Essays on Social Media:

Evidence from the Music Industry and the Stock Market

A Dissertation Proposal

Submitted to the Faculty

of

Purdue University

by

Hailiang Chen

October, 2010

Purdue University

West Lafayette, Indiana

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

This dissertation studies two aspects of the emerging social media. The first essay documents the interesting phenomenon that broadcasting in online social networks evolves into a new form of marketing. Using the activity stream data from MySpace and the sales rank data from Amazon, it employs a panel vector autoregressions model to quantify the effect of social network marketing on music sales. The second essay addresses the growing importance of social media and evaluates its influence above and beyond the traditional mainstream media in the stock market. By conducting textual analysis on the articles published on a popular social media platform Seeking Alpha and the Wall Street Journal, it establishes the significant correlation between social media sentiment and stock returns. The additional commenting data available from the Seeking Alpha website makes it possible to evaluate whether there is a causality relationship between media sentiment and stock returns.

CHAPTER 1. INTRODUCTION

Recently, there has been a surge of interest in studying social media due to its increasing popularity and ubiquity. This dissertation contributes to this stream of research by looking at the social media from two different perspectives. It picks two of the most popular forms of social media, namely, social networking and blogging, draws upon the literature of marketing and finance, respectively, and brings fresh insight into how social media is playing an important role in this information age.

The first essay examines artists' broadcasting behavior on the popular social networking site, MySpace, and studies the interactions between artists' online activities and music sales. We employ a panel vector autoregression (PVAR) model to analyze a dataset containing the activity stream data from MySpace and the music sales rank data from Amazon. This model allows us to treat our main variables as endogenous and to investigate the inter-relationship between social networking promotions and music sales. We find empirical evidence that an artist's network size plays a crucial role in moderating the effect of social networking activities on sales. For the artists with many friends, broadcasting activities on MySpace have a significant impact on music sales; this, however, is not true for the artists with only a few friends. Our results still hold after controlling for new album releases and artists' popularity over time. Finally, we observe that the relationship between MySpace activities and sales ranks can be nonlinear with the possibility of over-marketing.

The second essay investigates the role of both traditional and social media in the stock market and the possible causality link from financial blogs to stock returns. We conduct textual analysis on articles collected from the Wall Street Journal and a popular social media website Seeking Alpha, and find that social media sentiment has a significant impact on stock returns even after controlling for traditional media sentiment. We also observe that more intensely discussed articles on Seeking Alpha tend to have a stronger impact on stock returns.

CHAPTER 2. BROADCASTING IN ONLINE SOCIAL NETWORKS: AN EMPIRICAL STUDY OF ARTISTS' ACTIVITIES AND MUSIC SALES

2.1. Introduction

Social networking sites such as Facebook, MySpace, and Twitter have become an important component of people's daily life. In the past few years, social network technology has been widely adopted by companies to market their products. For instance, a *Wired* article (Silver 2009) notes, "Twitter, at first a place to tell everyone what you ate for breakfast, is now a place to promote yourself, your company or your product." Companies can broadcast information regarding their products to their Facebook "friends" or embed marketing messages in their "tweets" (a.k.a., status updates) to their Twitter "followers". Companies also hope that those consumers, with whom they directly communicate, may relay the marketing messages they receive to their own networks of friends. Because these social networking sites reach millions of users who often spend hours and hours on these sites, they can serve as a very powerful marketing channel for companies. For instance, according to a news report in *InformationWeek*, from 2007 to June 2009, Dell has generated a total of $2 million in direct sales of refurbished systems and $1 million in indirect sales of new systems from their Twitter presence @DellOutlet (Gonsalves 2009). Twitter is becoming such an effective marketing tool that advertising startups like Ad.ly and SponsoredTweets have signed up thousands of Twitter users and paid them up to $10,000 per tweet for sending advertising messages to their Twitter followers (Learmonth 2010).

MySpace, one of the world's largest social networking sites, is a prominent example of how social networks can be used by musicians for marketing purposes. Over eight million artists and bands have set up their profiles on MySpace (Owyang 2008). F. Vincent, the author of the book titled “*MySpace for Musicians*”, puts it this way: “with so many potential pairs of eyes and ears at your fingertips, it is becoming a necessity for any musical artist – whether signed and selling or unsigned and hopeful – to have a profile on MySpace.” MySpace offers many great free tools, such as bulletins and activity streams, and focuses on building a community of artists and music fans. Artists and bands can upload songs, show music videos, communicate with fans, and even sell MP3 downloads through MySpace. Moreover, there are different kinds of commercial software designed to help musicians promote their music on MySpace (e.g., MySpace Friend Pro: http://myspacefriendspro.com).

Despite the abundant anecdotal evidence indicating that social network marketing can be very effective in driving up product sales (e.g., Baker 2006, Gonsalves 2009, Vincent 2010), there is a lack of academic research that examines the dynamic relationship between companies’ social networking activities and their product sales. Do companies’ marketing activities on social networks really have an impact on their product sales? How does such an impact vary across companies? What are the key factors that determine the success of companies’ marketing activities on social networks? These questions remain largely unanswered to date.

To answer the above questions, we use artists’ marketing activities on MySpace as an example. We collect artists’ activity stream data from MySpace and combine it with the sales rank data from Amazon. We address the potential endogeneity problem by employing a panel vector autoregression (PVAR) model (Holtz-Eakin et al. 1988) estimated by the Generalized Method of Moments (GMM) (Binder et al. 2005), and study the dynamic relationship between artists’ activities and music sales. The activity stream data analyzed in this study is quite unique – it is the time series data on artists' marketing messages broadcasted to their friend networks. The PVAR model applied to this data enables us to treat all of these time series variables as endogenous and to avoid making any restrictive assumptions.

Our results suggest that artists' marketing efforts on MySpace do have a significant effect on their music sales. We also examine how this effect varies across different artists, by dividing the sample into two equal-size groups according to the average number of friends each artist has. Interestingly, we find that the significant impact of MySpace activities on music sales applies only to those artists who have a relatively large network. For the artists with a relatively small network, this impact is statistically insignificant.

We conduct a few additional analyses. Specifically, we use the information on artists' new album releases as an additional control as artists' marketing activities in traditional channels often occur around the time of new releases and the introduction of new albums itself could lead to a change in artists' sales ranks. We also use the consumer search volume index from Google Trends as an indirect measure of artist popularity and thus are able to account for time-varying individual characteristics in dynamic panel models. We find that our results are robust to these phenomena. In addition, we test the case when quadratic terms are included in the regression for social networking activities, to and find that the relationship between these activities and sales ranks can be nonlinear.

This study makes a number of contributions to the IS literature. First, we use a novel dataset on companies' social networking activities, i.e., we use company-generated content, rather than consumer activities or user-generated content (UGC). To date, IS research has studied consumers' activities on the Internet (e.g., their reviews, ratings, and blogs) and the impact of such activities on product sales (see, for example, Dellarocas 2003, Chevalier and Mayzlin 2006, Chen et al. 2007, Forman et al. 2008, Dewan and Ramaprasad 2009, and Zhu and Zhang 2010). To the best of our knowledge, there is no existing research that has examined how companies' activities on social networking sites can impact their product sales. Our study bridges this gap in the literature. Second, the study introduces a relatively new econometric methodology – the PVAR model, estimated using a GMM estimator – to the literature on social networks. Because endogeneity is often an important concern for this type of research, this methodology can be particularly useful in identifying the relationships, if any, in both directions. This study also has important managerial implications. In light of our findings, companies should enhance their efforts to utilize social networks as a marketing channel. In addition, they should pay close attention to their network size when planning marketing activities on social networks – the network size is a key factor that determines the success of companies' social-network-based marketing efforts. Finally, companies should be careful about not over-marketing on social networks.

The remainder of this chapter is organized as follows. In Section 2.2, we review recent IS studies that examine the relationship between consumers' online activities and sales, as well as the relevant literature on the impact of marketing effects on sales. Section 2.3 first discusses how MySpace helps artists promote their music, and then gives a brief overview of the data. We present our empirical analysis and main results in Section 2.4. The robustness checks are carried out in Section 2.5. Section 2.6 concludes the study and suggests some directions for future research.

2.2. Literature Review

Many papers have examined whether online word-of-mouth or user-generated content, such as consumers' reviews, ratings, and blogs, have an impact on sales. The earlier studies (see, for example, Dellarocas 2003, Chevalier and Mayzlin 2006) have tried to establish the relationship between online consumer reviews and product sales. The more recent studies have taken more nuanced approaches toward examining such a relationship. For instance, Chen et al. (2007) study how the number of helpful votes on reviews and the reputation of reviewers influence the relationship between book ratings and book sales. Forman et al. (2008) investigate the role of reviewer identity disclosure in affecting the relationship between consumer reviews and sales. Li and Hitt (2008) study the time series of consumer reviews and finds that early reviews are subject to self-selection biases and can have an influence on the long-term consumer behavior. Zhu and Zhang (2010) consider how product and consumer characteristics moderate the relationship between consumer reviews and product sales. In addition, researchers have started to pay attention to how consumer blogs can drive product sales. For instance, Dewan and Ramaprasad (2009) use the Granger Causality tests and two-stage least squares to study the causal relationship between blog buzz and music sales. All of these existing studies utilize data on user-generated content and address the impact of consumer behaviors on product sales. Our study is related to the existing literature that studies how online content can influence product sales. However, it differs greatly from this literature because we examine online content from a completely different angle and study companies' activities (or company-generated content) rather than consumers' activities (or user-generated content). To the best of our knowledge, no other paper has tried to quantify the value of social network marketing[1].

