of the routes out of the market but it is also well established that airport-level dominance often confers route-level benefits to airlines. For example, airlines that are dominant at a particular airport will often offer greater frequency service, operate larger planes (if the airport is a hub to that airline), and offer a more valuable frequent flyer program for travelers in that city. Thus, an airline's airport share captures both the feasibility and desirability of exiting from that airline.[13] We construct analogous measures of dominance at the city level for multi-airport cities.

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 as the opportunities for exit (measured by airport dominance) change. 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. Note that we are not

---

[13] There are a number of 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.

able to match individual tweets to particular flights. However, 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 as a consumer's scope for exit changes, 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". We also identify the subset of these tweets that occurred within two miles of an airport. Given the size of tarmac and runways at major airports, we interpret these as tweets that occurred at the airport itself. These tweets make up 51% of tweets with geocode stamps and we label these "Geocode tweets within two miles of an airport".

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 435,972 tweets (based on the latter two measures of location) and city-level information for 1,434,127 tweets (based on all three measures of location). As a check on the reliability of the different location measures, we examine the 199,412 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 199,412 tweets, the city and airport locations match 47.1% of the time. As a benchmark, if the measures perfectly captured the correct city and airport, we might expect them to match 50% of the time because of return trips. 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 98 airports in the airline-airport-day sample and 84 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).[14] 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 388,215 observations while the final airline-city-day dataset contains 334,919 observations.

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,

[14] 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. We also exclude all observations from the day of and days around Super Storm Sandy when delays and cancellations were widespread but few people were likely to be tweeting about airlines.

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.[15] 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. There is a spike in tweet volume in April 2014. This spike is largely driven by two tweets that generated a large number of comments and replies on Twitter: the first is a tweet from US Airways' official Twitter account that contained an explicit image and the second is a tweet from a Dutch teenager threatening a terrorist attack on American Airlines. Figure 1b shows the number of daily tweets by airline and confirms that the spike observed in April 2014 is driven by American Airlines and US Airways. It also 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, 4.5 tweets per airline-city- day compared to 0.95 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 49% of airline-city combinations, the airline operates less than 15% of flights from the city. For about 34% 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. In both datasets, about 20% of an airline's flights at an airport or in a city are delayed more than 15 minutes or cancelled on a given day, with most of these being delayed not cancelled (only about 1.9% of flights are cancelled on average in our data).

[15] 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.

Note that, 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.

For the city-level data, we also construct a number of variables to capture the content and sentiment of the tweets received. These are tweet-level characteristics that we aggregate to the airline-city-day level and they serve as more nuanced and details measures of voice. 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 to the airline whereas tweets about the airline are not. On average, an airline receives 2.9 tweets to its handle, on a given day from consumers associated with a given city. In addition, we construct variables that measure the number of tweets that mention on-time performance and the number of tweets that mention frequent flier programs. The means of these variables are, respectively, 0.74 and 0.24.[16]

Finally, 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.[17] The algorithm then identifies the probability the :) or :(symbol appears, given the appearance of the various word pairs ("bi-grams"). For example, the word pair ("again", "cancel") appears disproportionately

[16] 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.

[17] 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 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.

often with ":(" and the word pair ("great", "service") appears more often with ":)". Overall, the most negative word pair in our data is ("awful", "customer") and the most positive is ("add", "everyone"). 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. However, the algorithm does a better job with tweets that score very positive (below 0.1) or very negative (above 0.9). Therefore, we focus on distinguishing very positive and very negative tweets, rather than average score. We construct a variable that captures the average sentiment of the tweets an airline receives on a given city-day as well as variables that measure the number of very positive and number of very negative tweets received. On average, across airline-city-days, airlines receive 1.85 very positive tweets and 0.94 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 voice being used in response to quality deterioration. Fourth, we carry out a number of 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 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.

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 include airline-location fixed effects, which will control for factors that influence the average amount of voice that an airline receives from consumers in a particular market. Importantly, these fixed effects will capture the overall scale of an airline's operations at an airport or in a city, which is likely to impact the amount of voice generated. These fixed effects 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 the diffusion of Twitter during our sample period in a very flexible way, allowing the diffusion rate to differ across locations. Almost all of our analyses are estimated using a linear regression and we show robustness to a fixed effects Poisson specification.

Some of our evidence in support of specific theoretical model examines which tweets generate airline responses. This analysis is necessarily more speculative because we do not observe all types of responses. In particular, we only observe public tweets by the airlines on Twitter. Airlines can respond in other ways: Quality improvements, email, and direct messaging.[18] Airlines publicly respond on Twitter to 21% of tweets.

[18] At the time of our data, direct message was only possible to Twitter users who were following the user. Therefore, the initial tweets that we use to measure voice likely capture all voice on Twitter: The airlines are unlikely to follow the users who complain before the complaint occurs. However, the responses can be through direct messages if the users follow the airline. In this way, our measures of tweets to the airline capture most of the voice exercised through Twitter. Our measures of responses by the airline only capture of subset of such responses: The subset that the airline choose to keep public either because the user did not follow the airline or because the airline wanted the response to be publicly visible.

6. Results

a. *Motivating Analysis*

Before turning to the regression analysis, Table 4 illustrates the core variation that we exploit in our regression analysis. Using location in profile as the location definition, each cell shows the average number of tweets by airline-location-day. The columns show the percentage of delays or cancellations, increasing from left to right. The rows show the share of flights in the city by that airline, increasing from top to bottom. Two clear patterns emerge in Table 4 and both persist in the regression analysis below. First, in general, the number of tweets rises moving from left to right, as on-time performance deteriorates. Second, this increase in the number of tweets as on-time performance deteriorates is most prominent when an airline has a large share of flights in the city. A third pattern in Table 4, that the number of tweets rises moving from top to bottom, as the share of flights increases, is captured in the regressions by the airline-location fixed effects that we include.

b. *Tweets and On-Time Performance*

Table 5 estimates the relationship between tweets and on-time performance. The first row contains the coefficient of interest: the share of the airline's flights in a location that were delayed 15 minutes or more or canceled. If, as hypothesized, tweets are a response to quality deterioration, we would expect the coefficient to be positive. In this table, our main dependent variable is the 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 their opportunities for exit on future trips. We also show robustness to the alternative ways of matching tweets to locations.

Looking at Table 5, it is evident that there is a robust statistical relationship between on-time performance and tweet volume. Across ten different specifications, the point estimate is always positive, statistically significant, and large in magnitude (relative to a baseline of 4.1 tweets per day with location in profile information or 0.6 tweets per day with geocoded tweets). Column 1 includes controls for the number of flights that the airline has at that airport, and separate airline and city fixed effects.[19] As expected, having more flights from a location increases the number of tweets received from consumers from that city. Column 2 adds day-city and airline-city fixed effects. [20] This serves as our main empirical specification for the remainder of the paper.[21] Note that, once we add these fixed effects, the variable capturing the 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 either insignificant or marginally significant but much smaller in magnitude.

