Content and Context: Identifying the Impact of Qualitative Information on Consumer Choice

Completed Research Paper

Sinan Aral
NYU Stern
KMC 8-81 / 44 W. 4th St.
New York, NY 10012
sinan@stern.nyu.edu

Panos Ipeirotis
NYU Stern
KMC 8-84 / 44 W. 4th St.
New York, NY 10012
panos@stern.nyu.edu

Sean J. Taylor
NYU Stern
KMC 8-186A / 44 W. 4th St.
New York, NY 10012
staylor@stern.nyu.edu

Abstract

Managers and researchers alike suspect that the vast amounts of qualitative information in blogs, reviews, news stories, and experts' advice influence consumer behavior. But, does qualitative information impact or rather reflect consumer choices? We argue that because message content and consumer choice are endogenous, non-random selection and conflation of awareness and persuasion complicate causal estimation of the impact of message content on outcomes. We apply Latent Dirichlet Allocation to characterize the topics of transcribed content from 2,397 stock recommendations provided by Jim Cramer on his show Mad Money. We demonstrate that selection bias and audience prior awareness create measurable biases in estimates of the impact of content on stock prices. Comparing recommendation content to prior news, we show that he is less persuasive when he uses more novel arguments. The technique we develop can be applied in a variety of settings where marketers can present different messages depending on what consumers know.

Keywords: Econometric analyses, Text mining, Identification, Qualitative information

Introduction

Managers and researchers alike believe that the vast amounts of qualitative information available to consumers in online and offline environments impacts their decision making. Research has shown that the content of real estate listings affect which properties sell (Levitt and Syverson 2008); that feedback about sellers in electronic markets affects their future sales (Ghose et al. 2006); and that user-generated product reviews help determine which products consumers purchase online (Archak et al. 2007). But do these sources of qualitative information impact or rather reflect consumers' choices? We argue that, in most economic contexts, message content and consumer choices are endogenously determined. Real estate agents, bloggers, analysts and pundits know their audience and know what their audience wants to hear. Persuasive messages often target particular individuals or are sought by those who need the information most. We show that the endogeneity of content creation and consumption creates measurable bias in estimates of the change in consumer behavior caused by exposure to qualitative messages. Our estimates suggest this bias can be detected and mitigated by modeling the process that determines which individuals receive which information.

We further address another source of potential bias in estimates: omitting measures of prior attention which are correlated with the selection process or content variables. Content producers may choose to discuss or recommend based on the prior attention of their audiences, and the message they deliver may be affected by how aware their audiences are. Insufficient control for prior attention creates additional bias in estimates of the economic impact of message content because changes in consumer choice are wrongly attributed to the content of the message.

Both sources of endogeneity—selection effects and omitted prior attention—are rooted in consumers' prior knowledge. Individuals consuming qualitative content are not blank slates; rather, they read or listen to messages in the context of what they already know. After estimating biases due to endogeneity problems, we explore how exposure to prior information mediates the influence of qualitative content on consumers' economic decisions. In particular, we devise methods to estimate the quantitative impact of topical dimensions of information content on consumer choice while controlling for selection and attention as sources of endogeneity.

Our empirical context is the investment advice offered by Jim Cramer on his CNBC show *Mad Money*. Cramer's recommendations have been shown to influence market prices although he does not support his analysis with new news or quantitative data (Engelberg et al. 2010). Our estimates suggest that the topical content of Cramer's advice helps explain his price effect after accounting for his role in drawing attention to stocks and the bias introduced by his selection process. Further, the extent of his impact is affected by the prior knowledge of his audience—when Cramer supports his recommendations with words or phrases that have themselves recently been mentioned in the news, he has a smaller effect on changes in stock prices.

Our methodology attempts to isolate the economic impact of persuasive messages and to characterize what makes particular information and content influential by incorporating consumers' prior knowledge into estimates of the effect of message content on consumer behavior. We believe our findings suggest a new approach for marketers which will be enabled by information technology: understanding what consumers already know can help devise or select more persuasive messages. The technique we develop can therefore be put to practical use in a variety of settings where managers and marketers can present subjects with different messages depending on what they already know.

Theory

The Impact of Qualitative Information on Consumer Choice

Consumers are increasingly turning to qualitative information for expressive descriptions and quality assessments for the products and services they are considering purchasing. At the same time, researchers and firms are interested in drawing inferences about how such information affects consumer behavior. For instance, Das and Chen (2007) show that sentiment measures derived from the content of postings in finance message boards are correlated with stock prices, trading volumes, and volatility. Similar studies have demonstrated relationships between economic outcomes and measures derived from news (Tetlock

2007), Tweets (O'Connor, et al. 2010), and postings on Internet message boards (Antweiler and Frank 2004). Archak et al. (2007) econometrically model the impact of product reviews on consumer choice, arguing the textual review content significantly affects demand for products. Ghose et al. (2006) show that qualitative feedback in online marketplaces provides a signal of seller characteristics beyond numerical ratings, and that buyers pay attention to these signals when they decide from whom to purchase.

Despite the established predictive power of qualitative information, it is difficult to establish causal relationships between message content and consumer choices because review or feedback text may merely reflect unobserved attributes of products or sellers that affect message content and consumer choices simultaneously. Where previous work helped establish that textual content is predictive of economic outcomes, we point out some of the identification problems which could confound causal interpretation of these types of results. A causal interpretation is a prerequisite for making effective decisions about which qualitative information to present consumers.

Our analysis also extends work on the influence of textual information on consumer choice by relaxing the assumption that consumers process information *tabula rasa*. Accounting for the interaction between message content and information context has proven important in past research. For example, Berger et al. (2010) argue that the effect of negative publicity on a product's sales depends on how well individuals know the product. When awareness is low, they find that negative reviews actually increase sales by informing consumers of a product's existence. Our analysis extends this line of work by considering different, qualitative dimensions of prior knowledge.

Selection Bias in Content Production

Selection bias is likely to exist in content production across a broad set of contexts. For example, books recommended on a blog may undergo a substantial upswing in sales after recommendations are posted. But this effect may be explained in part by the fact that the blogger is more likely to reviews books that are gaining in popularity because she is more likely to hear about these types of books from her friends. Without a model of the blogger's book selection process, it is difficult to make causal inferences about the effect of her recommendations on book sales.