This study also draws upon the marketing literature that studies the effect of traditional advertising on sales. According to Dekimpe and Hanssens (1995), a marketing action can affect the sales performance of a brand or a firm in six ways: contemporaneous effects, carry-over effects, purchase reinforcement, feedback effects, firm-specific decision rules, and competitive reactions. In general, advertising often has an immediate effect on sales. A more subtle question to marketing managers is how long the cumulative effect of advertising persists. Early studies (e.g., Givon and Horsky 1990) in marketing have documented that the effect of advertising in one time period may be carried over, at least partially, into future periods. It can be argued that consumers remember past advertising messages, but this “goodwill” toward the advertised brand gradually decays because of forgetting and competitive advertising. Givon and Horsky (1990) and Horsky and Simon (1983) propose that advertising can also indirectly affect sales through purchase reinforcement. Advertising can encourage consumers to try a new product or a product they have not purchased for a long time; then, if they like their experience, they may purchase it again in the future. In addition, Simester et al. (2009) show that current advertising affects future sales through the two competing effects of brand-switching and inter-temporal substitution. These previous findings suggest that it is important to use appropriate models that capture the lagged effects when the effects of marketing activities on sales are studied.

---

[1] The only exception known to us is the work by Lin et al. (2009), who quantify the value of social networks in online peer-to-peer lending markets.

Previous studies (see, for example, Bass 1969, Hanssens 1980) also conclude that advertising efforts should not be treated as exogenous as they may be influenced by the current and past performance of sales. This phenomenon essentially calls for a research method that can treat both advertising and sales as endogenous variables. Failure to do this may result in a bias in estimation. In addition, competitive reactions could also have a major impact on the effectiveness of advertising. In the short run, marketing actions may prompt a positive sales response, but the long-run effect could be negligible depending on the nature of competitive reactions (Hanssens 1980).

We apply the insights from the studies mentioned above to the online social network context, and employ a time series model that allows us to investigate the effects of artists' online marketing actions on sales and, at the same time, address the endogeneity problem. The model we use is a variant of the vector autoregression (VAR) model developed in the seminal work by Sims (1980). The VAR model has been widely used to analyze time series in macroeconomics, finance, and other fields (see, for example, Bessler 1984, Sims 1992, McCarty and Schmidt 1997). In the social networking context, Trusov et al. (2009) employ a VAR model to study the impact of word-of-mouth referrals and traditional marketing on the number of sign-ups at a major social networking site[2]. In principle, VAR models can also be applied to study individual-level data. For example, Holtz-Eakin et al. (1988) is the first paper to estimate a panel vector autoregression model, revealing the dynamic relationship between individuals' hours of work and their wages. According to Pauwels et al. (2004), time series econometric models such as VAR-based models have made considerable contributions to the marketing literature.

---

[2] The authors (Trusov et al. 2009) do not disclose the site name in their paper.

2.3. Data

2.3.1. How Artists Promote Themselves on MySpace?

As a social networking site, MySpace allows bands and artists to set up profile pages that are different from the ones for normal users. This aims to provide a professional platform for artists to connect with their fans besides building their personal networks. On a typical artist profile page[3], aside from the basic profile information about the artists, there is usually a music player that allows visitors to play the songs in a playlist specified by the artists. The artists can also list the schedules of their upcoming shows or concert performances. Many artists also leave space for some recent blog entries and upload a few videos. Finally, a top friend list and some recent comments by friends appear at the end of the profile page.

The information provided on the profile page serves as only a starting point of online music promotion. If MySpace users do not visit the profile pages, they might not be aware of the artists' recent activities. MySpace thus provides other tools to exploit the benefits of social networks. One of these tools is the bulletin board. Bulletins are posts that everyone in the friend list can see. The commercial software MySpace Friend Pro introduces its bulletin poster feature starting with the following, "Once you have enough friends on myspace.com, posting bulletins can be very very effective and unlike commenting and messaging you do not need to send individual bulletins to all your friends…People read bulletins as long as the topic is catchy and somewhat

---

[3] For an example of an artist profile page, please visit http://www.myspace.com/onerepublic.

relevant[4]." Artists can post bulletins on a timely basis to tell their friends what they are doing right now, ask them to listen to their new songs, and invite them to post comments. Once a user becomes an artist's friend, then that user will automatically receive the artist's bulletins and be able to see them until they expire in 10 days.

Another major promotional tool is the friend updates that inform music fans of the artists' recent MySpace activities (MySpace recently renamed "friend updates" to "activity stream"). The benefit of the friend updates lies in the fact that any activities by the artists are automatically updated to their friends if they choose to receive these updates. MySpace, like other social networks, also periodically send emails to users to notify them of the recent friend updates. This feature thus enables artists to spread out a message to their fans very quickly and efficiently in online social networks, especially considering that the number of friends can easily reach hundreds of thousands for popular artists. There are several different types of activities that can show up in an activity stream. Among them, "add new blogs", "add new photos", "upload new track", and "add new friends" are the most common ones. By default, MySpace users can see their friends' updates in the past month after they login.

To summarize, the fundamental objective of promoting music is to keep music fans interested in them over time. To achieve this, MySpace provides the bulletin board and activity stream to automatically spread artists' messages across the network of their friends. Previous research has not studied the marketing impact of these automatic news feeds in virtual social networks. This research aims to examine the roles of these emerging promotional tools.

---

[4] HTTP://MYSPACEFRIENDSPRO.COM/BULLETIN.PHP.

2.3.2. Data Description

We use two data sources in this research, namely, the Amazon sales rank data and the MySpace activity stream data. Amazon tracks the sales ranks of all music artists that sell digital albums or songs in the Amazon MP3 store. The activity stream on MySpace records all the promotional activities by artists, which are then broadcasted to their' friends. To collect this information, we subscribed to a selected sample of artists as their friends and retrieved the daily updates of their MySpace activities. We also counted the number of friends for each artist every day.

One potential concern with the Amazon sales rank data is that it represents artists' digital music sales only. Although this is due to the limitation of data availability, we assess whether it creates any problem by comparing the Amazon music charts with the well-known Billboard charts, which are based on overall music sales. We retrieved the Billboard Hot 100 Songs and Amazon Top MP3 Songs on the same day and calculated the correlation between the sales ranks of the songs that appear on both lists. These two sales ranks turned out to be highly correlated, with a coefficient of 0.758. Thus, we conclude that artists' Amazon sales ranks are representative of their overall music sales ranks.

Another important problem that needs to be dealt with is how to select a sample of artists. Since there are over eight million artist profiles on MySpace Music and the list of all artists is not available to the authors, we selected the sample of artists through the lists of Amazon's daily Top MP3 Songs for all genres and 22 different individual genres (23 lists per day)[5]. We used the top song list instead of

---

[5] The list of these 22 genres is available at the following link: http://www.amazon.com/gp/bestsellers/dmusic/digital-music-track.

the top artist list to avoid selecting only the popular artists[6]. We selected three days (August 30, September 10, and September 20) in 2008 and downloaded 23 lists per day. Originally, 5,146 artists appeared in the lists, and both major labels and indie (i.e., independent or unsigned) artists were included. A program was written to search these artists' information on MySpace Music and find their corresponding profile pages automatically. To ensure accuracy, the artist name and the titles of uploaded songs on the profile page were used to match with the information on Amazon. There could be multiple MySpace profiles for the same artist name, but we selected the artists that had only one match on MySpace. Overall 1,604 artists' profile pages were exactly matched. Next, we tried to send a friend request to each of these 1,604 artists, and until October 18, 2008, successfully subscribed to 631 artists[7]. The actual data collection started on October 19, 2008 and ended on May 30, 2009; thus, we have the data for 32 weeks in total.

The main time series variables constructed for our analyses are artists' weekly average sales ranks (in logarithm), the number of bulletins, and the number of friend updates in each week. To study the role of network size, we also kept track of each artist's network size over time. Table 2.1 shows the descriptive statistics of these variables for all artists. The number of observations is 20,192, and the panel data set is strongly balanced. The sales ranks for artists range from 2 to 111,093, indicating that both popular and unpopular artists are included in the

---

[6] The top artist list is based on the total sales of all digital music; usually it is very difficult for obscure artists to get into this list, but it is much easier for them to show up in the top song list. In light of this, this we decided to use the top song list.

[7] If the friend request is accepted by the artist, then the person initiating the request becomes a friend of that artist and is able to observe his or her promotional activities.

sample. We also find that a wide range of genres are represented in the sample. The average number of bulletins in each week is only 0.67, but the maximum number could reach 48 (i.e., in the extreme case, artists post almost 7 bulletins per day). Similarly, the average number of friend updates in each week is also small (0.56), but the maximum could reach 12. The network size of each artist on MySpace is measured by the variable *AllFriends*. The median number of friends per artist is 43,349. The average is even larger (111,633), and the maximum number of friends is more than one million in our sample.