The coefficient estimate in column 2 suggests that an increase in the share of delayed or canceled flights of twenty percentage points (roughly one standard deviation) coincides with 0.35 more tweets against an average of 4.13 tweets. Column 3 separates delayed flights and canceled flights and shows that both have a positive coefficient, with cancellations having a much larger impact on tweets than delays. Column 4 shows robustness to associating tweets to locations using any of the three sources of location information. Columns 5 and 6 address potential skew in the data on the number of tweets. Column 5 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. Column 6 shows robustness to a fixed effects Poisson specification.

Columns 7 through 10 use the tweet count measures that are based on matching tweets to airports rather than cities. In column 7, tweets are matched to the airport closest to the user at the time the tweet was made and then aggregated to the airline-airport-day level. Column 8 is similar to column 7 but includes geocoded tweets that were made from a location that is within two miles of an airport. Column 9 only includes tweets that have an airport mentioned in the text of the tweet. Column 10 includes tweets that are either geocoded or that mention an airport. The coefficients in columns 7 to 10 are statistically significant though smaller in magnitude than in the city-level regressions, which is expected given that we can match many more tweets to cities than to airports. Overall, we view Table 5 as clearly indicating that there is a robust statistical relationship between

[19] The airline fixed effects are estimated and the city fixed effects are differenced out from means using Stata's xtreg, fe function.

[20] Here, and in the regressions below, the airline-city fixed effects are estimated, and the day-city fixed effects are differenced out using stata's xtreg, fe command.

[21] In appendix table A1, we show robustness to adding fixed effects for the day-airline, differencing them out using first differences.

tweets and quality deterioration which emerges across various location measures, fixed effect specifications, and functional forms.

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

In order 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 airport dominance. Table 6 estimates columns 2 through 10 of Table 5 with the added interactions. The first row shows the main effect of delays or cancellations, which captures the relationship between tweets and on-time performance when an airline operates less than 30% of the flights in a city. The second row shows the interaction with 30-50% share and the third row shows the interaction with over 50% share. (The exception is column 3 which separates delays and cancellations and shows the general result applies to both). Because we include airline-location fixed effects, the direct effect of dominance is not separately identified.

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 more dominant in a market.

The estimated marginal effects on the interactions in Table 2 are large. Column 1 suggests that, in markets in which an airline is not dominant, when its share of delays or cancelations rises by 20 percentage points (about one standard deviation), the number of tweets it receives that day increases by just 0.23. The average number of tweets across all days for this group is 3.08. In contrast, when an airline operates 30-50% of flights from a city, a one standard deviation increase in the share delayed or canceled flights increases the number of tweets received by 1.13 (1.135/5+4.495/5), compared to an average number of tweets for this group of 8.12. If the airline has >50% of flights, a 20 percentage point increase in the share delayed or canceled leads to 3.23 more tweets compared to an average for this group of 9.49.

Looking at columns 6 through 9, which use the airport-based location information to match tweets to locations, we see a similar pattern. As in Table 1, the marginal effects in terms of number of tweets are smaller for the airport-level analysis because we have more tweets with location information when location is defined at the city level.

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

In this section, we include three sets of additional analyses that assess the credibility of the results. In particular, we confirm that tweets about on-time performance rise when on-time performance deteriorates, that tweets are more negative when on-time performance deteriorates, and that longer delays are associated with more tweets. Together, we view these as suggesting that the increase in tweets is related to comments about quality deterioration, rather than some other reason (for example, a mechanical increase in tweeting because people have time to use Twitter while waiting at the airport).

Table 7 re-estimates column 1 from Table 6 using two alternative dependent variables: the number of tweets that mention on-time performance and the number of tweets that do not. On average, there are 0.74 tweets per airline-location-day that mention on-time performance, leaving a remainder of 3.39 that do not. Despite the much lower overall number of tweets about on-time performance, these tweets have a greater response elasticity to a deterioration in on-time performance. The first two columns of Table 7 show that both tweets about on-time performance and tweets not about on-time performance rise when performance deteriorates; however, given the lower baseline number of tweets about on-time performance, the proportional effect is larger.[22] Comparing columns 3 and 4 shows that as dominance rises, the greater responsiveness of tweets about on-time performance persists.

Table 8 explores tweet sentiment. The dependent variable in columns 1 and 2 is the average negative sentiment of tweets for that airline-location-day. The value is missing when there are no tweets in a data. These columns investigate whether on-time performance impacts the average sentiment of tweets received by an airline at a location-day. We find that the average negative sentiment of the tweets received is higher during delays, with no significant difference as market dominance increases. 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 position tweets increase when on-time performance is worse, with the

[22] An alternative way to communicate this relative increase would be to use the fraction of tweets with on-time performance for a given airline-location-day. We did not use this method because we could not determine a meaningful way to calculate this fraction for days without any tweets.

impact on very negative tweets larger in both absolute and relative terms. 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.

Table 9 considers whether the severity of on-time performance deterioration matters. We compare the responsiveness of tweets to delays of different lengths. Specifically, we include variables capturing the share of an airline's flights delayed by 15 minutes or more, the share of an airline's flights delayed by 60 minutes or more and the share delayed by 120 minutes or more. The delay measures are not mutually exclusive, and so 15-minute delays include 60 and 120 minute delays. This means that the coefficient on delayed 60 minutes, for example, is the increase in tweets for 60-minute delays relative to 15-minute delays. Column 1 shows that tweets increase with the severity of delay. Column 2 shows that, for each level of delay, the impact on tweets is generally larger when an airline is more dominant though we lose some power and have large standard errors for some coefficients.

Overall, the results in this section, though not surprising, indicate that the relationships we have uncovered are indeed evidence of that consumers use 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 support the relational contracting conceptualization of voice proposed above.

i. The model emphasizes the most valuable 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 by more profitable customers. We define profitability 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 a number of 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. 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 airline responses to tweets for all airlines except US Airways. Overall, 21.2% of tweets receive responses. Of tweets that mention the airline's handle, 34.5% receive responses. Figure 2a shows that the fraction of tweets that receive responses grew rapidly until June 2013, and then leveled out. Figure 2b shows that there is considerable variation in response rates by airline, with American being most responsive during this period and Southwest least responsive.

Before proceeding with the airline response analysis, it is important to recognize that there are multiple other ways airlines could respond to tweets, including direct messages and future quality improvements. We are unable to observe either of 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.

In Table 10, we estimate whether airlines are more likely to respond to tweets from their most valuable customers. For this empirical analysis, the level of observation is the tweet and the dependent variable is an indicator variable for whether the airline responded to a tweet. We estimate a logit model. We control for other keyword strings that might elicit an airline response including whether the tweet contains the airline's handle, whether the tweet contains a customer service keyword,[23] and whether the tweet contains an on-time performance keyword. We also 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 the airlines are more likely to respond to tweets from customers who live in a city in which the airline operates more than 30% of flights. We interpret this as indicating that airlines are more likely to respond to customers with a greater future value, either because the margins earned on them are larger and/or because they are more likely to be members of the airline's frequent flier program. The remaining rows show the impact of the other variables: airlines respond more to tweets with negative sentiment, to tweets to the

[23] 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".