Non-random selection may also imply non-random content production—what is said about a product once a content producer chooses to talk about it. For example, a product's popularity may affect what content providers say about the product. To expand on the example, the blogger may write longer reviews when she hears about the book from more friends. The more of her friends who know about a book, the more likely it is to become popular after her review, and we may falsely attribute greater sales to the length of the blogger's coverage. This example demonstrates why attributing causal interpretations to the effects of message content (such as length) on sales (or popularity or prices) is problematic—the content production process itself may be driven by the same factors that affect the outcomes of interest.

Research examining the kinds of information firms report has shown that profit-seeking can entail the production of biased information. For example, Gentzkow and Shapiro (2006) construct a model in which firms report news that falsely conforms to the prior expectations of their consumers in order to build a reputation for quality. Anand et al. (2007) derive a model which, assuming some facts are unverifiable, explains media bias as an equilibrium in a horizontally differentiated product market. Neither of these models directly addresses the question of how an information provider chooses what to talk about in order to maximize ratings or profits. In our particular context, it is possible to draw conclusions about which stocks it may be profitable for Cramer to recommend.

First, Cramer may select stocks that are trending in popularity or price momentum. Applying Gentzkow and Shapiro's argument, profit-maximizing content producers are likely to recommend subjects about which their audiences are optimistic. In the *Mad Money* context, recent returns or volume may either be an indication or cause of the audience's interest, and thus correlated with recommendations. As there is evidence of significant serial correlation in stock returns (Lo and MacKinlay 1998), if Cramer selects stocks with positive serial correlation that have recently trended upward in price, then this selection process will tend to bias measures of his effect on prices upward.

Second, information that affects the valuations of stocks is likely to diffuse non-uniformly and slowly through the consumer population (Merton 1987). Media sources compete by providing timely access to information about recent events or by accomplishing the costly process of deriving the salient implications of relevant information. All things equal, people prefer receiving higher quality information faster and media sources which process and deliver salient information faster than their competition are likely to thrive. Cramer's selections on *Mad Money* should therefore provide information which has not yet diffused through other sources and contains insights that have not been deduced elsewhere. In this case, Cramer would not be causing the price changes, but instead accelerating a price change which would have happened anyway.

Each of these explanations imply that, because he has a profit-maximizing motive to acquire and retain a loyal audience, Cramer will tend to recommend stocks that are more likely to experience positive overnight returns with or without his recommendation. We therefore propose Hypothesis 1: *Accounting for selection bias will reduce the estimated magnitude of the effect of Cramer's recommendations on the overnight returns of recommended stocks.*

Prior Awareness Bias

We conceptually separate the impact of Cramer's recommendation on a stock's price into an attentional component and a persuasive component. The former relies on the mechanism of alerting audience members to the existence of the stock, while the later relies on Cramer convincing them that a stock they already know about is undervalued. This distinction parallels earlier thinking on the economic effects of advertising in the marketing literature (e.g. Mitra and Lynch 1995).

Consumers are likely to have different levels of prior awareness for the stocks Cramer discusses, and this may affect price response because the more of his viewers who know about the stock, the fewer will need to buy it to follow his advice. Merton (1987) formalized this explanation for the effect of drawing attention to stocks in his "investor recognition hypothesis." In his model, stocks appreciate in value as more investors learn about their existence and subsequently purchase shares to properly diversify. If level of prior awareness is correlated with Cramer's selection process or with measures derived from content, then omitting controls for prior awareness may introduce an additional bias into estimates of the effect of qualitative information on price changes. Accordingly, we propose Hypothesis 2: *Stocks for which there is more investor attention before Cramer's recommendation should experience a smaller increase in overnight returns after Cramer's recommendation. If Cramer's selection or content is correlated with prior attention, then controlling for prior attention should significantly affect estimates of the effect of content on price changes.*

The Role of Prior Knowledge

Recipients of Cramer's investment advice probably also consult other sources of news to inform their trading decisions. In addition to being aware of a stock and considering it as an investment, individuals may have some prior knowledge, which they have used to form impressions of the stock's prospects. Allowing for this possibility, the effect of Cramer's recommendations is likely to be mediated by what his audience already knows. When the relevant state is the value of a single unknown random variable, modeling how a rational actor combines prior information with newly acquired data is an application of Bayesian statistics.[1] Standard results show that the change in expected value, from prior to posterior, is increasing in difference between the sample mean and the prior mean, and also in the sample size.

In the realm of qualitative information, we can conceptualize the novelty of the information as the distance between the sample mean and the prior mean (Aral and Van Alstyne 2011). The length of the text serves as a proxy for the sample size. The comparative statics of the Bayesian update derived above then become very intuitive hypotheses about how prior information and new data may interact to determine the size of Cramer's price effect. Hypothesis 3: *Recommendations which contain information that is more novel (content that is different from the audience's prior information) and a greater volume of*

[1] In the case of qualitative, linguistic information, the sample space is considerably more complex. Archak et al. (2007) posit that consumers have a set of prior distributions over a set of choice-relevant feature dimensions to model how textual content in reviews affects consumer choice. By using semi-supervised learning to identify ex-ante what these dimensions are and then assigning feature-modifier pairs to them, they are able to model the consumer's Bayesian updating.

novel information, should experience a greater increase in overnight returns after Cramer's recommendation.

Empirical Methods

Jim Cramer's Mad Money

CNBC's *Mad Money* began airing on March 14th, 2005 and remains a popular part of the network's lineup. Engelberg et al. (2010) report mean ratings from a 123 episode sample cover about 213,000 households, ranging from 73,000 to 304,000. The households in their sample are high-income, with almost three quarters of the mean households reporting income over $60,000 per year. Engelberg et al. (2010) observe an average overnight abnormal return exceeding 3% for their sample of 388 first-time buy recommendations. The large returns indicate substantial price changes generated from increased demand for the stocks Cramer discusses.

Mad Money presents a unique opportunity for investigating persuasion resulting from the consumption of qualitative information. Each of Cramer's recommendations discretely refers to a particular stock and is associated with a well-defined economic action: buying or selling the stock in the near future. Since the show occurs after trading, we can observe the aggregate response to Cramer's recommendations by measuring changes in stock prices in the opening price of the stock the next day.