Table 2.1 Descriptive Statistics

Variable	# Obs.	Mean	Std. Dev.	Median	Min	Max
SalesRank	20,192	5,261	9,232	2,116	2	111,093
Bulletins	20,192	0.67	1.94	0	0	48
FriendUpdates	20,192	0.57	1.06	0	0	12
AllFriends	20,192	111,633	179,987	43,349	55	1,385,929

2.4. Empirical Analysis

We estimate a panel vector autoregression (PVAR) model that examines the dynamic interactions among artists' sales ranks and promotional actions. To avoid making any assumptions that may be unreasonable, we adopt the reduced form of VAR models in which each dependent variable is endogenous and is a linear function of its own past values, the past values of all other dependent variables, a set of exogenous variables, and an error term[8]. The panel structure of the data allows us to control for unobserved individual heterogeneity. Toward that end, we introduce individual-specific effects to the model. Thus, we specify the following model:

$$\mathbf{y}_{it} = \mathbf{\Phi}_1\mathbf{y}_{i,t\text{-}1} + \mathbf{\Phi}_2\mathbf{y}_{i,t\text{-}2} + \ldots + \mathbf{\Phi}_p\mathbf{y}_{i,t\text{-}p} + \mathbf{\Lambda}\mathbf{X}_t + \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{it}\,, \quad (1)$$

where $\mathbf{y}_{it} = (\ln(SalesRank_{it}), Bulletins_{it}, FriendUpdates_{it})'$ is a three-element column vector for artist *i* at time *t*, containing the dependent variables of sales rank in logarithm, number of bulletins, and number of friend updates; $\mathbf{\Phi}$'s are 3×3 matrices of slope coefficients for endogenous variables; $\mathbf{X}_t$ is a column vector of exogenous variables with the coefficient matrix $\mathbf{\Lambda}$ (one example of the exogenous variables can be time dummies that control for seasonality or time effects); $\boldsymbol{\mu}_i$ is a three-element vector of unobserved individual effects; $\boldsymbol{\varepsilon}_{it}$ is a three-element vector of errors, assumed to be independently and identically distributed (i.i.d.); and *p* is the number of lags.

2.4.1. Model Estimation

Since the error term $\boldsymbol{\varepsilon}_{it}$ in equation (1) is correlated with $\mathbf{y}_{i,t+1}, \mathbf{y}_{i,t+2}, \ldots, \mathbf{y}_{i,T}$, which are the future values of the regressors $\mathbf{y}_{i,t\text{-}1}, \mathbf{y}_{i,t\text{-}2}, \ldots, \mathbf{y}_{i,t\text{-}p}$, these lagged dependent variables are correlated with the average error term $\bar{\boldsymbol{\varepsilon}}_i$ in the within-group estimator (i.e., the least-squares estimator after subtracting the individual means of the observations) through the terms $\boldsymbol{\varepsilon}_{i1}$, ..., $\boldsymbol{\varepsilon}_{it-1}$. Thus, the within-group

---

[8] Recursive or structural forms are possible, but they require one to impose restrictions on causality or contemporaneous relationships between variables (Stock and Watson 2001).

estimator for the fixed effects model will be biased for this type of dynamic panel model (Nickell 1981, Arellano 2003). We estimate the model by the Generalized Method of Moments (GMM) (see Hansen 1982, Hamilton 1994, Hayashi 2000 for a detailed review of this method). A consistent GMM estimator that uses dependent variables which are lagged two or more periods as instruments has been developed in the econometrics literature. Estimation methods for single equations in dynamic panel models are discussed in a series of papers such as Arellano and Bond (1991), Ahn and Schmidt (1995, 1997), Arellano and Bover (1995), Blundell and Bond (1998), etc. Binder et al. (2005) summarize the GMM estimation procedure for Panel VAR models with only one lag. The derivation of the standard GMM estimator for multiple lags is provided in the appendix. To the best of our knowledge, there are no commercial statistical or econometric software packages that implement the estimation functions for this type of model. We have, therefore, written a Matlab program to estimate our model.

The estimation results for three lags are presented in Table 2.2. Our main goal is to examine the relationship between artists' marketing actions and their sales ranks; we are less interested in the inter-relationship between the bulletins and friend updates variables. We first look at how MySpace promotions affect artists' music sales. The coefficients for bulletins and friend updates at lag 1 in the *ln(SalesRank)* equation are negative (-0.008 and -0.011) and statistically significant at 1%, indicating that the dependent variable, sales rank, will decrease (i.e., sales will go up) next week when the number of bulletins or friend updates in this week increases. From an economic perspective, if an artist posts one additional bulletin this week, the sales rank will be decreased by 0.8% in the following week. Likewise, if the artist generates one additional friend update this week, the sales rank will be decreased by 1.1%. We also try to estimate these impacts in terms of sales volumes, thereby quantifying the economic value of social network marketing. Without the actual sales data, we resort to the news and press releases from Nielsen SoundScan and use the method adopted by Brynjolfsson et al. (2003) to translate sales ranks into sales units. They fit sales and the sales rank to the following regression relationship:

$$\log(Sales) = \alpha + \beta \log(Rank) + \varepsilon \, , \tag{2}$$

where α determines the location of the sales curve and β determines the shape of the curve. Given that we are estimating the weekly digital track sales of an artist, we assume the β coefficient to be -0.87, same as what is provided in Brynjolfsson et al. (2003), and then obtain an estimate of 12.6 for α according to the sales data published in the Nielsen Company 2009 Year-End Music Industry Report (2010). These estimates indicate that the topmost artist has a digital track sale of 296,559 units per week and that an artist with a rank of 1,000 has a digital track sale of 728 units per week. Based on these estimates, we show the impacts of posting one additional bulletin or friend update on artists' digital track sales in Table 2.3. Depending on the sales rank of an artist, the increase in sales from the MySpace promotions can range from a few to a couple of thousands each week.

Actually, the PVAR model estimates also reveal whether artists' marketing actions on MySpace can be affected by music sales. The coefficients for sales rank (in logarithm) at lags 1 to 3 in the *Bulletins* and *FriendUpdates* equations in Table 2.2 are all insignificant at 5%, suggesting that artists' promotional activities are unlikely to be affected by their sales ranks. Therefore, we focus more on how MySpace promotional activities affect sales ranks in the following analyses and do not report the results of the second and third equations, unless necessary.

Table 2.2 Panel VAR Coefficient Estimates

Independent Variable	Dependent Variable		
	ln(SalesRank)	Bulletins	FriendUpdates
$\ln(SalesRank)_{i(t-1)}$	0.542** (0.012)	0.024 (0.040)	0.010 (0.028)
$\ln(SalesRank)_{i(t-2)}$	0.050** (0.009)	-0.035 (0.033)	0.005 (0.024)
$\ln(SalesRank)_{i(t-3)}$	0.046** (0.009)	-0.025 (0.030)	-0.030 (0.022)
$Bulletins_{i(t-1)}$	-0.008** (0.002)	0.374** (0.009)	0.029** (0.006)
$Bulletins_{i(t-2)}$	-0.0002 (0.002)	0.127** (0.009)	-0.011 (0.006)
$Bulletins_{i(t-3)}$	-0.0004 (0.002)	0.138** (0.008)	-0.004 (0.006)
$FriendUpdates_{i(t-1)}$	-0.011** (0.004)	-0.030* (0.013)	0.155** (0.009)
$FriendUpdates_{i(t-2)}$	-0.003 (0.003)	-0.015 (0.012)	0.081** (0.008)
$FriendUpdates_{i(t-3)}$	0.001 (0.003)	-0.038** (0.012)	0.047** (0.008)

Notes: Number of observations is 20,192; three lags of endogenous variables are used in the estimation; numbers in parentheses are standard errors; ** and * denote significance at 1% and 5%, respectively.

Table 2.3 Estimated Impacts of One Additional Bulletin or Friend Update on Weekly Digital Track Sales

Sales Rank	1	10	100	1,000	10,000	100,000
Weekly Digital Track Sales	296,559	40,005	5,396	728	98	13
Impact of Bulletin (0.8%)	2,372.5	320.0	43.2	5.8	0.8	0.1
Impact of Friend Update (1.1%)	3,262.1	440.1	59.4	8.0	1.1	0.1

Notes: This table presents the economic impacts of the coefficient estimates in Table 2.2. The same method can be applied to the coefficient estimates shown in other tables such as Tables 2.5, 2.6, 2.7, and 2.8.

2.4.2. Specification Tests

Lag selection is critical in estimating Panel VAR models as this determines how much information should be used and whether the instruments are valid. The GMM estimation of the Panel VAR model involves the use of dependent variables lagged two periods or more as instruments in the equations in their first-difference form. The Sargan test (Sargan 1958, 1959) is the most common specification test applied in this context (see Dahlberg and Johansson 2000). The Sargan statistic is the objective function evaluated at the estimated parameters when forming the overidentifying restriction test. Under the null hypothesis, this statistic follows a chi square distribution with a degree of freedom (df) equal to the number of instruments minus the number of parameters. To test specific hypotheses, the difference Sargan statistic is formed by estimating both the restricted and unrestricted models and then calculating the difference of the two Sargan statistics. Under the null hypothesis, this difference statistic follows a chi square distribution with $df = df_R - df_U$. In this study, we utilize the difference Sargan statistic to test the specification of each equation in the Panel VAR model. We start with a relatively large number of lags and gradually restrict the model by reducing the number of lags by one each time until it reaches zero. If the p-value is very small (<0.01), then we reject the null hypothesis and prefer the unrestricted model with the larger number of lags.