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.[24]

Column 2 of the table adds interactions with tweet sentiment and shows no consistent relationship in terms of response rates. Column 3 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 because airline social media managers will have easy access to the tweet content. The location information is harder to find. Again, the result suggests that airlines respond more to tweets with frequent flier keywords. Column 4 adds an interaction between frequent flier keywords and negative sentiment. In this case we find that airlines are particularly likely to respond to tweets that have frequent flier keywords and a negative sentiment. Column 5 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 most 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 as a way 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 airline delays and cancellations. Here, we provide evidence that suggests venting is not the primary motivation for the voice that we observe.

If a tweet is made 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 message about the airline directed to the sender's Twitter followers. Table 11 compares the impact of on-time performance deterioration on tweets to the handle and tweets not to the handle. Columns 1 and 2

[24] 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.

show that when there are delays and cancellation, both tweets to the handle rise and tweets not to the handle rise. The relative increase is roughly proportional to the overall ratio of tweets of each type. 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. We view this as suggesting that the customers want to communicate with the airline.

Table 11, however, does not address the point that even a tweet to an airline's handle is public and therefore the public nature of the tweet might be what is driving the consumer's decision to exercise voice. We explore this in two ways. First, Figure 3 looks at the share of flights delayed to the relevant airline-location-day by number of followers of the twitter user who tweeted. Figure 3 shows no consistent relationship between the fraction delayed and the number of followers. The underlying correlation coefficient between fraction delayed and number of followers is -0.008. We interpret as suggesting that people with more followers do not tweet disproportionately more when there are delays or cancellations.

Second, returning to Table 10, which estimated the airline response models, it is apparent that there is no clear relationship between number of followers and response rates. This indicates that airlines are not disproportionately likely to respond to people with more followers. We interpret Table 11, Figure 3, and the followers results in Table 10 to suggest that tweets about airlines during delays are often communications to the airline, especially in situations in which the airline is dominant.

iii. The model implies responses should lead to future tweets

Finally, in Table 12, we look at whether twitter users who receive a response from an airline are more likely to tweet in the future. Many of the twitter users in our data tweet multiple times to an airline. The 4,003,326 tweets in the data are made by 1,340,734 different users. Of these, 520,807 tweet more than once. The median number of tweets is 1, the 75th percentile is 2, the 99th percentile is 26 and the maximum is 6635.

Table 12 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 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 the results of this section 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 airline response when given by the highest value customers, rather than by the customers that have the most 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 the choice between exit and voice is influenced by market structure. Our model predicts that voice is more likely to emerge in concentrated markets because the value to firms of retaining consumers in concentrated markets 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 exploits that 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. Our empirical results show that consumers use voice to express disappointment with quality and are more likely to use voice when they have fewer opportunities for exit. The observed increase in voice is most pronounced for tweets that mention on-time performance and that are negative in sentiment. Consistent with the relational contracting model proposed, firms appear to respond more often to their most valuable customers, the public nature of a tweet does not seem to be the major driver of the tweets or responses by airlines, and users are more likely to tweet again to an airline if their first tweet received a response.

There are, of course, a number of limitations to this paper, which suggest fruitful avenues for future research in this area. First, our setting does not allow us to investigate how social media has affected the quantity and nature of complaints since we do not observe complaints directed to airlines through a means other than Twitter. Identifying an empirical setting in which one could study voice before-and-after the introduction of social media as a channel for complaints would be interesting. Second, rather than observe exit directly, we infer exit options based on market structure. An interesting avenue for future research would be studying the choice between voice and exit, at the customer level. Third, we do not observe quality changes that result from voice besides responses to tweets and so we cannot directly investigate whether voice leads to quality improvements. This would require a setting in which one could measure equilibrium quality with and without voice. Since voice is always available in our setting, we are unable to explore this. Overall, however, we view our results as indicating that a relational contracting model provides a useful way to think about consumer exercise of voice.

In an interview, Eric Maskin (at the time, the Albert O. Hirschman Professor of Social Science at the Institute for Advanced Study) argued that recent advances in economic theory may lead to a revival of interest in models that relate to *Exit, Voice, and Loyalty* (Adelman 2013, p. 615). In addition, new communication technologies both lower the costs of voice and enable researchers to measure consumer voice. Our research takes advantage of these advances and demonstrates that consumers do indeed use voice to attempt to influence market outcomes, especially when their exit options are limited. Given the link between voice and market power, and the new opportunities provided by digital communication, we believe that this suggests that voice is a fruitful area of future research in industrial organization.

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.

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, R. 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.

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	4,003,326	0.3582	0.4795	0	1
Airport mentioned in tweet	4,003,326	0.0426	0.0426	0	1
Geocode stamp on tweet	4,003,326	0.0715	0.0715	0	1
Any location information	4,003,326	0.4162	0.4162	0	1
Airport in tweet or geocode	4,003,326	0.1089	0.1089	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	0.2562	0.2449	0.2565
Alaska Airlines	0.0284	0.0263	0.0337
JetBlue	0.1177	0.1268	0.1373
Delta Air Lines	0.1258	0.1486	0.1328
United Airlines	0.2438	0.2364	0.2047
US Airways	0.1138	0.1043	0.1077
Southwest Airlines	0.1143	0.1126	0.1276

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

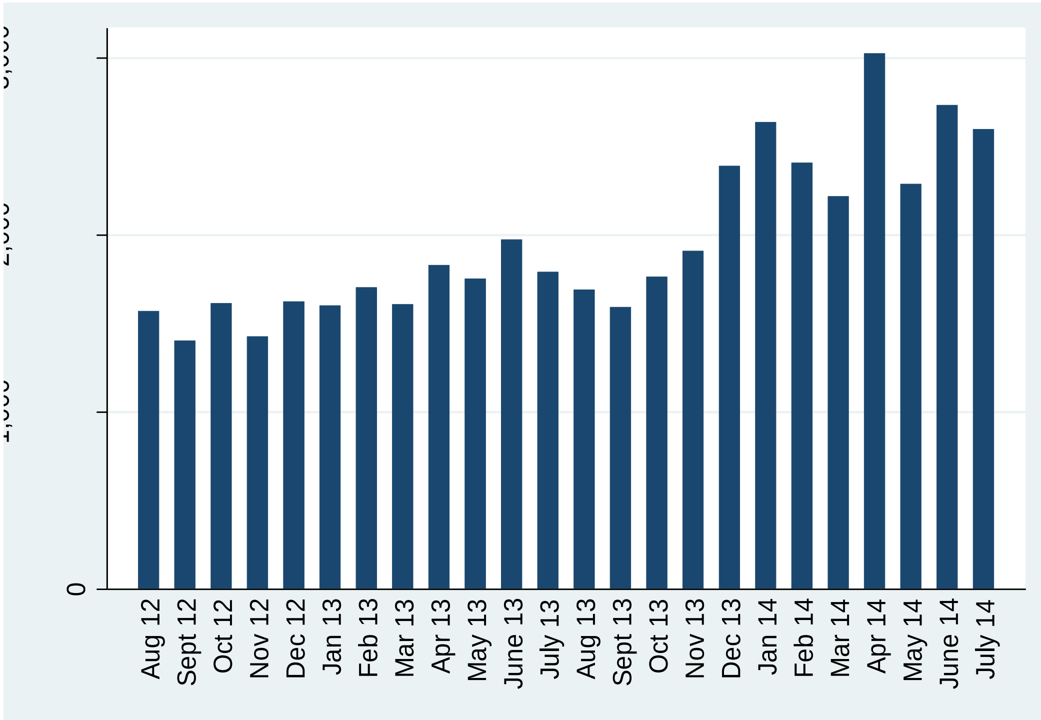

Figure 1b: Average Daily Tweets by Month by Airline (Data with city information)