A second order effect of Jim Cramer's show is the stream of research it has spawned in an attempt to characterize the magnitude and nature of his effect on financial markets (Engelberg et al. 2010; Neumann and Kenny 2007; Lim and Rosario 2008; Karniouchina et al. 2009). As far as we know, we are the first to econometrically identify the magnitude of Cramer's influence. Other studies have estimated pick returns adjusted for risk, but none have addressed identification. We are also the first to conduct a content analysis of Cramer's speech. Karniouchina et al. (2009) employ a similar data set to examine some message factors such as primacy and recency, but do not analyze the textual content.

Data

We analyze transcripts of Cramer's spoken content provided by the website www.madmoneyrecap.com, which records verbatim records of Cramer's recommendation content. We watched a random sample of transcribed shows and found the accuracy of the transcriptions to be very high. The transcripts for each show are segmented into comments about a particular stock, either that Cramer has chosen to discuss or about which a caller has inquired (during the "lightning round" of the show). We call these segments "recommendation events," which is our unit of analysis. We consider a Cramer-initiated recommendation to buy a stock as the treatment, matching each recommendation to the overnight return for the stock.

Our sample is censored in a few important ways. First, we only keep recommendations containing at least 20 words of text after stop words and tickers are removed. This ensures that we only analyze cases where message content is likely to be important. Second, we consider only "buy" recommendations. "Buy" recommendations outnumber "sell" recommendations by about 2-1 in the sample and "sell" outcomes may be systematically different due to higher transaction costs of short-selling. Third, our sample is limited to stocks which are listed in the CRSP and Computstat databases and we require return data that covers a 155 day estimation window leading up to the recommendation to estimate pricing models. In practice, this omits newly listed stocks and stocks traded on foreign exchanges. 21% of our observations are censored as a result of this limitation.

The resulting data set consists of 2397 recommendation events for 850 distinct stocks occurring during 612 episodes of *Mad Money* from 11/7/2005 to 11/7/2008. We call these events treated and create a set of counterfactual events for stock-episode pairs where Cramer did not make a recommendation of any kind, using all episode dates and mentioned stocks plus stocks in the S&P 1500 as of 1/1/2006 not mentioned on the show, yielding 2,076 stocks. This is an impractically large set of events, so we construct a stratified sample where 1/32 events are treated cases and the remainder are a random sample of counterfactuals. Table 1 reports summary statistics for selected variables in both treated and untreated subsamples. We observe a 1% mean overnight return for treated events, with median overnight and daily return of about 0.5%. This is compared to a mean and median of 0% for both overnight and daily returns in the untreated cases.

Our analysis proceeds in four steps. First, we address the endogeneity of selection for treatment and outcome by modeling Cramer's selection process using a Heckman (1979) sample selection model to adjust our naive estimates of the treatment effect. Second, we use Latent Dirichlet Allocation (LDA) (Blei et al. 2003) to estimate a set of topics and allocate the terms of each recommendation across these topics. Third, we consider how Cramer's role in drawing attention to stocks may bias estimates of the coefficients of topic variables. Finally, we compute a measure of information novelty using recent financial news about each stock and show this measure interacts with the length of recommendations to explain overnight returns. The next four sections describe these methods and results of associated analyses in detail.

Table 1: Summary Statistics			Percentiles				
	Mean	Std. Dev.	0%	25%	50%	75%	100%
Treated Events (N=2397)							
Overnight Return	0.94	2.981	-66.401	-0.139	0.426	1.379	28.539
Return	0.837	3.718	-27.448	-0.953	0.507	2.248	35.759
Stock Price	55.492	64.743	1.95	23.42	40.79	66.05	842
Market Cap.	29796	53709	45	1778	9210	30988	504240
Lightning Round	0.464						
Number Words	67.645	76.238	20	31	39	59	758
News Words	335.662	394.49	0	0	202	535	2842
Untreated Events (N=74307)							
Overnight Return	0.01	2.008	-112.072	-0.39	0	0.439	111.474
Return	-0.024	3.147	-40.231	-1.248	0	1.168	160.333
Stock Price	36.605	34.295	0.161	17.99	30.84	47.2	797
Market Cap.	9408	27230	29	811	2059	6780	504240

Modeling Selection Bias in Content Production

We are interested in estimating the effect of Cramer's advice on the value of stocks, as measured by the overnight return[2]—the change in the stocks' price from the close of trading on the day preceding the show to the opening price the following day, $OR_{it} = \log(\frac{open_{it+1}}{close_{it}})$. Let c_{it} be a set of measures associated with recommendation text, where i indexes the stock and t indexes the show date. The overnight return may also depend on some attributes of the stock which also depend on time, which we call p_{it}, and a set of measures of prior attention a_{it}. A simple linear specification could be used to model this:

$$OR_{it} = \alpha + c'_{it}\beta_c + p'_{it}\beta_p + a'_{it}\beta_a + \epsilon_{it} \qquad [1]$$

The key assumption that allows unbiased estimation of these coefficients is that $E[\epsilon_{it}|c_{it}, p_{it}, a_{it}] = 0$. The problem is that we only observe c_{it} for events where Cramer recommends the stock, so our estimation is conditional on Cramer deciding to discuss a stock, meaning we now require that $E[\epsilon_{it}|c_{it}, p_{it}, a_{it}, Cramer\ discusses\ stock\ i\ at\ t] = 0$. This assumption is more questionable because Cramer may have some insight into what will happen to the stock in overnight trading or he may select stocks which performed well today and for which there is some autocorrelation in performance. In

[2] We use overnight return as our dependent variable for a number of reasons. First, the more time that elapses between the event and the measured price effect, the more likely that there will be intervening events which could confound estimation. Second, price effects with shorter time horizons are better proxies for Cramer's effects on consumer's choices, while measuring effects in longer horizons may be better proxies for the quality of his recommendations, which we are less interested in here. Third, the longer the time horizon, the more noise there is in estimates of Cramer's effect. We also conducted our analysis with one day abnormal returns as the dependent variable with qualitatively similar results.

either case, the conditional expectation of ϵ_{it} is unlikely to be zero, necessitating some further assumptions.

