Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2009 Proceedings

International Conference on Information Systems (ICIS)

2009

MySpace Killed the Radio Star? The Impact of Online Sampling on Song Sales

Ramnath K. Chellappa
Emory University, ram@bus.emory.edu

Conny Chen
Emory University, conny.chen@emory.edu

Follow this and additional works at: http://aisel.aisnet.org/icis2009

Recommended Citation

Chellappa, Ramnath K. and Chen, Conny, "MySpace Killed the Radio Star? The Impact of Online Sampling on Song Sales" (2009). *ICIS 2009 Proceedings*. 157.
http://aisel.aisnet.org/icis2009/157

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2009 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

MYSPACE KILLED THE RADIO STAR?

THE IMPACT OF ONLINE SAMPLING ON SONG SALES

Ramnath K. Chellappa and Conny Chen
Goizueta Business School, Emory University
1300 Clifton Road
Atlanta, GA 30322
ram@bus.emory.edu, conny.chen@emory.edu

Abstract

The recent emergence of online feedback and sampling mechanisms has created new avenues where consumers can learn about information goods such as books, music, and movies. An increasing number of studies evaluate the impact of these technological innovations on purchasing behavior. These studies often face estimation issues such as unobserved heterogeneity and simultaneity. This paper reviews the strategies commonly implemented to address various sources of endogeneity and proposes the use of a difference- and system-Generalized Method of Moments (GMM) estimator. The paper also stresses the need to consider the pattern of sales over time. As an empirical example, we explore the relationship between individual track sales and sampling of songs on the radio and MySpace. Our results suggest that while radio exposure continues to be an important predictor of song sales, online sampling of songs has a nearly equivalent effect on sales.

Keywords: Music industry, online sampling, digital goods, Generalized Method of Moments

Introduction

The digitization of media and other technological developments have created new avenues where consumers can learn about information goods such as books, music, movies, and software. With the emergence of online feedback and sampling mechanisms, buyers are able to read user reviews or to preview music, movie, or book samples to learn about a product before making a purchase decision. There has been increasing interest among researchers in information systems and marketing in evaluating the impact of these technological innovations on consumer purchasing behavior. Recent studies have investigated the influence of online word-of-mouth (WOM) on sales of books (Chevalier and Mayzlin 2006) and box office revenues (Duan et al. 2008; Liu 2006), and the effects of album downloads on music sales (Bhattacharjee et al. 2007; Liebowitz 2006, 2008a; Oberholzer-Gee et al. 2007).

In this paper, we discuss two estimation issues that arise in examining the effects of online activity on sales: simultaneity and unobserved heterogeneity. First, there is a potential simultaneity issue between online activity and sales. For example, the volume of online word-of-mouth has been shown to have a positive effect on box office revenues in the movie industry (Duan et al. 2008; Liu 2006); however, a greater level of ticket sales may lead to more users posting online reviews. Failure to account for this feedback mechanism between WOM and sales will yield inaccurate estimates. Second, it is likely that both user activity and sales are influenced by unobservable individual-specific characteristics. In this case, correlation between the individual effects and user activity will cause OLS estimates to be biased. These individual effects are difficult to address in cross-section data, and necessitates the use of an exogenous instrument. But if these characteristics are constant over time, for instance, qualities such as genre or debut album, the problem of unobserved heterogeneity can be addressed in panel data by using fixed effects or first-differencing.

We review the various strategies that have been implemented in recent work to address these issues: among them, difference-in-differences (Chevalier and Mayzlin 2006), simultaneous equations (Duan et al. 2008), and fixed effects with instrumental variables (Oberholzer-Gee and Strumpf 2007). While these methods may address specific endogeneity concerns, one aspect that is often overlooked is the presence of time-varying unobserved heterogeneity. For example, the popularity of a musical artist will likely influence both the number of illegally downloaded and legitimately purchased albums. Artist popularity may be changing over time; for instance, if an artist is featured on a television program, boosting downloads and sales. While standard panel data techniques can address individual fixed effects, such methods will yield inconsistent estimates in the presence of time-varying effects.