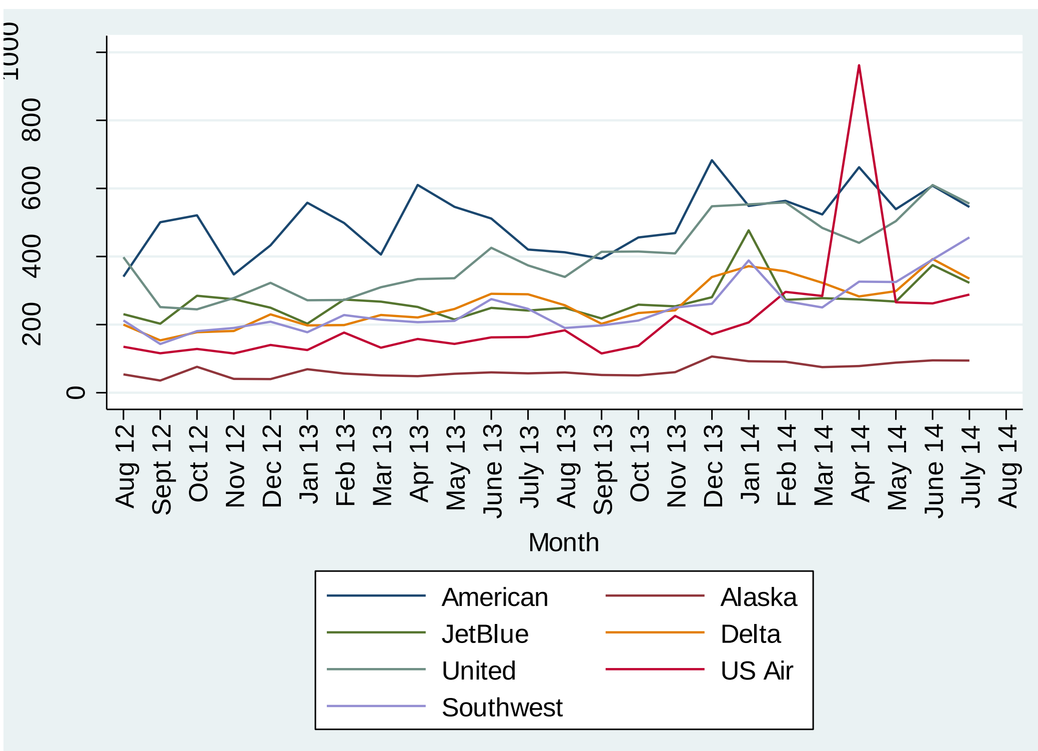

Table 2: Location-Airline-Day Descriptive Statistics

CITY LEVEL DATA					
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Min**	**Max**
# tweets (location given in profile)	334,919	4.1338	13.5282	0	2365
# tweets (any location definition)	334,919	4.4844	14.2791	0	2407
Airline-airport flights/week	334,919	192.6008	435.6203	1	4453
Airline share of flights in city					
Under 15%	334,919	0.4895	0.4999	0	1
15-30%	334,919	0.3404	0.4739	0	1
30-50%	334,919	0.1228	0.3282	0	1
Over 50%	334,919	0.0473	0.2124	0	1
Share delayed					
Departure delay over 15 minutes	334,919	0.1812	0.1938	0	1
Departure delay over 60 minutes	334,919	0.0728	0.1306	0	1
Departure delay over 120 minutes	334,919	0.0392	0.0993	0	1
Departures canceled	334,919	0.0190	0.0728	0	1
Dep. delay > 15 min. or canceled	334,919	0.1966	0.2177	0	1
Tweet content (for location in profile tweets)					
# tweets to handle	334,919	2.8560	9.2281	0	1141
Average sentiment	334,919	0.3571	0.2915	0	1
# tweets mention on time performance	334,919	0.7429	2.8029	0	452
# tweets mention frequent flier program	334,919	0.2434	1.0849	0	126
# very positive tweets	334,919	1.8490	6.3114	0	1145
# very negative tweets	334,919	0.9412	3.6095	0	587
AIRPORT LEVEL DATA					
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Min**	**Max**
# tweets (geocode stamp)	388,215	0.5882	1.8788	0	97
# tweets (geocode within 2 miles)	388,215	0.3712	1.4778	0	70
# tweets (airport mentioned in tweet)	388,215	0.4128	1.8662	0	368
# tweets (geocode or mentioned)	388,215	0.9509	3.0374	0	371
Airline-airport flights/week	388,215	166.0982	393.2921	1	4453
Airline share of flights at airport					
Under 15%	388,215	0.5245	0.4994	0	1
15-30%	388,215	0.3034	0.4597	0	1
30-50%	388,215	0.1001	0.3001	0	1
Over 50%	388,215	0.0720	0.2584	0	1
Share delayed					
Departure delay over 15 minutes	388,215	0.1800	0.2584	0	1
Departure delay over 60 minutes	388,215	0.0722	0.1306	0	1
Departure delay over 120 minutes	388,215	0.0388	0.0990	0	1
Departures canceled	388,215	0.0187	0.0722	0	1
Dep. delay > 15 min. or canceled	388,215	0.1952	0.2180	0	1

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.
wow my @united premier upgrade cleared 3 days before ny flt to hnl! maybe united will grow on me
@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: Average Number of Tweets to Airline per Day, by Dominance and On-Time Performance

Airline share of flights at airports in the city	No delays or cancelations	0.01% - 10% of flights delayed or canceled	10% - 25% of flights delayed or canceled	Over 25% of flights delayed or canceled
Under 15%	1.01	5.83	5.14	4.88
15-30%	0.75	3.52	3.20	3.23
30-50%	0.68	6.42	8.84	11.46
Over 50%	1.01	7.54	10.06	11.78

Location identified as location in profile.