If we can model how Cramer chooses to recommend a stock, we can characterize the distribution of ϵ_{it} in our sample and construct a variable which adjusts for the bias of our estimates of β_c (Heckman 1979). Let z_{it} be a vector of data about the stock and the market conditions available prior to the airing of *Mad Money*. Cramer's selection process may be modeled as a Probit:

$$d_{it}^{*} = z_{it}'\gamma + \nu_{it};\ d_{it} = 1 \text{ if } d_{it}^{*} > 0; 0 \text{ otherwise},$$

where d_{it} is an indicator describing whether stock i is discussed at time t and ν_{it} is normally distributed. By further assuming that ϵ_{it} and ν_{it} are bivariate normal with covariance matrix $\Sigma = \begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon\nu} \\ \sigma_{\epsilon\nu} & \sigma_\nu^2 \end{pmatrix}$, we can simultaneously estimate the selection and outcome models using maximum likelihood. As long as our parametric assumptions hold, the estimates are consistent. We can identify $\beta_c, \beta_p, \beta_a$ as long as we have a valid exclusion restriction—information about the selection process which plausibly does not affect the outcome (discussed below). Identification also requires that we normalize the covariance matrix so that $\sigma_\nu = 1$. We can then test the null hypothesis of no sample selection bias using a likelihood ratio (LR) test on the restriction that the coefficient $\rho = \frac{\sigma_{\epsilon\nu}}{\sigma_\epsilon \sigma_\nu} = 0$.

Our ability to reliably correct for non-random selection depends on how accurately we can model the process through which a stock is mentioned on *Mad Money*. Accordingly, we use a variety of data in the vector z_{it} to reflect stock and time-varying characteristics. The predictors in the selection equation are a superset of the stock characteristics and measures of attention, including variables for stock price, the market capitalization of the company, and the coefficients from a Fama-French four-factor pricing model to concisely account for a stock's risk exposure. Cramer's selection may also be based on trends in news coverage, trading activity, or investor attention. Therefore, we include five lags for each of the following: 1) financial data (returns, abnormal returns, log trading volume, and normalized log trading volume); 2) Google search volume indices for the stock's ticker; and 3) sentiment indicators derived from news about the stock, weighted by volume (as measured by number of articles). The financial data accounts for prior investor interest in purchasing the stock, the Google search volume data accounts for prior investor interest in researching the company or stock, and text based sentiment indicators derived from recent prior news about the stock account for attention the stock has received from the financial media. For reference, Table 2 provides a summary of the data used in the selection model.

Table 2: Summary of Variables Used

Variable	Source	Use
Stock Returns	CRSP	Selection
Abnormal Returns	Fama-French 4-Factor Residuals[3]	Selection
Fama-French betas	Fama-French model coefficiens	Stock Controls
Trade Volume	CRSP	Selection/Attention
Market Cap.	Compustat	Selection/Attention
Search Volume	Google Insights	Selection/Attention
Sentiment Indexes	Reuters NewsScope Sentiment	Selection
Mad Money Text	http://madmoneyrecap.com	Content
News Content	Reuters NewsScope Archive	Prior Knowledge

[3] Abnormal returns are measured using a Fama-French four-factor pricing model to adjust returns for exposure to price changes in four risk-factor portfolios (Fama and French 1992), estimated over the period [t-155,t-5) where t is the day of the recommendation date. Estimated abnormal returns adjust for common risk factors, and the coefficients from the pricing model may be interpreted as quantifying the stock's exposure to different types of market risk.

The sample selection model is identified because the selection equation includes financial variables which should not affect future price changes of the selected stocks. Recent stock returns and trade volumes should not affect future prices because the current price incorporates their information. The lagged financial data included in the selection equation therefore constitutes a valid exclusion restriction which identifies the coefficients of interest in the outcome model. Additionally, even if the exclusion restriction on lags of stock returns is invalidated by serial correlation in returns, lags of trading volume remain valid.

We use three measures of investor attention in the selection model: normalized log trading volume, Google search volume indices, and weighted news volume allocated into positive, negative, and neutral sentiment. Trading volume is normalized by the mean and standard deviation of the stock's log trading volume over the same 150 day pre-event window. Google search and news measures are described below.

A Generative Model of the Topical Content of Stock Recommendations

We use a vector-space model to numerically represent Cramer's recommendation text. Documents are encoded as vectors whose value at an index is the frequency of a particular term. Implicitly, this representation makes a word exchangeability assumption, meaning we do not consider two documents with the same terms but ordered differently to be different.

We preprocess the text by removing annotations; tokenizing the sentences into distinct terms; lemmatizing terms into their roots; and removing stop words using a standard list. The result of this process is a set of vectors k_{it} of length W, the number of unique terms in our sample. We compute term counts $L_{it} = \sum_{w=1}^{W} k_{itw}$, where k_{itw} is the number of terms at index w for recommendation event.

If arguments can vary in their persuasiveness for reasons other than the number of terms, the variance must be due to the distribution of the terms in the recommendation. We use Latent Dirichlet Allocation (LDA) (Blei et al. 2003), an automated machine learning procedure for content coding. LDA is an unsupervised machine learning technique which assumes each document is created as a mixture of core topics, each topic having its own distinct term distribution. LDA estimation simultaneously infers the topic distributions (multinomial over terms) and allocates terms across these topics. We sum these by topic terms to compute a topic measure for each recommendation,

$$TOPIC_{itk} = \sum_{j=1}^{L_i} 1\left(Z_{itj} = k\right),\ k = 1, \ldots, 20;$$

where 1 is one if the statement is true and zero otherwise. The topic variables measure the amount of time Cramer has dedicated to various topics. While our ultimate task is to explain the variation in overnight returns from Cramer's recommendations, it is worthwhile to examine if we are able to draw some qualitative conclusions about topical sources of Cramer's influence by summarizing our estimated topic distributions. We do so by examining the terms with the highest likelihood ratios for appearing in the topic distribution as compared with the topic.

Table 3 shows a sample of the terms with the highest likelihood ratios for each of our 20 topics and the labels we create based on them. Some clear categories emerge from the representative terms, indicated by the presence of descriptive keywords. Topics range from advice based on trading strategies, for instance based on recent growth (Topic 4), mergers (Topic 17), or companies' management (Topics 12, 16), to recommendations based on industry plays (e.g. Shipping–Topic 1; Oil and Gas–Topic 2; Retail–Topic 20).

We have constructed 20 topic variables based on numbers of words related to a particular topic appearing in the recommendation text. If we sum each of these 20 variables, the result is equal to the total number of words. Since LDA-estimated topic allocations are a higher-dimensional projection of the recommendation text, including provides a nested model and we can test their joint significance using a Likelihood ratio test.