Table 2.4 Test of Lag Length

Dependent Variable	Lag reduction	Difference Sargan Statistic	Degree of Freedom	p-value
ln(SalesRank)	5→4	2.7	15	0.9998
	4→3	5.0	12	0.958
	3→2	4.7	9	0.860
	2→1	7.5	6	0.277
	1→0	314.1	3	0.000
Bulletins	5→4	301.6	15	0.000
	4→3	871.4	12	0.000
	3→2	504.9	9	0.000
	2→1	1073.7	6	0.000
	1→0	5462.8	3	0.000
FriendUpdates	5→4	40.3	15	0.000
	4→3	26.6	12	0.009
	3→2	38.6	9	0.000
	2→1	87.3	6	0.000
	1→0	336.7	3	0.000

In Table 2.4, we report the specification tests starting with 5 lags. The history of an artist’s bulletins and friend updates is only observable for as long as one month on MySpace, the influence of the promotions 5 weeks ago is thus negligible. Starting with the *ln(SalesRank)* equation, we see that the lag length can be reduced from five to four, as the difference Sargan statistic is only 2.7 and the p-value is close to 1. Continuing on, we find that the lag length can be further shortened to one, but not from one to zero (the difference Sargan statistic is 314.1 and the null can be rejected at any reasonable significance level). Turning to the *Bulletins* and *FriendUpdates* equations, we see that we reject all the null hypotheses and can not reduce the lag length as the p-values are very small. These results indicate that for the *ln(SalesRank)* equation, one lag is the best specification and that for the *Bulletins* and *FriendUpdates* equations, longer lags are preferred. Thus, from now on, we use only one lag for the *ln(SalesRank)* equation.

2.4.3. How Network Size Moderates Marketing Effects

Now we consider how different network sizes could moderate the relationship between artists' promotional activities and music sales. The underlying reasoning is that the effect of promotions depends on not only how hard artists promote themselves on social networking sites but also their network size, which determines the target size of their promotional activities. With a small network, artists' promoting messages can be received by only a few people, thus limiting the potential benefits of these broadcasting efforts. However, with a large network, a fair amount of effort in broadcasting may attract a huge amount of attention from music fans and eventually lead to a lot of music purchases by them.

To study this moderating effect of network size, we divide the sample into two equal-size groups based on the average number of friends for each artist. We then estimate the Panel VAR model for each group using one lag for the *ln(SalesRank)* equation and compare the results together with the whole sample in Table 2.5. To evaluate the consistence of the GMM estimator, we also compare our results with the within-group estimates from the two-way fixed effects models (artist and time effects). As shown in Table 2.5, the coefficient estimates of the within-group estimator in columns (1)-(3) are much lager than the corresponding ones of the GMM estimator in columns (4)-(5); thus, the within-group estimator for this dynamic panel model tends to overestimate the impact of MySpace promotional activities on music sales. However, the signs of the coefficients are all in the right directions.

Table 2.5 Coefficient Estimates for the ln(SalesRank) Equation

	Within-Group Estimator			GMM		
		Network Size			Network Size	
	All Artists (1)	≥Median (2)	<Median (3)	All Artists (4)	≥Median (5)	<Median (6)
$\ln(SalesRank)_{i(t-1)}$	0.680** (0.005)	0.734** (0.007)	0.593** (0.008)	0.548** (0.013)	0.665** (0.015)	0.442** (0.016)
$Bulletins_{i(t-1)}$	-0.014** (0.002)	-0.013** (0.002)	-0.012** (0.003)	-0.008** (0.003)	-0.010** (0.003)	-0.007 (0.005)
$FriendUpdates_{i(t-1)}$	-0.021** (0.003)	-0.026** (0.004)	-0.014 (0.004)	-0.008* (0.004)	-0.019** (0.004)	-0.003 (0.005)
Number of Obs.	20,192	10,112	10,080	20,192	10,112	10,080

Notes: Individual and time fixed effects are included, and only one lag of the endogenous variables is used, in the estimation according to the results shown in the specification tests; numbers in parentheses are standard errors; ** and * denote significance at 1% and 5%, respectively.

Focusing only on the GMM estimates, we have the following observations. First, when we study the artists with a relatively large network size (i.e., larger than or equal to the median network size), the coefficients for both bulletins and friend updates in column (5) are more negative (compared with the coefficients in column (4) when studying the whole sample) and statistically significant at 1% level. In particular, the impact of friend updates increases from 0.8% for the whole sample to 1.9% for the artists with a relatively large network size, while the impact of bulletins increases from 0.8% to 1%. This implies that the marketing effect of bulletins and friend updates on sales is much larger for the artists with more friends. Second, for the artists with a relatively small network size (i.e., smaller than the median), the coefficients for both bulletins and friend updates in column (6) become less negative and no longer significant at even 5% level. The opposing results exhibited by these two groups reveal that network size is an important factor in determining the relationship between artists' online promotional activities and music sales. We also divide the sample at the 25th and 75th percentiles and estimate the models for the top and bottom 25% of the data; the results are very similar to those presented here, ensuring that the conclusion is not affected by how we divide the sample.

2.5. Robustness Checks and Additional Analyses

Many artists, especially major labels, also promote their music through the traditional channels such as radio and television. To examine whether these social networking promotions have an incremental impact on music sales, we need to study both the traditional and MySpace promotions at the same time in the ideal case. However, due to the unavailability of the traditional channel data, we are not able to completely control for the impact of traditional channel marketing on sales ranks. Alternatively, we can use artists' new album releases as an additional control. This is because artists generally advertise their music intensely right around the time of a new album release. A second potential issue is that fixed effects models only enable us to control for time-invariant individual characteristics. If there are some time-varying individual characteristics that could influence both the sales rank and MySpace promotions, then we might be establishing a strong relationship that is nonexistent in reality. For example, one important factor that is usually unobservable, but could play a major factor in determining an artist's music sales, is his or her popularity among consumers over time. In short, there could be a concern that we might obtain spurious regression findings. Yet another potential issue is that companies may over advertise on social networks because of the low cost. Then, using only a linear representation may not be sufficient enough to model the complex relationship between artists' social networking activities and their music sales. In this section, we address the above potential issues to check the robustness of our results and gain further insights. First, we incorporate the new album releases into the PVAR model and see if our main results still hold. Second, we control for time-varying individual characteristics by including a measure of artist popularity and address the issue of spurious regression. Third, we test a model specification in which the sales rank is a nonlinear function of social networking activities.

2.5.1. New Album Releases

We collect data on the new album releases of each artist in our sample from the Amazon MP3 store. Amazon provides a list of all MP3 albums released by an artist on their profile page. Single song releases are also included in this list as single-song albums. We collect the list of all MP3 albums released during our study period and create a dummy variable named $NewRelease_{it}$, which is 1 if there is at least one album release from artist i in week t, and 0 if there is no release. It is possible that an artist releases multiple MP3 albums in the same week as one album can have different versions such as Explicit or Deluxe versions and they can be released on different days. During our study period of 32 weeks, 266 out of 631 artists had at least one album release.

Table 2.6 Coefficient Estimates for the ln(SalesRank) Equation after Controlling for New Album Releases

	GMM Estimator			
	(1)	(2)	(3)	(4)
$ln(SalesRank)_{i(t-1)}$	0.548** (0.013)	0.553** (0.013)	0.544** (0.013)	0.540** (0.013)
$Bulletins_{i(t-1)}$	-0.008** (0.003)	-0.009** (0.003)	-0.007* (0.003)	-0.007* (0.003)
$FriendUpdates_{i(t-1)}$	-0.008* (0.004)	-0.012** (0.003)	-0.005 (0.004)	-0.005 (0.004)
$NewRelease_{it}$		-0.119** (0.022)	-0.133** (0.006)	-0.145** (0.007)
$NewRelease_{i(t-1)}$			-0.237** (0.006)	-0.257** (0.006)
$NewRelease_{i(t-2)}$				-0.024** (0.007)

Notes: Number of observations is 20,192; individual and time fixed effects are included, and only one lag of the endogenous variables is used, in the estimation; numbers in parentheses are standard errors; ** and * denote significance at 1% and 5%, respectively; column (1) is the same as column (4) in Table 2.5.

To control for the impact of new album releases on the sales rank, we add the dummy variable $NewRelease_{it}$ as an exogenous variable to the basic Panel VAR model and estimate the new model using GMM. This is because new album releases are usually pre-scheduled events. Table 2.6 reports the coefficient estimates for the *ln(SalesRank)* equation with one lag of the endogenous variables. Usually, an artist's sales rank will have an instant decrease (i.e., sales increase) once a new album is released, and this effect can last for a few weeks. Thus, we test different specifications by including contemporaneous, and then lagged, new release dummies. For the sake of simplicity, we only report the GMM results up to two lags (we also test longer lags, but the sign and significance of the coefficients do not change).

From Table 2.6, we have the following findings. First, after controlling for the contemporaneous new album release, the coefficients for both bulletins and friend updates in column (2) are still negative and statistically significant at 1%. If we add the lagged new album releases too, then the coefficients for bulletins in columns (3) and (4) are only significant at 5% level, but the coefficients for friend updates in columns (3) and (4) become insignificant, though still negative. We expect that these promotional effects become smaller after accounting for the impact of new album releases. Second, the coefficients for contemporaneous and lagged new releases by the GMM estimator are negative and statistically significant at 1% in columns (2)-(4). Consistent with the typical observations in the music industry, new album releases have a huge impact on the sales rank – on average, the rank decreases by more than 10% in the week of a new release and more than 20% in the following week.