Table 5: Tweets and On-Time Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	City-level location in profile only	City-level location in profile only	City-level location in profile only	City-level all three location measures	Log(City-level location in profile only+1)	FE poisson City-level location in profile only	Airport-level Geocoded tweets	Airport-level Geocoded tweets within 2 miles of airport	Airport-level Airport mentioned in tweet	Airport-level Geocoded and airport mentioned
Share flights delayed or canceled	3.072***	1.732***		1.798***	0.165***	0.635***	0.249***	0.153***	0.288***	0.512***
	(0.822)	(0.422)		(0.340)	(0.015)	(0.007)	(0.040)	(0.030)	(0.058)	(0.084)
Share flights delayed 15 minutes or more			0.845***							
			(0.186)							
Share flights canceled			5.501**							
			(1.654)							
Airline flights departing that location	0.216***	0.036	0.037	0.007*	0.003	0.002***	0.040+	0.030	0.008	0.046+
	(0.035)	(0.037)	(0.037)	(0.002)	(0.003)	(0.000)	(0.023)	(0.020)	(0.009)	(0.027)
Fixed effects	Airline, Location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location
Avg. value dep. var.	4.13	4.13	4.13	4.48	4.13	4.13	0.59	0.37	0.41	0.95
N	334919	334919	334919	334919	334919	280806	388215	388215	388215	388215
R-sq/LL	0.123	0.357	0.358	0.383	0.444		0.409	0.458	0.319	0.455

Dependent variable identified in column headers. Unit of observation is the location-airline-day. In columns 1-6, location is defined by city. In columns 7-10, 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 6: Tweets, On-Time Performance, and Market Dominance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	City-level location in profile only	City-level all three location measures	City-level location in profile only	Log (City-level location in profile only +1)	FE Poisson City-level location in profile only	Airport-level Geocoded tweets	Airport-level Geocoded tweets within 2 miles of airport	Airport-level Airport mentioned in tweet	Airport-level Geocoded and airport mentioned
Share flights delayed or canceled	1.135***	1.062***		0.151***	0.619***	0.079**	0.023	0.093***	0.162***
	(0.321)	(0.279)		(0.014)	(0.007)	(0.026)	(0.016)	(0.025)	(0.043)
Share Delayed or Canceled x Airline 30-50% share city	4.495*	5.609**		0.091***	0.034***	0.871*	0.593*	1.120+	1.914*
	(1.857)	(2.191)		(0.026)	(0.010)	(0.360)	(0.292)	(0.617)	(0.897)
Share Delayed or Canceled x Airline >50% share city	15.023**	16.941***		0.445***	0.311***	4.360***	3.524***	4.649***	8.592***
	(5.612)	(5.761)		(0.072)	(0.020)	(0.729)	(0.635)	(1.198)	(1.449)
Airline flights departing that airport	0.026	0.007***	0.029	0.003	0.002***	0.034	0.025	0.001	0.035
	(0.037)	(0.002)	(0.037)	(0.003)	(0.000)	(0.021)	(0.019)	(0.009)	(0.025)
Share flights delayed 15 minutes or more			0.380**						
			(0.142)						
Share Delayed 15 minutes x Airline 30-50% share city			4.473**						
			(1.443)						
Share Delayed 15 minutes x Airline >50% share city			4.091**						
			(1.324)						
Share flights canceled			7.506***						
			(1.955)						
Share Canceled x Airline 30-50% share city			5.516						
			(4.464)						
Share Canceled x Airline >50% share city			48.367*						
			(23.006)						
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	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location
Average value dep. var.	4.13	4.48	4.13	4.13	4.13	0.59	0.37	0.41	0.95
N	334919	334919	334919	334919	280806	388215	388215	388215	388215
R-sq/LL	0.360	0.386	0.361	0.444		0.420	0.469	0.332	0.471

Dependent variable identified in column headers. Unit of observation is the location-airline-day. In columns 1-5, location is defined by city. In columns 6-9, location is defined by airport. Robust standard errors clustered by location 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 7: On-Time Performance Mentioned in Tweet

	(1)	(3)	(2)	(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
Share flights delayed or canceled	0.740***	0.923***	0.483***	0.550**
	(0.160)	(0.211)	(0.113)	(0.162)
Share delayed or canceled x Airline 30-50% share city			2.020*	2.873*
			(0.814)	(1.122)
Share delayed or canceled x Airline >50% share city			5.494*	8.388**
			(2.097)	(2.978)
Airline flights departing that location	0.001***	0.004**	0.001**	0.004*
	(0.000)	(0.001)	(0.000)	(0.001)
Fixed effects	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location
Average value dep. var.	0.74	3.39	0.74	3.39
N	334919	334919	334919	334919
R-sq	0.283	0.335	0.290	0.336

Dependent variable identified in column headers. Unit of observation is the location-airline-day. Location defined by city in profile. Robust standard errors clustered by location in parentheses. Airline-location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. Day-airline fixed effects are differenced out with first differences. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

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
Share flights delayed or canceled	0.093***	0.092***	0.892***	0.208***	0.620***	0.096+
	(0.006)	(0.006)	(0.217)	(0.058)	(0.166)	(0.048)
Share Delayed or Canceled x Airline 30-50% share city		0.009			2.069*	0.857*
		(0.013)			(0.879)	(0.358)
Share Delayed or Canceled x Airline >50% share city		0.020			6.728*	2.780**
		(0.016)			(2.658)	(0.906)
Airline flights departing that location	-0.000	-0.000+	0.002***	0.002*	0.002***	0.002*
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
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
Average value dep. var.	0.36	0.36	0.94	1.85	0.94	1.85
N	179484	179484	334919	334919	334919	334919
R-sq	0.079	0.079	0.275	0.335	0.282	0.335

Dependent variable identified in column headers. Unit of observation is the location-airline-day. Location defined by city in profile. Robust standard errors clustered by location in parentheses. Airline- location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. Day-airline fixed effects are differenced out with first differences. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table 9: By length of delay

	(1)	(2)
Share flights delayed 15 minutes	0.343**	-0.005
	(0.112)	(0.123)
Share delayed 15 minutes x Airline 30-50% share city		2.950**
		(0.906)
Share delayed 15 minutes x Airline >50% share city		5.677*
		(2.218)
Share flights delayed 60 minutes	0.860***	0.638**
	(0.246)	(0.224)
Share Delayed 60 minutes x Airline 30-50% share city		1.659+
		(0.969)
Share Delayed 60 minutes x Airline >50% share city		2.518
		(2.207)
Share flights delayed 120 minutes	1.406**	1.265***
	(0.430)	(0.353)
Share Delayed 120 minutes x Airline 30-50% share city		2.547
		(2.266)
Share Delayed 120 minutes x Airline >50% share city		3.080
		(9.190)
Share flights canceled	3.667**	2.887*
	(1.312)	(1.138)
Share Canceled x Airline 30-50% share city		2.273
		(2.853)
Share Canceled x Airline >50% share city		44.411+
		(26.449)
Airline flights departing that airport	0.006**	0.006**
	(0.002)	(0.002)
Fixed effects	Day-location, Airline-location	Day-location, Airline-location
Average value dep. var.	4.13	4.13
N	334919	334919
R-sq	0.358	0.361

Dependent variable is city-level tweets with the location in profile known. Unit of observation is the location-airline-day. Location is defined by city. Robust standard errors clustered by location 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