Measuring the Prior Awareness of Consumers

If the content of Cramer's message is correlated with the prior awareness of his audience and prior awareness affects stock prices (i.e. $E[a_{it}\epsilon_{it}] > 0$ and $E[a_{it}' c_{it}] > 0$), then estimates of the effects of message factors will be biased unless we control for prior awareness. To address this potential omitted variable bias, we use four variables to control for investor knowledge of the recommended stocks. First, since stocks which are traded more frequently are likely to have greater investor awareness, we include the log of total trading volume for the stock for the previous five trading days. Second, we include a dummy variable which indicates whether Cramer has discussed this stock on a previous *Mad Money* episode. If Cramer has mentioned the stock to his audience before, then they are more likely to know about it and less likely to be learning about it for the first time.

Table 3: Representative Words for LDA Topics

	Category	Representative Words
1	Shipping	railroad, engine, chassis, auto, truck
2	Oil and Gas	gas, oil, energy, drilling, crude
3	Commodities	gold, steel, grain, mineral, commodity
4	Growth	growth, growing, betting, impressive, fastest-growing
5	Quant/Options	numbers, means, analysts, calls, contracts
6	Emotion	comfortable, insane, forgotten, desperate, theme
7	Government	usa, health, congress, regulation, aid
8	Gadgets and Games	iphone, gadget, research, rebuild, nintendo
9	Healthcare/Pharma	medical, treat, cancer, patent, pharma
10	Warnings	pain, decelerating, bandwagon, addiction, dangerous
11	Infrastructure	restructuring, subsidiary, modern, aging, funds
12	Managers (personal)	idiot, silly, wild, relentless, tough
13	Telecom	cable, networking, phone, network, broadband
14	Competition	gaming, playing, slaughtered, dramatically, enemy
15	Fixed Problem	recover, new-high, solved, bright, impossible
16	Executives (negative)	ceos, exec, former, clearing, shame
17	Mergers and Acquisitions	lbo, takeover, aquire, private, equity
18	Value	appreciation, dividend, value, yield, buffett
19	Construction	construction, caterpillar, infrastructure, rubble, asbestos
20	Retail	retailer, retail, high-end, store, mall

Third, stocks which have been written about in the financial press are likely to be fresh in retail investors' minds. We therefore include measures of the volume and sentiment of articles published about the companies in our sample. Our corporate news data is provided by the Reuters NewsScope Archive, which contains a comprehensive set of business news for the entire period of our study. To measure sentiment, we average the Reuters sentiment scores (between 0 and 1) for the articles in each day, weighting them by relevance, to compute positive, neutral, and negative sentiment indicators for each company over every air date. Fourth, search volume is a direct measure of active attention because it indicates that investors are seeking information about a stock they may be considering. Search volume has been shown to predict a number of consumer decisions including car purchases (Choi and Varian 2009), home purchases (Wu and Brynjolfsson 2009), and video game sales (Goel et al. 2010). We therefore include an aggregate of the Google Insights search volume index for the stock's ticker over the 5 days prior to a recommendation.

Measuring the Prior Knowledge of Consumers

Given that an audience member knows that a stock exists, we theorize that what they already know about the stock may affect whether they find the message content persuasive or not. Instead of measuring prior knowledge explicitly, we measure the cosine distance (a measure common in the information retrieval literature) from Cramer's recommendations to text of news articles about each company provided by Reuters, resulting in a measure of the novelty of his message content for his viewers. More novel recommendations will contain rarer words and words which were not present in prior news articles.

For each event in our sample, we selected up to ten English-language articles tagged as relevant to the company Cramer recommended. To ensure the timeliness of the content, we only selected articles published at most 5 days before the start of a show. For 74% of the sample, at least one article was written about the recommended stock during that time interval. We compute tf-idf weighted cosine distance, letting $\beta = (\beta_w)_{w=1}^{W}$ be a vector of term frequencies,

$$Sample_{it} = \frac{k_{it}}{\log(\beta)}\,;\; Prior_{it} = \frac{ARTS_{it}}{\log(\beta)};\; Cosine\ Dist.(it) = \cos^{-1}\left(\frac{Sample_{it} \cdot Prior_{it}}{\|Sample_{it}\|\|Prior_{it}\|}\right).$$

Results

The Effect of Selection Bias on Estimates of Cramer's Price Effect

We estimate a specification of the linear model in equation [1] to measure the effect of the number of words in Cramer's recommendation on overnight return, including controls for weekday, how far into the sample the recommendation occurred, and a dummy variable indicating whether the recommendation was made during the "lightning round" segment on *Mad Money*.

Model 1 in Table 4 provides OLS estimates of the baseline model, demonstrating that there is a clear information volume effect to Cramer's recommendations. An additional 100 words included in a recommendation is associated with a 0.6% increase in the overnight return of the stocks in our sample—an economically significant value. The effect is weaker for stocks mentioned in the lightning round, and that Cramer's effect is diminishing over time. Inexpensive stocks and stocks for smaller companies exhibit an increased response to Cramer's recommendation. Several important results from this estimation also describe how Cramer's selection process biases naïve estimates of his price effect.

Model 2 of Table 4 shows maximum likelihood estimates of the same model in the Heckman framework. The selection adjustment is significant, which implies we can reject the null hypothesis of no sample selection bias providing evidence in favor of Hypothesis 1. The maximum likelihood estimates provided in Table 4 show compelling evidence of correlation between the residuals of the selection and outcome models. Likelihood ratio tests strongly reject the null hypothesis that $\rho = 0$ ($\chi^2[1] = 671; p < 0.001$), providing strong evidence for the presence of selection bias. We therefore use the sample selection estimation framework in all ensuing analysis. The differences between the adjusted and unadjusted coefficient estimates provide insight into how Cramer's choice biases naïve model estimates.

First, the volume of his speech – the number of words he uses to recommend a stock – has a 50% smaller positive effect on the stock's price than implied by naïve estimates. Though estimates for specifications 1 and 2 each show a clear information volume effect to Cramer's recommendations, the measurement is biased upward when we do not account for selection. In naïve estimates, an additional 100 words in a recommendation is associated with a 0.6% increase in the overnight return of the stocks in our sample. Accounting for the selection effect diminishes this estimate by half—a statistically and economically significant difference. A simple interpretation of this result is that Cramer chooses to speak longer about stocks that are more likely to go up in overnight returns with or without his recommendation, creating an upward bias in estimates of his price effect. This finding supports Hypothesis 1.