2.5.2. Artist Popularity

Since there is no direct measure of artist popularity, we resort to the Google Trends data, which measures the volume of consumer searches on specific artists. This data source represents the level of consumer interest in different artists over time; so, it can be a good indirect measure of artist popularity. Consumer searches on artists are also highly correlated with news events and media coverage of artists; therefore, by including the Google Trends measures, we also control for these previously unobservable factors that could have an influence on music sales.

Table 2.7 Panel VAR Coefficient Estimates

Independent Variable	Dependent Variable			
	ln(SalesRank)	Bulletins	FriendUpdates	SearchIndex
$\ln(\text{SalesRank})_{i(t-1)}$	0.642** (0.014)	-0.001 (0.057)	0.003 (0.039)	-0.006** (0.002)
$\text{Bulletins}_{i(t-1)}$	-0.013** (0.003)	0.279** (0.013)	0.043** (0.009)	0.001** (0.0005)
$\text{FriendUpdates}_{i(t-1)}$	-0.017** (0.005)	0.022 (0.019)	0.178** (0.013)	0.002** (0.001)
$\text{SearchIndex}_{i(t-1)}$	-0.431** (0.085)	-0.212 (0.366)	-0.092 (0.230)	0.366** (0.015)

Notes: Number of observations is 8,896; individual and time fixed effects are included, and only one lag of the endogenous variables is used, in the estimation; numbers in parentheses are standard errors; ** and * denote significance at 1% and 5%, respectively.

Google Trends provides data starting in January 2004, which is publicly available. We use the name of the artist as the search term and download the weekly search volume index for each artist. There is no absolute count of consumer searches on a specific artist name as the data is scaled based on the average search traffic of the term entered. However, Google Trends allows one to compare the popularity of two or more search terms over time. Hence, given the same benchmark, it will not matter if we only have the data on scaled search traffic. To collect the search volume indices on all the artists in our sample, we thus need to choose a benchmark artist so that the data for all artists are scaled based on the same search traffic and hence comparable. We first select the popular female artist Beyoncé as our benchmark and collect the search volume index data using this benchmark.

One potential hazard with this data source is that it is censored. The minimal value available for display is 0.01 (note that Google scales the search volumes over time so that the average search traffic for Beyoncé is 1.0); so, any value smaller than 0.01 but still larger than zero may be displayed as zero. This creates a censorship problem as the data for obscure artists may be inaccurate because we can not distinguish observed zero values from true zero values. To counter this problem, we limit our sample to only those artists who always have a positive search volume index compared to Beyoncé over the study period. This ensures that the search volume data is accurate for each artist. We are able to find 297 artists that satisfy this constraint. To be conservative, we also exclude the artists whose names are too general, such as Bond, CSS, Ivy, Panda, etc., because such general names may be confused with other terms, making the search volume index inaccurate (i.e., we cannot be sure if the user is searching for these artists or other unrelated things). This leaves us with 278 artists in total. The correlation between the search volume index and the log of sales rank is -0.278 for these 278 artists, which implies that highly ranked artists tend to attract more consumer searches.

To incorporate the search volume index into the basic model, we treat this new variable *SearchIndex* as endogenous and estimate a Panel VAR model with four equations. Each equation is a function of the lagged endogenous variables and other exogenous variables. To illustrate all the relationships between different variables, we report the GMM coefficient estimates for all four equations in Table 2.7. We have the following findings. First, in the *ln(SalesRank)* equation, the coefficient for bulletins is -1.3% and significant at 1% level, while the coefficient for friend updates is -1.7% and significant at 1% level. This means that even after controlling for consumer interest on artists, the impact of social network marketing on sales ranks is still significant. Second, the coefficient for search volume index in the *ln(SalesRank)* equation is also negative and statistically significant, indicating that more consumer searches lead to larger music sales. Third, the coefficients for *ln(SalesRank)* in the *Bulletins* and *FriendUpdates* equations are both insignificant, which is consistent with the previous result that the sales rank is unlikely to have an impact on artists' promotional activities. Fourth, in the *SearchIndex* equation, the coefficient for *ln(SalesRank)* is negative and statistically significant at 1%, suggesting that the volume of consumer searches next week will decrease when the sales rank of the artist increases (i.e., sales decrease) this week. Relating to the coefficient for search volume index in the *ln(SalesRank)* equation, we conclude that music sales and consumer searches influence each other. Finally, the coefficients for bulletins and friend updates in the *ln(SalesRank)* equation are positive and significant, implying that intense promotional activities in social networks can drive consumers' search volumes.

To address the concern that Beyoncé may be too popular to be used as a benchmark, we also download the data using Nickel Creek, an obscure music group with a unique name, as a benchmark. Our sample then increases to 320 artists, after excluding those with general names as before. The results remain consistent with those presented in Table 2.7 and thus are omitted here.

Table 2.8 Coefficient Estimates for the ln(SalesRank) Equation

	GMM Estimator	
	(1)	(2)
$\ln(SalesRank)_{i(t-1)}$	0.548** (0.013)	0.558** (0.012)
$Bulletins_{i(t-1)}$	-0.008** (0.003)	-0.016** (0.004)
$FriendUpdates_{i(t-1)}$	-0.008* (0.004)	-0.008 (0.006)
$Bulletins^2_{i(t-1)}$		0.0003* (0.0002)
$FriendUpdates^2_{i(t-1)}$		0.0001 (0.001)

Notes: Number of observations is 20,192; individual and time fixed effects are included, and only one lag of the endogenous variables is used, in the estimation; numbers in parentheses are standard errors; ** and * denote significance at 1% and 5%, respectively; column (1) is the same as column (4) in Table 2.5.

2.5.3. A Nonlinear Specification

To specify a nonlinear relationship between social networking activities and sales ranks, we add two quadratic terms, the squares of the number of bulletins and friend updates, as endogenous variables to the basic model. We can still fit a linear regression, although the sales rank is now nonlinear in the variables of social networking activities. The estimation results are presented in Table 2.8.

We find that the coefficient estimate for the quadratic term $Bulletins^2_{i(t-1)}$ is positive (0.0003) and statistically significant at 5% level. However, the coefficients of both $FriendUpdates_{i(t-1)}$ and $FriendUpdates^2_{i(t-1)}$ are insignificant. This suggests that the sales rank can be a convex function of the number of bulletins, and that there exists an optimal level of promotion using bulletins on MySpace. Setting the first order condition with respect to the number of bulletins to zero, we find that the optimal number of bulletins is roughly 27 in one week ($\frac{0.016}{2 \times 0.0003} = 26.7$). Thus, this number is the threshold above which the sales rank starts to increase (i.e., sales start to decrease). In our sample, we do observe that some artists are over advertising as the maximum number of bulletins per week is 48, which is far more than the optimal threshold.

2.6. Conclusion and Discussion

Because of its fast speed and low cost, broadcasting in social networks has evolved into a new form of internet marketing recently. On MySpace, artists regularly update their profile pages to keep music fans interested in their music. Their activities are automatically broadcasted to their friends by the bulletin board and activity stream features. Using the activity stream data from MySpace and the sales rank data from Amazon, we employ a panel vector autoregression model to quantify the effect of MySpace marketing on music sales. The proposed Panel VAR model utilizes time series data and is able to address the endogeneity problem among marketing activities, sales ranks, and other focal variables such as artist popularity. We find that marketing activities on social networking sites yield significant benefits only for the artists with many friends. To translate these effects into actual sales, we estimate that one additional bulletin or friend update by an artist can lead to an increase of as many as 3,000 of digital track sales each week.

This study provides important managerial implications. First, our results demonstrate that artists' marketing activities on MySpace have a significant impact on the sales of music. Although we have only studied MySpace and the music industry, the models we use and the insights we obtain can be easily applied to many other social networking sites and many other industries. In light of our findings, companies across industries should carefully plan their marketing activities on social networks, besides such activities in traditional channels. Second, building up the fan base and enlarging the network size is the key to the success of social network marketing. Not all firms that engage in social network marketing will be successful in their marketing efforts. The same amount of effort may yield very different payoffs depending on the size of the firm's virtual network. Firms should focus on attracting more people to join their networks at early stages of implementation and gradually enlarge the network size over time for successful social network marketing. To get a rough estimate of how large the network should be to enjoy the benefits, we divide the sample into four equal-size groups by splitting it at the 25th, 50th, and 75th percentiles of the average number of friends for each artist. We find that the coefficient estimates on MySpace activities for the first two groups are statistically insignificant and that those for the last two groups are statistically significant. Thus, we estimate that firms should at least aim for a network size of between 44,043 (50th percentile) and 121,672 (75th percentile). Third, we also find that the relationship between marketing activities on MySpace and sales ranks can be nonlinear. Companies should thus be careful that they do not overuse social network technology just because it is relatively inexpensive to adopt.