Figure 2a: Response rates over time

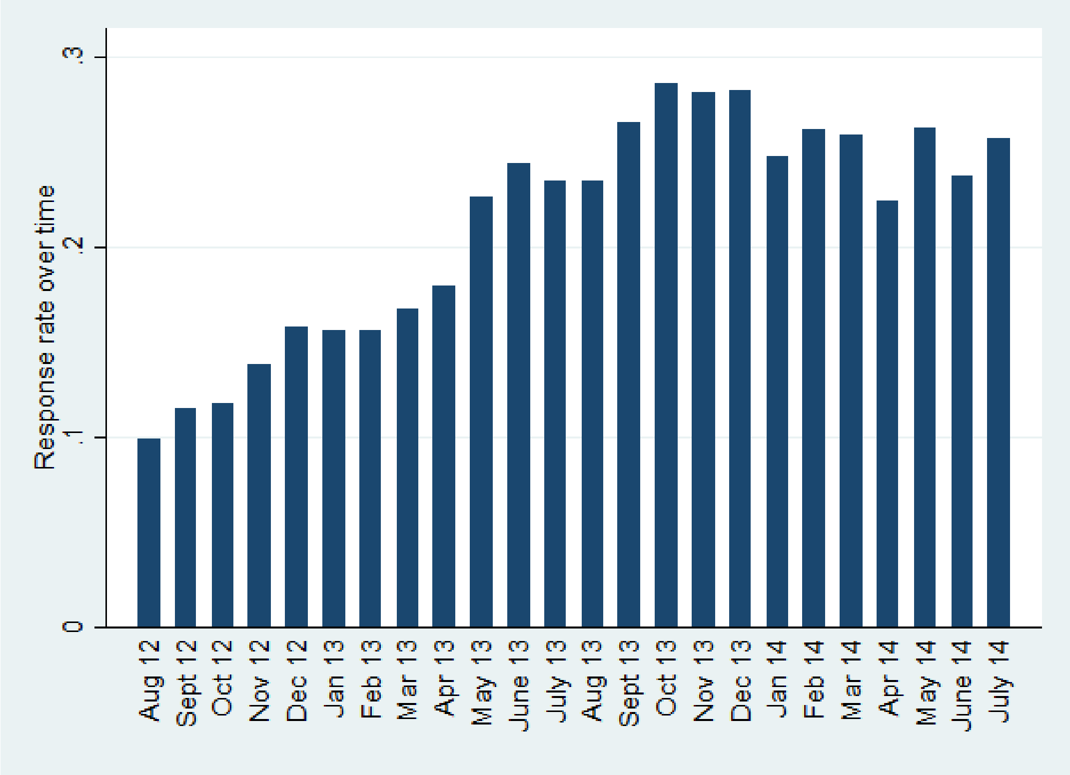

Figure 2b: Response rates by airline over time

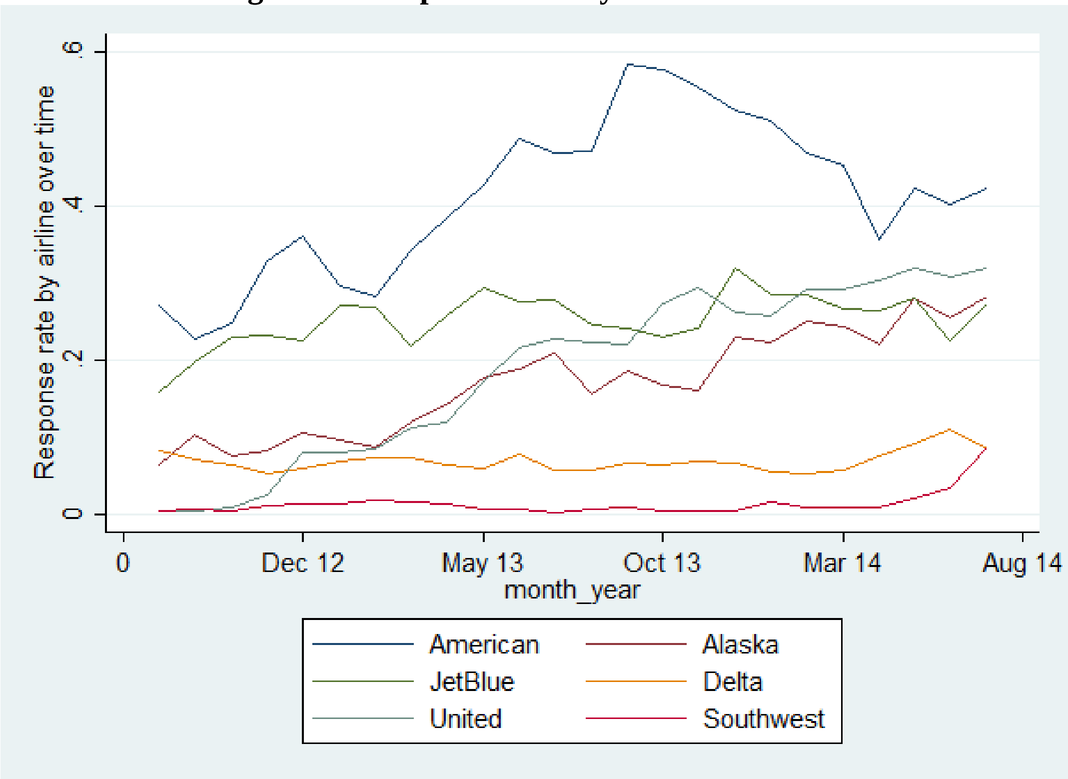

Table 10: Response Rates, Delays, and Dominance

	(1)	(2)	(3)	(4)	(5)
Airline 30-50% share city	0.248***	0.268***			0.246***
	(0.009)	(0.011)			(0.009)
Airline >50% share city	0.186***	0.183***			0.183***
	(0.014)	(0.016)			(0.014)
Frequent flier keyword			0.274***	0.164***	0.270***
			(0.027)	(0.040)	(0.027)
Probability sentiment is negative	0.039*	0.044*	0.052**	0.031+	0.054**
	(0.017)	(0.018)	(0.017)	(0.017)	(0.017)
Frequent flier keyword x				0.355***	
Probability sentiment is negative				(0.054)	
Airline 30-50% share city x		-0.046**			
Probability sentiment is negative		(0.016)			
Airline >50% share city x		0.007			
Probability sentiment is negative		(0.024)			
Number of followers, 25th -50th	0.035***	0.036***	0.051***	0.051***	0.037***
percentile	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Number of followers, 50th -75th	-0.063***	-0.063***	-0.043***	-0.044***	-0.061***
percentile	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Number of followers, 75th -99th	-0.129***	-0.129***	-0.105***	-0.106***	-0.128***
percentile	(0.014)	(0.014)	(0.013)	(0.013)	(0.014)
Number of followers, over 99th	0.136***	0.136***	0.155***	0.154***	0.138***
percentile	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Handle	3.135***	3.134***	3.144***	3.144***	3.130***
	(0.035)	(0.035)	(0.034)	(0.034)	(0.035)
Customer service keyword	0.402***	0.402***	0.409***	0.408***	0.408***
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)
On time performance keyword	0.493***	0.493***	0.501***	0.503***	0.497***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
American Airlines	3.999***	3.999***	3.973***	3.973***	3.993***
	(0.072)	(0.072)	(0.072)	(0.072)	(0.072)
Alaska Airlines	2.640***	2.640***	2.650***	2.648***	2.638***
	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)
JetBlue	3.367***	3.367***	3.349***	3.348***	3.370***
	(0.073)	(0.073)	(0.073)	(0.073)	(0.073)
Delta Air Lines	1.403***	1.403***	1.390***	1.389***	1.387***
	(0.070)	(0.070)	(0.070)	(0.070)	(0.071)
United Airlines	2.822***	2.822***	2.821***	2.821***	2.806***
	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)
Date	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	3,548,933	3,548,933	3,548,933	3,548,933	3,548,933
Log Likelihood	-1,253,521.8	-1,253,514.0	-1,253,999.3	-1,253,766.9	-1,252,636.5