Second, Cramer's price effect is more influential in the lightning round than predicted by naïve estimates that do not account for selection. Lightning round stocks are suggested for discussion by callers to the show rather than by Cramer. It is not surprising, therefore, that Cramer's price effect for lightning round stocks is lower than for stocks he chooses (see Model 2). Our selection model suggests that Cramer's selection process to some extent explains which stocks are selected in the lightning round as well. The positive effect of his recommendations in the lightning round are on average higher than predicted by the naïve model that does not consider his selection process, meaning he is more influential than we would think in the lightning round if we did not account for his selection process. Cramer's selection process is correlated with lightning round selection, perhaps because callers' choices are correlated with Cramer's choices or because of a caller screening process during which Mad Money employees choose between callers based on what Cramer is interested in talking about.

Third, we observe that Cramer's effect diminishes over time, but that this decline is overestimated in the naïve model which does not account for selection. At least a portion of the decline in Cramer's effect on overnight returns is explained by the fact that his selection process is changing over time. Finally, stocks with large market capitalizations exhibit a diminished response, likely because of higher liquidity. When the selection adjustment is included, we can see that our estimate of this liquidity effect is even stronger, which is likely because Cramer's selection process is related to company size. Inexpensive stocks also exhibit an increased response and this coefficient estimate is unaffected by the adjustment.

Table 4: Baseline Influence Model

Dependent Variable	Overnight Return	
	(1)	(2)
Lightning Round	-1.143***	-0.677***
	(0.130)	(0.108)
Days/365	-0.664***	-0.351***
	(0.075)	(0.063)
Log Price	-0.370***	-0.327***
	(0.070)	(0.074)
Log Market Cap	-0.337***	-0.878***
	(0.041)	(0.046)
Rho		-0.903***
		(0.007)
Number Words	0.665***	0.332***
	(0.082)	(0.072)
Weekday Dummies	Yes	yes
Fama French Coefs.	Yes	yes
Heckman ML	No	yes
R-squared	0.209	
Log-likelihood	-5737.824	-14444.735
N	2397	76704

Notes: OLS and MLE for specifications 1 and 2 respectively. The estimate of σ_ϵ is significant but omitted. 2 includes all 2,397 treated observations.

Table 6: Models Estimating the Effect of Novelty

Dependent Variable	Overnight Return	
	(1)	(2)
Lightning Round	-0.377***	-0.377***
	(0.112)	(0.112)
Days/365	-0.159*	-0.162*
	(0.064)	(0.064)
Log Price	-1.344***	-1.326***
	(0.101)	(0.101)
Number Words	0.672***	-1.060
	(0.172)	(0.646)
Number Words^2	-0.091*	-0.072
	(0.037)	(0.039)
Cosine Distance		-0.609
		(0.696)
Cosine Dist. X Num. Words		1.153**
		(0.407)
Rho	-0.929***	-0.928***
	(0.005)	(0.005)
Weekday Dummies	yes	yes
Fama French Coefs.	yes	yes
Attention Controls	yes	yes
Log-likelihood	-14291.876	-14287.442
N	76704	76704

Notes: Maximum likelihood estimation. The estimate of σ_ϵ is significant but omitted for brevity.

The Effects of Message Content on Stock Price Movements

So far we have used only one dimension of the qualitative content of Cramer's recommendations to estimate his effect on investor decision making: the number of words he uses. In this section, we analyze the topical content of his speech and show that some topics are significant and that they jointly improve the fit of our influence models. The effects of topical content are shown in Table 5, Models 1-3, which contain estimates of the baseline model including topic variables.

In specification 1 of Table 5, 6 of the 20 topic variables are significant at the 10% level or higher (Topics 3, 5, 6, 8, 9, and 13), indicating that LDA has identified distinct word distributions that vary with respect to their level of influence over *Mad Money* viewers. A likelihood ratio test for the joint significance of the topic variables rejects the null hypothesis that parameter estimates of topic variables are zero ($\chi^2[19] = 58, p < 0.001$). The topics Cramer chooses to discuss during his recommendations appear to significantly affect viewer response. Though we are reluctant to interpret how or why some topic words tend to be more influential, it is apparent that topics vary drastically in their impact on overnight returns. Words in Topics 3, 9, and 13 have point estimates over six times larger than the baseline words estimate, while those in Topic 8 actually appear to undermine Cramer's persuasiveness.

Table 5. Topic Influence Models (Dependent Variable = Overnight return)

	(1)		(2)		(3)		(4)		(5)	
Lightning Rd.	-0.531***	(0.598)	-0.306*	(0.119)	-0.209*	(0.089)	-0.455***	(0.108)	-0.372***	(0.111)
Days /365	-0.218**	(0.111)	-0.282***	(0.077)			-0.182**	(0.063)	-0.101	(0.072)
Log Price	-0.286***	(0.072)			-0.272***	(0.065)	-1.355***	(0.102)	-1.298***	(0.102)
Log Mkt. Cap	-0.841***	(0.074)			-0.578***	(0.039)	0.092	(0.074)	0.089	(0.073)
Log Trade Vol.							-1.180***	(0.082)	-1.165***	(0.082)
Prev. Mention							-0.532***	(0.161)	-0.505**	(0.162)
Log News Vol.							-0.055	(0.045)	-0.051	(0.045)
Search Vol.							-0.062	(0.059)	-0.043	(0.059)
Num. Words							0.291***	(0.070)		
Topic 1	0.147	(0.046)	1.659*	(0.708)	0.255	(0.555)			-0.433	(0.639)
Topic 3	1.979**	(0.662)	2.618***	(0.647)	2.142***	(0.570)			1.740**	(0.655)
Topic 5	0.660*	(0.273)	1.484***	(0.315)	0.528*	(0.219)			0.579*	(0.273)
Topic 6	0.905***	(0.261)	0.991***	(0.275)	1.149***	(0.217)			0.863***	(0.256)
Topic 8	-0.759**	(0.295)	-0.506	(0.305)	-0.503*	(0.252)			-0.093	(0.298)
Topic 9	1.908**	(0.589)	2.574***	(0.727)	1.603**	(0.494)			1.021	(0.585)
Topic 10	-0.958	(0.641)	-1.500*	(0.692)	0.069	(0.571)			-0.598	(0.629)
Topic 12	-0.617	(0.328)	-0.827*	(0.344)	-0.148	(0.277)			-0.271	(0.320)
Topic 13	2.282**	(0.770)	1.617*	(0.782)	3.037***	(0.623)			2.297**	(0.781)
Topic 17	0.526	(0.732)	1.948*	(0.834)	1.036	(0.648)			-0.108	(0.704)
Topic 18	0.469	(0.448)	1.180**	(0.441)	0.583	(0.374)			-0.028	(0.424)
Rho	-0.902***	(0.007)	-0.697***	(0.027)	-0.798***	(0.016)	-0.930***	(0.005)	-0.930***	(0.005)
Stock F.E.	No		Yes		No		No		No	
Show F.E.	No		No		Yes		No		No	
Log-like.	-14415.753		-11906.723		-13750.638		-14294.760		-14274.132	
N	76704		76290		76641		76704		76704	