There are a number of ways for future researchers to extend the findings in this study. Since MySpace is generally the most popular social network for artists to market their music, we have chosen it to be our data source. However, it is likely that artists have a significant presence at other social networking sites such as Facebook and Bebo as well. Activities on these sites could also affect artists' sales ranks on Amazon. Due to data limitations, we are not able to study the interactions between music sales and the overall marketing intensity across different social networking sites. Future research is warranted in this area. The following are some interesting research questions for artists, record companies, as well as academics: How to effectively manage marketing activities across different social networking sites? How to allocate limited advertising budgets across different social networks? Similarly, this study use new album and song releases to control for artists' marketing campaigns carried out in traditional channels. If a researcher is able to obtain the actual data on traditional marketing activities, it would be interesting to study how marketing activities in traditional channels and those on social networks should be coordinated, as well as to compare the effectiveness of social network marketing and that of traditional marketing. Such results would provide more insights to artists and record companies on how to manage the tradeoffs between different types of marketing channels.

In summary, social network marketing is more of a two-way communication in nature, rather than a one-way flow of information. Its success lies in the engagement and interaction with customers. The goal of this study is to provide insights on the effectiveness of social network marketing and address the business implications of social networks. It is still early to claim that social network marketing is the future of marketing, and more research is necessary to understand this new and interesting phenomenon.

CHAPTER 3. THE ROLE OF SOCIAL MEDIA IN THE STOCK MARKET

3.1. Introduction

Finance is built on the premise that stock price changes represent the market's interpretation of "novel" information. What qualifies as "novel" information and how information disseminates through agents into security prices are, however, not well understood. This essay attempts to characterize a particular type of information dissemination and the information so transmitted. Specifically, we extract stock opinions from a popular social media website *Seeking Alpha* (www.seekingalpha.com); we study how these opinions are formed, how they relate to the views expressed in more traditional media outlets (i.e., the Wall Street Journal), and how they pertain to investor trading and security prices. Our setting allows us to produce fresh insights on information aggregation and price formation related to stocks.

Seeking Alpha is a popular social media platform for investors to voice their market opinions and exchange investment ideas. It differs from internet message boards and forum discussions in at least two ways. First, submitted opinions in the form of articles are reviewed by a panel and subject to editorial changes. Such a review process intends to improve the quality of published articles, but not to interfere with the authors' original opinions. Second, authors on Seeking Alpha are required to reveal their identity and positions in discussed stocks. They also have a genuine incentive to gain a large audience by increasing their network of followers, who later could become their clients or paid subscribers to their financial blogs.

To quantify and study the information disseminated through Seeking Alpha and examine the role of social media in the stock market, we use textual analysis.

Specifically, we build on prior literature (e.g., Antweiler and Frank 2004, Tetlock 2007, Engelberg 2008, Li 2008, and Tetlock, Saar-Tsechansky, and Macskassy 2008), which suggests that negative word classifications[9] measure the tone of a report, and use the word classification schemes complied by Loughran and McDonald (2010) to characterize the sentiments revealed in articles and the ensuing information exchange.

We find that the fraction of negative words (i.e., the number of negative words divided by the number of all words) contained in articles on Seeking Alpha negatively correlates with stock returns. This is true even after controlling for the media sentiment revealed in traditional outlets. In addition, the effect is stronger for articles that receive more comments by readers and presumably are followed by a wider audience, pointing to a causal link from financial blogs to stock returns.

3.2. Data

The dataset we study in this paper includes articles in two different media sources and the related financial information from companies. We first download all the articles that appeared between 2006 and 2009[10] on *Seeking Alpha* (SA). To categorize articles, SA allows authors to tag each article with one or more stock tickers. Single-ticker articles focus solely on one stock; so, it is (relatively) easy to extract the author's opinion on that specific stock ticker from these

---

[9] There are also other types of word classifications available, e.g., positive words, uncertainty words, etc. However, the finance literature suggests that negative words are more suitable for studying stock returns (see, for example, Tetlock 2007).

[10] Seeking Alpha was established in 2004, but in its first two years, it did not publish many articles.

articles. Multiple-ticker articles, on the other hand, discuss multiple stocks in the same report, which makes extraction of different opinions on each of the tagged stock tickers extremely difficult, if not impossible. Thus, we focus our attention on single-ticker articles, which comprise roughly one third of all articles. In total, 2,899 U.S. common stocks are covered by these single-ticker articles.

We then collect the data on these 2,899 stocks' news coverage in Wall Street Journal (WSJ) through the Factiva database. Since WSJ news articles are not tagged by company name or stock ticker, we formulate a search query to find matched news articles for each stock in 2006-2009. We start with each company's name from the Center for Research on Security Prices (CRSP) data as the search term and restrict that this formal company name has to show up at least once in the first 50 words of the news article. We also adjust the CRSP company name to match with Factiva's coding of company names, e.g., changing "Company" to "Co", "Intl" to "International", etc. Within our study period, if a company ever changed its name, we query all possible names and combine the search results for that stock (e.g., former Apple Computer, Inc. and current Apple, Inc.). To make sure that one news article mainly talks about only one company stock, we exclude those articles that can be possibly matched with two or more stock tickers in our sample.

We obtain stock price data from CRSP and companies' accounting information from Compustat. All variables are defined in Table 3.1, and their descriptive statistics are shown in Table 3.2. To get a picture of the companies covered by both traditional and social media, we also listed the top 15 companies that are most written about on WSJ and SA. This list is presented in Table 3.3. We can see that all the companies are big companies and they come from a diverse industry base such as Technology, Banking, Retailing, Automotive, etc.

Table 3.1 Variable Definition

Variable	Definition
t	Time period that can be one week or one day depending on the level of analysis.
Neg	Number of negative words divided by the number of total words.
$Article\_WSJ_t$	Number of WSJ news articles on a specific stock in time period t.
$LogWord\_WSJ_t$	Natural logarithm of the average number of words in each WSJ news article on a specific stock in time period t.
Neg_WSJ_t	Average value of Neg for all WSJ news articles on a specific stock in time period t.
$Article\_SA_t$	Number of single-ticker articles on a specific stock appearing in SA in time period t.
$LogWord\_SA_t$	Natural logarithm of the average number of words in each single-ticker SA article on a specific stock in time period t.
Neg_SA_t	Average value of Neg for all single-ticker SA articles on a specific stock in time period t.
$Report_t$	Dummy variable, set to one if the company made an earnings announcement in time period t and to zero otherwise.
$Return_t$	Return of a stock in time period t, which is the difference of the stock prices in time period t and t-1, divided by the stock price in time period t-1.

Table 3.2 Summary Statistics for Selected Sample

	Weekly Data (N=29,730)			Daily Data (N=36,808)		
Variable	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
$Article\_WSJ_t$	0.6748	0	0.9298	0.4944	0	0.5719
$LogWord\_WSJ_t$	5.8137	5.8972	0.7513	5.8332	5.9108	0.7781
Neg_WSJ_t	0.0202	0.0170	0.0161	0.0202	0.0167	0.0166
$Article\_SA_t$	1.0149	1	1.5505	0.7419	1	0.7852
$LogWord\_SA_t$	5.9457	5.9135	0.6648	5.9243	5.8833	0.6704
Neg_SA_t	0.0152	0.0132	0.0111	0.0155	0.0132	0.0116
$Report_t$	0.1862	0	0.3893	0.0446	0	0.2063
$Return_t$	0.0044	0.0005	0.1478	0.0006	0.0000	0.0690

3.3. Analysis and Results

We borrow research methods from studies in the finance literature (e.g., Tetlock 2007, Tetlock, Saar-Tsechansky, and Macskassy 2008, Loughran and McDonald 2010) that examine how qualitative information contained in news stories can impact firms' stock returns. While these studies concentrate on traditional media outlets, such as WSJ, we focus on the emerging social media, which allows us to extend prior studies, but also enables us to look at new questions and add fresh insights to the information systems, economics, and finance literatures.

Table 3.3 List of Top 15 Companies Covered by Both WSJ and SA

Weekly Data			Daily Data		
Company	# WSJ Articles	# SA Articles	Company	# WSJ Articles	# SA Articles
Google	310	902	Google	77	141
Microsoft	312	480	Apple	58	147
Apple	209	849	Citigroup	56	114
Citigroup	281	427	AIG	64	125
Wal-Mart	129	116	Microsoft	42	48
GE	130	162	Bank of America	25	40
AIG	150	191	Goldman Sachs	17	36
Bank of America	100	173	Yahoo	13	32
Boeing	76	56	Wal-Mart	12	16
Goldman Sachs	80	182	Lehman Brothers	36	45
Yahoo	57	150	CIT	14	33
Intel	48	54	Hewlett Packard	16	15
Ford	87	73	Morgan Stanley	15	11
Dell	41	102	GE	8	16
eBay	31	92	CFC	10	14

Note: Companies are ranked by the number of effective observations in the regression.

3.3.1. Relationship between stock market return and media sentiment

Following prior research (e.g., Loughran and McDonald 2010), we use the fraction of negative words to measure media sentiment toward a specific company stock. We regress stock returns on variables constructed from WSJ and SA articles and a dummy variable for companies' earnings announcement. The results are presented in Table 3.4. We conduct our analyses at both weekly and daily levels.