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: Handles

	(1)	(2)	(3)	(4)
	Number tweets to handle	Number tweets not to handle	Number tweets to handle	Number tweets not to handle
Share flights delayed or canceled	1.221***	0.507***	0.704***	0.430***
	(0.294)	(0.132)	(0.201)	(0.125)
Share Delayed or Canceled x Airline 30-50% share city			3.979*	0.510
			(1.609)	(0.324)
Share Delayed or Canceled x Airline >50% share city			12.411*	2.465**
			(4.732)	(0.920)
Airline flights departing that location that week	0.005**	0.001	0.004**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Fixed effects	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location	Day-location, Airline-location
Average value Dep. Var	2.86	1.28	2.86	1.28
N	334919	334919	334919	334919
R-sq	0.421	0.164	0.425	0.164

Dependent variable identified in column headers. Unit of observation is the location-airline-day. Location defined by city in profile. Robust standard errors clustered by location in parentheses. Airline- location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. Day-airline fixed effects are differenced out with first differences. +p<0.10, *p<0.05, **p<0.01, ***p<0.001

Figure 3: Share flights delayed or cancelled when tweet, by number of followers

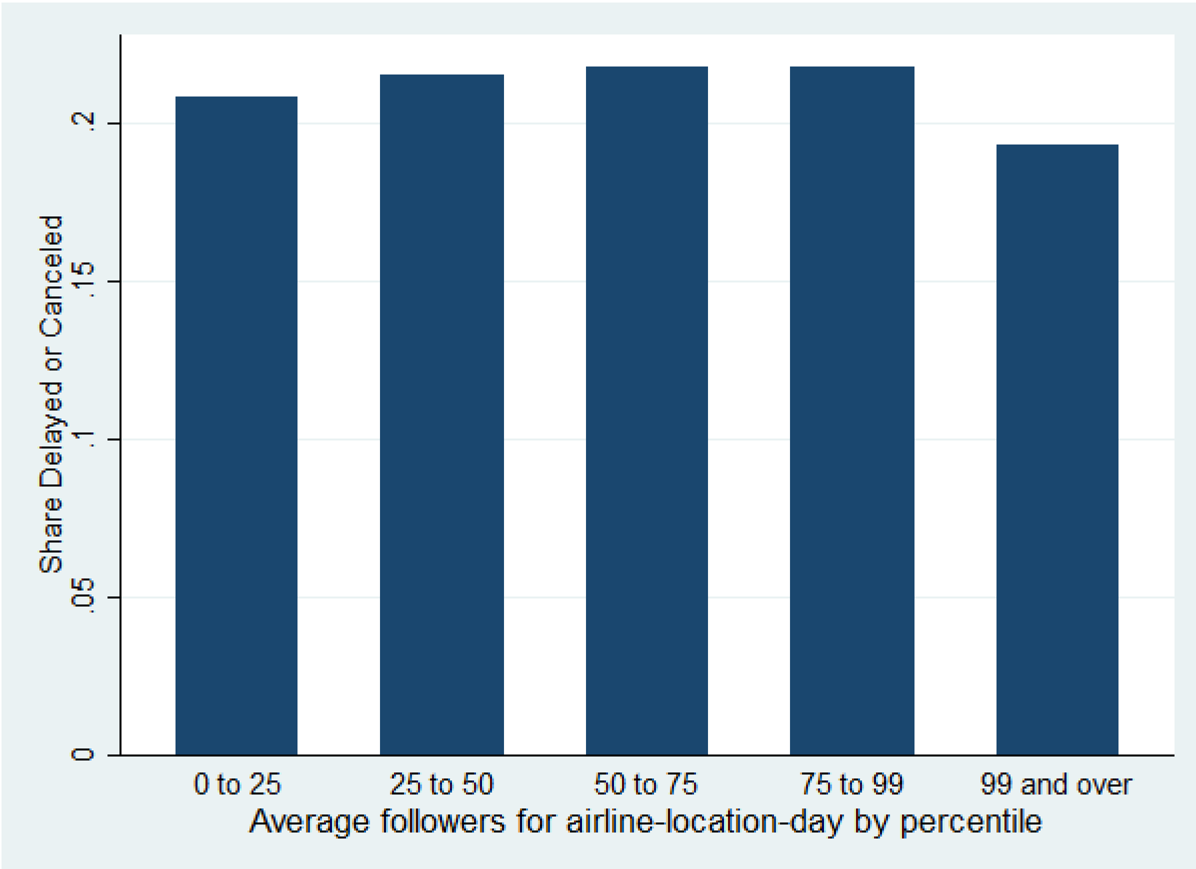

Table 12: Repeat tweeters and airline responses

	Tweet again, after first tweet		Tweet in 2013 or 2014, given first tweet in 2012	
	(1)	(2)	(3)	(4)
Airline responded to first tweet	0.988***	0.801***	0.497***	0.269***
	(0.005)	(0.006)	(0.014)	(0.016)
Frequent flier keyword		0.147***		0.385***
		(0.010)		(0.020)
Airline 30-50% share city		0.254***		0.370***
		(0.007)		(0.014)
Airline >50% share city		0.258***		0.442***
		(0.012)		(0.023)
Probability sentiment is negative		0.138***		-0.097***
		(0.005)		(0.011)
Number of followers, 25^{th} -50^{th} percentile		0.089***		0.373***
		(0.005)		(0.011)
Number of followers, 50^{th} -75^{th} percentile		0.258***		0.666***
		(0.005)		(0.011)
Number of followers, 75^{th} -99^{th} percentile		0.597***		1.139***
		(0.006)		(0.013)
Number of followers, over 99^{th} percentile		0.857***		1.739***
		(0.033)		(0.064)
Handle		0.652***		0.525***
		(0.004)		(0.009)
Customer service keyword		0.132***		0.091***
		(0.007)		(0.014)
On time performance keyword		0.140***		0.071***
		(0.006)		(0.012)
American Airlines		0.126***		-0.118***
		(0.007)		(0.015)
Alaska Airlines		0.134***		0.110***
		(0.013)		(0.027)
JetBlue		0.196***		-0.052***
		(0.008)		(0.016)
Delta Air Lines		-0.124***		-0.170***
		(0.007)		(0.016)
United Airlines		0.095***		-0.080***
		(0.007)		(0.014)
Date		-0.002***		-0.003***
		(0.000009)		(0.000009)
Constant		39.083***	-0.451***	52.671***
		(0.182)	(0.004)	(1.736)
N	1,216,865	1,216,865	268,968	268,968
Log Likelihood	-790,462.9	-746,682.2	-180,364.0	-171,736.8