Notes: Maximum likelihood estimation. The estimate of σ_ϵ is significant but omitted for brevity. Models 1,4 and 5 include all 2,397 treated observations. Models 2 and 3 contain treated 1,983 and 2,334 observations respectively. Topic variables which were not significant in at least one model are not reported. Specifications without stock fixed-effects contain Fama French coefficients. Specifications without show fixed-effects contain weekday dummy variables.

A potential alternative explanation is that estimated topics merely reflect the characteristics of the recommended stocks or of particular show dates, which then exhibit a different response depending on their attributes (e.g. unobserved stock or show characteristics). To test for this, we exploit the fact that Cramer recommends 441 of the stocks in our sample more than once, representing 83% of the total treated events (N=1983). We estimate a sample selection model with either stock or show fixed-effects, omitting the stock- or show-invariant characteristics from the regressors. Thus, after controlling for any company-specific or show-specific response, the same likelihood ratio test can determine if there is an effect of the topics controlling for the selection of topics based on characteristics of the stock.

Models 2 and 3 in Table 5 provide estimates of the model after including stock and show fixed-effects, respectively. After ruling out the alternative explanation of a stock- or show-specific response and selection of content based on characteristics of the stock, we can again reject the null hypothesis in a likelihood ratio test for the joint significance of the topic variables ($\chi^2[19] = 75, p < 0.001; X^2[19] = 61, p < 0.001$). While controlling for company fixed-effects, five of the six topic variables which were significant in Model 1 remain significant at the 10% or lower level. All six topic variables remain significant in the show fixed-effects specification.

These results indicate that the topics Cramer discusses when recommending a stock influence Cramer's price effect, while ruling out the possible bias resulting from omitting observable and unobservable characteristics of stocks and shows which may be correlated with Cramer's choice of the topics.

The Effects of Prior Attention on Cramer's Price Effect

Models 4 and 5 in Table 5 report estimates from models after the awareness controls have been included. In both specifications, each of the four controls for prior awareness have the correct sign, and log trading volume and the dummy variable for a previous mention are significant at the 5% level or greater. The negative and significant coefficient estimates for the attention variables support Hypothesis 2—stocks with high levels of prior trading volume and that have been discussed in a previous episode exhibit a diminished response to Cramer's recommendations.

Model 4 includes only the number of words variable as a measure of content. The control variables effectively improve the fit of the model ($\chi^2[4] = 300, p < 0.0001$) but they do not significantly alter the estimate of the coefficient on the number of words. We may be tempted to conclude that the attention measures do not substantively change the results.

However, when we compare estimates on the topic coefficients once the attention controls are included (comparing Model 5 to Models 1-3) our results are suggestive of an omitted variables bias. Topics 8 and 9 are no longer significant at the 10% level. The decline of the point estimate for Topic 9, which includes words related to pharmaceutical and healthcare, is an example of how this bias may occur. If Cramer decides to make arguments about the pharmaceutical industry only for companies which are relatively unknown to his audience, then we will tend to overestimate the impact of this type of recommendation. The topics are still jointly significant ($\chi^2[19] = 41, p < 0.01$), but the controls have diminished our confidence in the inference that some of the topics significantly influence investor choices.

The Effects of Prior Knowledge on Cramer's Price Effect

Model 1 in Table 6 provides estimates of models of overnight returns which depend on the volume of information and its quadratic term. The negative and significant coefficient on the squared number of words provides evidence of the decreasing effectiveness of more information. This finding is consistent with the prediction that the change in the posterior mean is marginally decreasing in sample size.

Model 2 in Table 6 is the main result of this section. Notice that neither number of words nor its quadratic term is significant at the 10% level or lower. Cosine distance, which is our measure for novelty, is also not statistically distinguishable from zero. However, the interaction between number of words and cosine distance is positive, large, and significant at the 1% level ($t[1] = 2.8$). The positive measurement of the interaction between volume and novelty suggests that neither novelty nor volume of information alone is sufficient to influence the audience more than average. This result suggests that novel information acts synergistically with volume to affect influence, a finding consistent with Hypothesis 3.

Discussion

When individuals make complex decisions, they are likely to utilize qualitative information which can express relevant details about alternatives they are comparing. In markets where product and service offerings have become diverse, such as retail e-commerce (Brynjolfsson et al. 2006), consumers increasingly rely on this kind of information. With the advent of machine learning techniques which can quantify text such as LDA and the growing availability of data on qualitative information which is tied to individual decisions, there has never been greater potential for use of text analysis by managers and marketers. This line of research has obvious and immediate applications in the design or delivery of persuasive messages, but these are exactly the kinds of applications in which modeling that does not address causality is unlikely to generalize.

This study highlights some of the sources of endogeneity which may affect estimation of the impact of message content on consumer behavior and specifically on stock prices, factors which are likely to confound causal inference. Observational data on content consumption and subsequent behavior are useful because they span a variety of message content and contain a record of individuals' behavior "in the wild." However, what we observe in observational data is qualitative message content that is endogenously produced and consumed.

In support of Hypothesis 1, we showed that explicitly modeling how a stock is chosen to be mentioned on *Mad Money* can drastically change estimates of the effect of the number of words used to recommend the stock on its overnight returns. This is only one type of non-random sampling that can occur with observational data. In the general case, an observed treatment is a combination of the decision of a producer about what to discuss and what content to create and the consumer's choice to consume that content.