Model 1 examines the impact of WSJ sentiments on stock returns (at the weekly level). The coefficient estimate on the *Neg_WSJ* variable is -0.5059 and statistically significant at 1% level. This result is consistent with prior research. In terms of the economic impact, this coefficient suggests that, on average, a company's weekly stock return decreases by 0.5059% if the fraction of negative words in WSJ news articles increases by 1% in magnitude (i.e., one more negative word out of 100 words). Model 2 studies the impact of SA sentiments on stock returns (on a weekly level). The coefficient estimate on the *Neg_SA* variable is -0.9142 and statistically significant at 1% level. The number of words is also significant at the 5% level, indicating that SA article length is negatively associated with stock returns. Model 3 studies instances in which the company is covered by both WSJ and SA at the same time. We find that only the sentiment revealed from WSJ has a significant impact on the stock return at the weekly level, although the sign of the *Neg_SA* variable is still negative.

At the daily level, we observe similar results except for one major difference: The coefficients of both *Neg_WSJ* and *Neg_SA* are negative and statistically significant at 1% level[11]. This suggests that the articles posted on SA have a

[11] The coefficient estimate on *Neg_SA* is (slightly) larger than that on *Neg_WSJ* (-0.3111 vs. -0.2653).

comparable influence on stock returns as those appearing in WSJ when both media outlets cover the same company stocks at the same time. This shows that social media play an important role in the stock market.

Interestingly, the coefficient estimates on the *Report* variable in Model 2 and 5 are both negative and statistically significant at 5% level. This implies that stock returns of the companies covered in SA, on average, decrease when they make an earnings announcement.

3.3.2. Ongoing Analyses

We are currently conducting further analyses in the following two directions. First, the setup of the SA website makes it possible for us to study whether there could be a causality link between media sentiment and stock returns. SA has a commenting feature that enables readers to make comments on each article published on the website, which often sparks a dialogue between readers (and between readers and authors). These data are also available historically (e.g., we can count the number of comments made on each article one day after it was published). If media sentiment was causing the stock returns to change, then we would expect to see that those intensely discussed articles on average should have a stronger impact on stock returns.

Second, since the major readers on SA are small investors, we also hypothesize that the relationship between media sentiment and stock returns is stronger for stocks that are mainly held by small investors. If a company stock is mainly held by institutions, then it is less likely that the media sentiment revealed on SA will have a large impact on stock returns. However, the correlation between these two should be stronger and more significant for stocks whose shareholders are mainly small investors.

Table 3.4 Contemporaneous Regressions

	$Return_t$					
	Weekly Level			Daily Level		
	1	2	3	4	5	6
$Article\_WSJ_t$	-0.0025 (-0.69)	——	-0.0058 (-1.24)	0.0004 (0.07)	——	-0.0061 (-0.53)
$LogWord\_WSJ_t$	-0.0011 (-0.70)	——	-0.0039 (-1.12)	0.0004 (0.77)	——	0.0010 (0.58)
Neg_WSJ_t	-0.5059 (-6.25)	——	-0.7072 (-5.02)	-0.1582 (-5.89)	——	-0.2653 (-3.17)
$Article\_SA_t$	——	-0.0001 (-0.03)	0.0022 (0.99)	——	-0.0011 (-0.72)	0.0008 (0.32)
$LogWord\_SA_t$	——	-0.0033 (-2.01)	-0.0013 (-0.31)	——	-0.0002 (-0.37)	0.0010 (0.49)
Neg_SA_t	——	-0.9142 (-6.72)	-0.3491 (-1.17)	——	-0.3408 (-7.23)	-0.3111 (-2.59)
$Report_t$	0.0011 (0.33)	-0.0086 (-2.27)	-0.0057 (-1.03)	-0.0003 (-0.06)	-0.0035 (-1.98)	0.0043 (0.75)
Observations	14696	19789	4907	17039	22856	3140
Clusters	207	207	207	1003	1003	930
Adjusted R^2	0.0031	0.0045	0.0049	0.0021	0.0029	0.0079

Note: This table shows the relationship between stock return and contemporaneous WSJ and SA variables, including the number of articles, natural logarithm of the number of words, and media sentiment measured by the proportion of negative words. Whether the company issued an earnings report or not is also included in the independent variable list. The same regressions are run on both weekly and daily data. Standard errors are calculated after clustering by time. The robust t-statistics are in parentheses, and the coefficients in bold are statistically significant at 5% level.

3.4. Conclusion

This essay examines the role of social media in the stock market and provides evidence that implies a causality link between media sentiment and stock returns. We show that social media sentiment measured by the fraction of negative words has a significant impact on stock returns and that this result still holds after controlling for traditional media sentiment. We also find that those intensely discussed articles on SA tend to have a stronger and more significant impact on stock returns.

CHAPTER 4. CONCLUSIONS

The rapid evolution of information technology has recently stimulated more interest in studying social media. This dissertation draws upon the literature of marketing and finance and examines how social media is playing an important role from two different perspectives.

The first essay investigates how social networks, or social media in general, can be employed as a marketing channel for both companies and individuals. We use the music industry as an example and quantify the value of social network marketing for record companies and music artists. We also find that an artist's social network size is moderating the effect of social networking activities on music sales. Despite the fact that social network marketing can be relatively cheaper, we caution that over-marketing in this channel yields diminishing or even opposite effect.

The second essay turns to the stock market and studies the role of both social media and traditional media. We first establish the significant correlation between social media sentiment and stock returns, even after taking the traditional media sentiment into account. Then we utilize the commenting data available on SA to study the causality link between media sentiment and stock returns. The analysis on how the ownership of stocks influences the focal relationship provides additional implications on this.

BIBLIOGRAPHY

Ahn, S.C., P. Schmidt. 1995. Efficient estimation of models for dynamic panel data. *Journal of Econometrics* **68**(1), 5-27.

Ahn, S.C., P. Schmidt. 1997. Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation. *Journal of Econometrics* **76**(1-2), 309-321.

Antweiler, W., M. Z. Frank. 2004. Is all that talk just noise? The information content of Internet Stock message boards, *Journal of Finance* 59, 1259-1294.

Arellano, M. 2003. *Panel Data Econometrics.* Oxford University Press, Oxford.

Arellano, M., S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* **58**(2) 277-297.

Arellano, M., O. Bover. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* **68**(1) 29-51.

Bass, F. M. 1969. A simultaneous equation regression study of advertising and sales of cigarettes. *Journal of Marketing Research* **6**(3) 291-300.

Baker, B. 2006. MySpace music marketing: How to promote & sell your music on the world's biggest networking web site. http://www.bob-baker.com/myspace/music.html.

Bessler, D. A. 1984. An analysis of dynamic economic relationships: an application to the U.S. hog market. *Canadian Journal of Agricultural Economics* **32**(1) 109-124.

Binder, M., C. Hsiao, M. H. Pesaran. 2005. Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* **21**(4) 795-837.

Blundell, R., S. R. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* **87**(1) 115-143.

Brynjolfsson, E., Y. J. Hu, M. D. Smith. 2003. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science* **49**(11) 1580-1596.

Chen, P., S. Dhanasobhon, M. D. Smith. 2007. An analysis of the differential impact of reviews and reviewers at Amazon.com. *Proceedings of the 28th International Conference on Information Systems,* Montreal, Quebec, Canada.

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

Dahlberg, M., E. Johansson. 2000. An examination of the dynamic behaviour of local governments using GMM bootstrapping methods. *Journal of Applied Econometrics* **15**(4) 401-416.

Dekimpe, M. G., D. M. Hanssens. 1995. The persistence of marketing effects on sales. *Marketing Science* **14**(1) 1-21.

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

Dewan, S., J. Ramprasad. 2009. Chicken and egg? Interplay between music blog buzz and album sales, *PACIS 2009 Proceedings,* Paper 87.

Engelberg, J. 2008. Costly information processing: Evidence from earnings announcement, Working paper, Northwestern University.

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

Givon, M., D. Horsky. 1990. Untangling the effects of purchase reinforcement and advertising carryover. *Marketing Science* **9**(2) 171-187.

Gonsalves, A. 2009. Dell makes $3 million from Twitter-related sales. *InformationWeek*, June 12.

Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**(3) 424-438.

Hamilton, J. 1994. *Time series analysis*. Princeton University Press, Princeton, NJ.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* **50**(4) 1029-1054.

Hanssens, D. M. 1980. Market response, competitive behavior and time-series analysis. *Journal of Marketing Research* **17**(4) 470-485.

Hayashi, F. 2000. *Econometrics*. Princeton University Press, Princeton, NJ.

Holtz-Eakin, D., W. Newey, H. S. Rosen. 1988. Estimating vector autoregressions with panel data. *Econometrica* **56**(6) 1371-1395.

Horsky, D., L. S. Simon. 1983. Advertising and the diffusion of new products. *Marketing Science* **2**(1) 1-17.

Learmonth, M. 2010. New Twitter rules clamp down on ad networks. Advertising Age, May 24.

Li, F., 2008. Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45, 221-247.

Li, X., L. M. Hitt. 2008. Self-selection and information role of online product reviews. *Information Systems Research* **19**(4) 456-474.

Lin, M., N. R. Prabhala, S. Viswanathan. 2009. Can social networks help mitigate information asymmetry in online markets? *Proceedings of the 30th International Conference on Information Systems,* Phoenix, Arizona.

Loughran, T., B. McDonald. 2010. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance*, forthcoming.

McCarty, T. A., S. J. Schmidt. 1997. A vector-autoregression analysis of state-government expenditure. *American Economic Review* **87**(2) 278-282.

Nickell, S. 1981. Biases in dynamic models with fixed effects. *Econometrica* **49**(6) 1417-1426.