Dependent variable in columns 1 and 2 is whether tweeted again. Dependent variable in columns 3 and 4 is whether tweeted 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

Table A1: Descriptive Statistics for Response Data

Variable	Obs.	Mean	Std. Dev.	Min	Max
Airline replied	3,548,933	0.2122	0.4089	0	1
Airline replied if tweet to airline handle	2,074,393	0.3451	0.4754	0	1
Frequent flier keyword	3,548,933	0.0533	0.2246	0	1
Airline 30-50% share city	3,548,933	0.0951	0.2933	0	1
Airline >50% share city	3,548,933	0.0364	0.1873	0	1
Probability sentiment is negative	3,548,933	0.3614	0.3954	0	1
Number of followers, 25^{th} -50^{th} percentile	3,548,933	0.2518	0.4340	0	1
Number of followers, 50^{th} -75^{th} percentile	3,548,933	0.2506	0.4334	0	1
Number of followers, 75^{th} -99^{th} percentile	3,548,933	0.2384	0.4261	0	1
Number of followers, over 99^{th} percentile	3,548,933	0.0101	0.0999	0	1
Handle	3,548,933	0.5845	0.4928	0	1
Customer service keyword	3,548,933	0.1036	0.3048	0	1
On time performance keyword	3,548,933	0.1580	0.3647	0	1
American Airlines	3,548,933	0.2890	0.4533	0	1
Alaska Airlines	3,548,933	0.0320	0.1761	0	1
JetBlue	3,548,933	0.1329	0.3395	0	1
Delta Air Lines	3,548,933	0.1420	0.3490	0	1
United Airlines	3,548,933	0.2751	0.4466	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	1,216,865	0.380105	0.485413	0	1
Airline replied	1,216,865	0.141916	0.348964	0	1
Frequent flier keyword	1,216,865	0.04214	0.200909	0	1
Airline 30-50% share city	1,216,865	0.077724	0.267737	0	1
Airline >50% share city	1,216,865	0.028162	0.165435	0	1
Probability sentiment is negative	1,216,865	0.361716	0.39428	0	1
Number of followers, 25^{th} -50^{th} percentile	1,216,865	0.295047	0.456064	0	1
Number of followers, 50^{th} -75^{th} percentile	1,216,865	0.229437	0.420471	0	1
Number of followers, 75^{th} -99^{th} percentile	1,216,865	0.137377	0.344245	0	1
Number of followers, over 99^{th} percentile	1,216,865	0.003519	0.059216	0	1
Handle	1,216,865	0.490423	0.499909	0	1
Customer service keyword	1,216,865	0.098778	0.298363	0	1
On time performance keyword	1,216,865	0.155583	0.362459	0	1
American Airlines	1,216,865	0.251919	0.434115	0	1
Alaska Airlines	1,216,865	0.02847	0.166311	0	1
JetBlue	1,216,865	0.12947	0.33572	0	1
Delta Air Lines	1,216,865	0.170376	0.375963	0	1
United Airlines	1,216,865	0.273254	0.44563	0	1
First tweet for 2012 tweets					
Tweeted in 2013 or 2014	268,968	0.4001	0.4899	0	1
Airline replied	268,968	0.0888	0.2844	0	1
Frequent flier keyword	268,968	0.0406	0.1973	0	1
Airline 30-50% share city	268,968	0.0883	0.2838	0	1
Airline >50% share city	268,968	0.0313	0.1742	0	1
Probability sentiment is negative	268,968	0.3536	0.3922	0	1
Number of followers, 25^{th} -50^{th} percentile	268,968	0.2998	0.4582	0	1
Number of followers, 50^{th} -75^{th} percentile	268,968	0.2357	0.4245	0	1
Number of followers, 75^{th} -99^{th} percentile	268,968	0.1653	0.3715	0	1
Number of followers, over 99^{th} percentile	268,968	0.0046	0.0677	0	1
Handle	268,968	0.3949	0.4888	0	1
Customer service keyword	268,968	0.0928	0.2902	0	1
On time performance keyword	268,968	0.1522	0.3592	0	1
American Airlines	268,968	0.2549	0.4358	0	1
Alaska Airlines	268,968	0.0270	0.1621	0	1
JetBlue	268,968	0.1603	0.3669	0	1
Delta Air Lines	268,968	0.1540	0.3610	0	1
United Airlines	268,968	0.2811	0.4495	0	1

Table A3: Fixed effects for day-airline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	City-level location in profile only	City-level location in profile only	City-level all three location measures	Airport-level Geocoded tweets	Airport-level Geocoded tweets	Airport-level Geocoded tweets and airport mentioned in tweet	Airport-level Airport mentioned in tweet
Share flights delayed or canceled	1.724***	1.179**	0.119***	0.283***	0.111***	0.200***	0.096**
	(0.442)	(0.410)	(0.029)	(0.048)	(0.032)	(0.052)	(0.032)
Share Delayed or Canceled x Airline 30-50% share city		4.153+	0.485+		0.769*	1.896+	1.207+
		(2.274)	(0.289)		(0.340)	(0.973)	(0.708)
Share Delayed or Canceled x Airline >50% share city		14.263***	4.624***		5.392***	9.996***	5.075***
		(3.815)	(1.182)		(0.859)	(1.444)	(0.881)
Airline flights departing that airport	0.055	0.044	0.016+	0.048**	0.036*	0.039*	0.004
	(0.038)	(0.038)	(0.008)	(0.018)	(0.016)	(0.019)	(0.007)
Fixed effects	Day-location, Airline-location, Day-airline	Day-location, Airline-location, Day-airline	Day-location, Airline-location, Day-airline	Day-location, Airline-location, Day-airline	Day-location, Airline-location, Day-airline	Day-location, Airline-location, Day-airline	Day-location, Airline-location, Day-airline
Average value dep. var.	4.13	4.13		0.59	0.59	0.41	0.95
N	269271	269271	269271	310930	310930	310930	310930
R-sq	0.372	0.373	0.418	0.425	0.439	0.469	0.292

Dependent variable identified in column headers. Unit of observation is the location-airline-day. In columns 1-3, location is defined by city. In columns 4-7, location is defined by airport. Robust standard errors clustered by location in parentheses. Airline-location fixed effects are estimated directly. Day-location fixed effects are differenced out using stata's xtreg, fe command. Day-airline fixed effects are differenced out with first differences. +p<0.10, *p<0.05, **p<0.01, ***p<0.001