In addition to selection bias on the part of the content producer, our findings also provide compelling evidence that individuals do not process message content *tabula rasa*. What they know prior to consuming qualitative information mediates its effect. Individuals' prior awareness and knowledge present two potentially omitted variables which can bias estimates. We found evidence in support of Hypothesis 2; more prior attention to the stock reduced the impact of a recommendation to buy the stock. We also demonstrated that measures of prior attention are correlated with message content variables, and that their omission biases estimates of the effect of topical content. The empirical observation that some dimensions of content are correlated with an individual's prior awareness supports the hypothesis that content is produced with some notion of how likely it is that the audience knows about its subject, providing another example of how content may be endogenously produced.

Conclusion

Qualitative content is abundant and individuals rely on it to inform important decisions. Causal inference about the effect of qualitative information is an important research topic for both managers and researchers: without understanding why content is persuasive, it is difficult to effectively devise more influential content or to better understand how individuals process information.

Persuasive messages are often delivered in order to target certain individuals, pertain to particular products or items, and sought by those who need the information most. Generically, there is a non-random selection process that could lead to biased estimates of the effect of content variables. As we show, this selection bias can be significant and misleading. The omission of control variables for prior awareness causes a similar inferential problem. When awareness of the existence of products is correlated with attributes of message content—for example when more compelling arguments are used to discuss newer products—it becomes important to properly measure prior awareness in order to mitigate this additional source of endogeneity.

Besides drawing attention to the issues surrounding causal inference on the effect of message content, our paper provides suggestive evidence that recipients' prior knowledge interacts with message content to inform their choices. Marketing campaigns take great care to target the right individuals with relevant offers or advertisements. These types of campaigns typically target on the basis of age, gender or other demographic factors. But the persistent attributes of the message recipient may not be as important as what they know prior to processing a message. Next generation marketers, enabled by an unprecedented amount of behavioral data (Lazer et al 2009), may seek design or select information content to fit the prior knowledge of the customers to whom they are marketing.

References

Anand, B., R. Di Tella, and A. Galetovic. (2007) "Information or opinion? Media bias as product differentiation," *Journal of Economics & Management Strategy*, Vol. 16, No. 3, 635-682.

Antweiler, W. and M.Z. Frank. (2004) "The market impact of corporate news stories," *Working Paper*.

Aral, S. and Van Alstyne, M. (2011) "Networks, Information and Brokerage: The Diversity-Bandwidth Tradeoff." *American Journal of Sociology*.

Archak, N., A. Ghose, and P.G. Ipeirotis. (2007) "Deriving the pricing power of product features by mining consumer reviews," NET Institute Working Paper No. 07-36.

Berger, J., A.T. Sorenson, and S.J. Rasmussen. (2010) "Positive effects of negative publicity: when negative reviews increase sales," *Marketing Science*, Vol. 29, Issue 5, 815-827.

Blei, D.M., A. Ng, and M. Jordan. (2003) "Latent Dirichlet allocation," *Journal of Machine Learning Research*, 3:993-1022.

Brynjolfsson, E.,Y. Hu, and M.D. Smith (2006) "From niches to riches: Anatomy of the long tail," *MIT Sloan Management Review,* Vol. 47, No. 4.

Choi, H. and H.R. Varian. (2009) "Predicting the present with Google trends," *Google Research Blog*. http://ssrn.com/abstract=1659302

Das, S.R., and M.Y. Chen. (2007) "Yahoo! for Amazon: Sentiment extraction from small talk on the web," *Management Science*, Vol. 53, No. 9, 1375-1388.

Engelberg, J., C. Sasseville, and J. Williams. (2010) "Market madness? The case of Mad Money," *SSRN Working Paper*. http://ssrn.com/abstract=870498

Fama, E. and K. French. (1993) "Common risk factors in the returns of stocks and bonds," *Journal of Financial Economics*, Vol. 25, 383-417.

Gentzkow, M. and J.M. Shapiro. (2006) "Media bias and reputation," *Journal of Political Economy*, Vol. 114, No. 2, 280-316.

Ghose, A., P.G. Ipeirotis, and A. Sundararajan. (2006) "The dimensions of reputation in electronic markets," *NYU Center for Digital Economy Research Working Paper* No. CeDER-06-02.

Heckman, J. (1979) "Sample selection bias as a specification error," *Econometrica*, 47 (1): 153-161.

Goel S., Hofman J.M., Lahaie, S., Pennock, D.M., and D.J. Watts (2010) "What can search predict," *WWW 2010*.

Karniouchina, EV, WL Moore, and KJ Cooney. (2009) "Impact of Mad Money stock recommendations: merging financial and marketing perspectives," *Journal of Marketing*, 73: 244-6.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, A., King, G., Macy, M., Roy, D., Van Alstyne, M. 2009. "Computational Social Science," Science, February 6: 721-722.

Levitt, S.D. and C. Syverson. (2008) "Market distortions when agents are better informed: The value of information in real estate transactions," *The Review of Economics and Statistics*, 90(4): 599-611.

Lo, A.W. and MacKinlay, A.C. (1988) "Stock market prices do not follow random walks: evidence from a simple specification test," *Review of Financial Studies*, Vol. 1, No. 1, 41-66.

Lim, B. and J. Rosario. (2010) "The performance and impact of stock picks mentioned on 'Mad Money'," *Applied Financial Economics*, Vol. 20, No. 14. 1113-1124.

Merton, R.C. (1987) "A simple model of capital market equilibrium with incomplete information," *Journal of Finance*, Vol. 42, No. 3, 483-510.

Mitra, A. and J.G. Lynch, Jr. (1995) "Toward a reconciliation of market power and information theories of advertising effects on price elasticity," *The Journal of Consumer Research*, 21(4): 644-659.

Neuman, J.J., and P.M. Kenny. (2007) "Does Mad Money make the market go mad?" *The Quarterly Review of Economics and Finance*, 47: 602-615.

O'Connor, B., R. Balasubramanyan, B.R. Routledge, and N.A. Smith. (2010) "From tweets to polls: linking text sentiment to public opinion time series," *Proceeedings of the Fourth International AAAI Conference on Weblogs and Social Media*.

Tetlock, P.C. (2007) "Giving content to investor sentiment: the role of media in the stock market," *The Journal of Finance*, Vol. 62, Issue 3, 1139-1168.

Wu, L. and E. Brynjolfsson. (2009) "The future of prediction: how google searches foreshadow housing prices and quantities," *ICIS 2009 Proceedings*.