Nielsen Company 2009 Year-End Music Industry Report. 2010. 2009 top ten selling digital artists. (Available at http://blog.nielsen.com/nielsenwire/wp-content/uploads/2010/01/Nielsen-Music-2009-Year-End-Press-Release.pdf.)

Owyang, J. 2008. Social network stats: Facebook, MySpace, reunion (Jan, 2008). http://www.web-strategist.com/blog/2008/01/09/social-network-stats-facebook-myspace-reunion-jan-2008/.

Pauwels, K., I. Currim, M. G. Dekimpe, D. M. Hanssens, N. Mizik, E. Ghysels, P. Naik. 2004. Modeling marketing dynamics by time series econometrics. *Marketing Letters* **15**(4) 167-183.

Sargan, J. D. 1958. The estimation of economic relationships using instrumental variables. *Econometrica* **26**(3) 393-415.

Sargan, J. D. 1959. The estimation of relationships with autocorrelated residuals by the use of the instrumental variables. *Journal of the Royal Statistical Society, Series B (Methodological)* **21**(1) 91-105.

Silver, C. 2009. Organized chaos: Viral marketing, meet social media. Retrieved February 23, 2010, http://www.wired.com/epicenter/2009/10/organized-chaos-2/.

Sims, C. A. 1980. Macroeconomics and reality. *Econometrica* **48**(1) 1-48.

Sims, C. A. 1992. Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review* **36**(5) 975-1000.

Stock, J. H., M. W. Watson. 2001. Vector Autoregressions. *Journal of Economic Perspectives* **15**(4) 101-115.

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

Tetlock, P. C., M. Saar-Tsechansky, S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance* 63, 1437-1467.

Trusov, M., R. E. Bucklin, 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.

Vincent, F. 2010. *MySpace for Musicians*. Course Technology PTR.

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

APPENDIX. GMM Estimation of Panel VAR models

Assume that we are estimating a Panel VAR model with *p* lags. Let $\mathbf{y}_{it}$ be an $m \times 1$ vector of endogenous variables for artist *i* at time *t*. Without considering exogenous variables, the model can be written as

$$\mathbf{y}_{it} = \mathbf{\Phi}_1 \mathbf{y}_{i,t\text{-}1} + \mathbf{\Phi}_2 \mathbf{y}_{i,t\text{-}2} + \mathbf{...} + \mathbf{\Phi}_p \mathbf{y}_{i,t\text{-}p} + \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{it} \tag{A1}$$

The main step involved in a GMM estimator is to identify the orthogonality conditions for instrumental variables. Arellano and Bond (1991) propose the standard GMM estimator that employs lagged level variables (specifically, dependent variables which are lagged two or more periods) as instruments that are orthogonal to the error term of the first-differenced form of the model. First, we eliminate the individual effect $\boldsymbol{\mu}_i$ by first-differencing equation (1),

$$\Delta \mathbf{y}_{it} = \mathbf{\Phi}_1 \Delta \mathbf{y}_{i,t\text{-}1} + \mathbf{\Phi}_2 \Delta \mathbf{y}_{i,t\text{-}2} + \mathbf{...} + \mathbf{\Phi}_p \Delta \mathbf{y}_{i,t\text{-}p} + \Delta \boldsymbol{\varepsilon}_{it} \tag{A2}$$

for $t = 2,3,...,T.$ The orthogonality conditions then can be written as

$$E[(\Delta \mathbf{y}_{it} - \sum_{j=1}^{p} \mathbf{\Phi}_j \Delta \mathbf{y}_{i,t\text{-}j}) \cdot \mathbf{q}'_{i,t}] = \mathbf{0} \qquad \text{for } t = p+1, p+2, ...T, \tag{A3}$$

where $\mathbf{q}_{it}$ is the $m(t-2) \times 1$vector of instruments defined by

$$\mathbf{q}_{it} = (\mathbf{y}'_{i1}, \mathbf{y}'_{i2}, ..., \mathbf{y}'_{i,t-2})'. \tag{A4}$$

We can rewrite equation (A3) in stacked form as

$$E[\mathbf{Q}'_i(\mathbf{\Delta Y}_i - \sum_{j=1}^{p} \mathbf{\Delta Y}_{i,\text{-}j} \mathbf{\Phi}_j)] = \mathbf{0}\ , \tag{A5}$$

where $\mathbf{Q}'_i$ is a matrix of dimension $\frac{m(T+p-2)(T-p-1)}{2} \times (T-p-1)$ constructed from the vectors of instruments,

$$\mathbf{Q}'_i = \begin{pmatrix} \mathbf{q}_{i,p+2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{q}_{i,p+3} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{q}_{i,T} \end{pmatrix}, \tag{A6}$$

and $\mathbf{\Delta Y}_i$ and $\mathbf{\Delta Y}_{i,-j}$'s are $(T-p-1) \times m$ matrices defined as

$$\mathbf{\Delta Y}_i = (\mathbf{\Delta y}'_{i,p+2}, \mathbf{\Delta y}'_{i,p+3}, ..., \mathbf{\Delta y}'_{i,T})', \ \mathbf{\Delta Y}_{i,-1} = (\mathbf{\Delta y}'_{i,p+1}, \mathbf{\Delta y}'_{i,p+2}, ..., \mathbf{\Delta y}'_{i,T-1})', \ldots$$

and $\mathbf{\Delta Y}_{i,-p} = (\mathbf{\Delta y}'_{i,2}, \mathbf{\Delta y}'_{i,3}, ..., \mathbf{\Delta y}'_{i,T-p})'$.

By solving the first-order condition of (A5), the standard GMM estimator is expressed as

$$\hat{\boldsymbol{\varphi}}_{GMM} = (\mathbf{S}'_{\mathbf{ZX}}\mathbf{A}^{-1}\mathbf{S}_{\mathbf{ZX}})^{-1}\mathbf{S}'_{\mathbf{ZX}}\mathbf{A}^{-1}\mathbf{S}_{\mathbf{Zy}}, \tag{A7}$$

where

$$\mathbf{S}_{\mathbf{ZX}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}'_i\mathbf{X}_i, \ \mathbf{S}_{\mathbf{Zy}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}'_i\mathbf{y}_i, \ \mathbf{A} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}'_i\mathbf{\Omega}\mathbf{Z}_i,$$

$$\mathbf{Z}'_i = \mathbf{Q}'_i \otimes \mathbf{I}_m, \ \mathbf{X}'_i = [\mathbf{\Delta Y}_{i,-1} \quad \mathbf{\Delta Y}_{i,-2} \quad \ldots \quad \mathbf{\Delta Y}_{i,-p}] \otimes \mathbf{I}_m, \ \mathbf{y}_i = \mathrm{vec}(\mathbf{\Delta Y}'_i),$$

and $\mathbf{\Omega} = \mathbf{V} \otimes \mathbf{I}_m = \begin{pmatrix} 2 & -1 & 0 & & & & \\ -1 & 2 & -1 & & & 0 & \\ & & & \ddots & & & \\ & 0 & & & -1 & 2 & -1 \\ & & & & 0 & -1 & 2 \end{pmatrix} \otimes \mathbf{I}_m$.

A consistent estimate of the variance matrix of $\hat{\boldsymbol{\varphi}}_{GMM}$ can be obtained as

$$\widehat{Var}(\hat{\boldsymbol{\varphi}}_{GMM}) = (\mathbf{S}'_{\mathbf{ZX}}\mathbf{A}^{-1}\mathbf{S}_{\mathbf{ZX}})^{-1}\mathbf{S}'_{\mathbf{ZX}}\mathbf{A}^{-1}\mathbf{\Psi}\mathbf{A}^{-1}\mathbf{S}_{\mathbf{ZX}}(\mathbf{S}'_{\mathbf{ZX}}\mathbf{A}^{-1}\mathbf{S}_{\mathbf{ZX}})^{-1}, \tag{A8}$$

where $\mathbf{\Psi} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}'_i\hat{\mathbf{\Omega}}\mathbf{Z}_i$, $\hat{\mathbf{\Omega}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{e}}_i\hat{\mathbf{e}}'_i$, and $\hat{\mathbf{e}}_i = \mathrm{vec}([\mathbf{\Delta Y}_i - \sum_{j=1}^{p}\mathbf{\Delta Y}_{i,-j}\hat{\mathbf{\Phi}}_{j,GMM}]')$.

Note that $\mathbf{\Omega}$ is a weighting matrix that can be replaced with some alternatives, such as $\mathbf{\Omega}_1 = \frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{e}}_i\hat{\mathbf{e}}'_i$ or $\mathbf{\Omega}_2 = \hat{\mathbf{e}}_i\hat{\mathbf{e}}'_i$, where $\hat{\mathbf{e}}_i = \mathrm{vec}([\mathbf{\Delta Y}_i - \sum_{j=1}^{p}\mathbf{\Delta Y}_{i,-j}\hat{\mathbf{\Phi}}_j]')$, and the $\hat{\mathbf{\Phi}}_j$'s are initial coefficient estimates by specifying an arbitrary weighting matrix $\mathbf{\Omega}_a$. In this case, the variance matrix of the estimator is simplified to $\frac{1}{N}(\mathbf{S}'_{\mathbf{ZX}}\tilde{\mathbf{A}}^{-1}\mathbf{S}_{\mathbf{ZX}})^{-1}$, where $\tilde{\mathbf{A}}$ varies depending on the chosen weighting matrix.