Moreover, in addition to issues of endogeneity, it is also necessary to consider the pattern of sales over time in developing an empirical specification. Demand for information goods like movies, books, and music usually peaks upon release and declines rapidly in the weeks following. Figure 1 presents an example of the typical exponential decaying sales pattern of song sales for Kanye West's *Stronger.* Both the level of sales and the number of user plays of the song display this pattern of significant decay. Given this consistent declining pattern of sales over time, it is necessary to address the differences in release dates across titles. Figure 2 depicts the timeline for selected song sales in our sample. There is a notable drop in sales following the release date for each song. Because of the differences in release dates across titles, if we look at a given week, one title may be just released and at its peak of sales, while another title may be at the end of its life cycle with very low sales. In order to accurately assess the impact of online activity on sales across titles, we need to somehow account for these differences.

Our objective is to explicitly address these issues to identify the effect of user activity on sales. We propose the use of the difference- and system- Generalized Method of Moments (GMM) estimator as a possible approach to address time-varying unobserved heterogeneity and other endogeneity concerns. We present an empirical example in the context of the music industry and explore the relationship between sampling of songs on MySpace artist pages and on the radio and song sales. Proposed by Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998), this estimation methodology was developed to address time-varying individual effects in panel data models with small numbers of time-series observations. In addition, this method allows for consistent estimation of dynamic panel data, in which lagged values of the dependent variable are included as regressors. While extensively used in the economics literature, use of difference- and system-GMM estimation has been relatively limited in information systems research. In order to account for the consistent decay pattern in sales, we follow a recent study by Hendricks and Sorensen (2009) by anchoring our data set to the week of album release. The inclusion of indicator variables for the weeks following release in our specification allows us to capture the exponential decline in sales. We also allow for persistence in the level, which further necessitates the use of dynamic panel data techniques.

Our paper is different from previous studies that examine the effects of user online activity on sales in several respects. First, we explicitly consider the impact of time-varying unobserved heterogeneity. Second, previous studies do not fully consider the effect of title release and the dynamic pattern of sales in their estimation strategies. Third, we consider the effects of online sampling in addition to the impact of traditional sampling methods such as radio, which has not yet been included in previous work examining online activities. Finally, the move towards digital distribution of content has enabled the sale of individual song tracks through retailers such as iTunes and Amazon. Previous work on the music industry has so far only examined the effects of consumer behavior on album sales. Our study uses song-level data, which allows us to examine the effect of song sampling directly on song sales without confounding factors involved with album purchases.

In the next section, we discuss the empirical context. Following this, we review several common estimation strategies employed in examining the relationship between online activity and sales. Next, we describe the data used in the empirical analysis. In the following section, we present the empirical methodology and outline our estimation approach and discuss our results thereafter. The final section concludes and provides recommendations for researchers facing issues of time-varying unobserved heterogeneity.

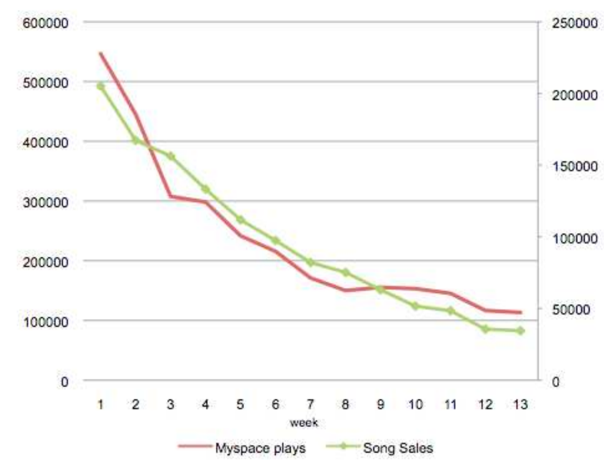

Figure 1. Pattern of song sales and MySpace plays for Kanye West's *Stronger*

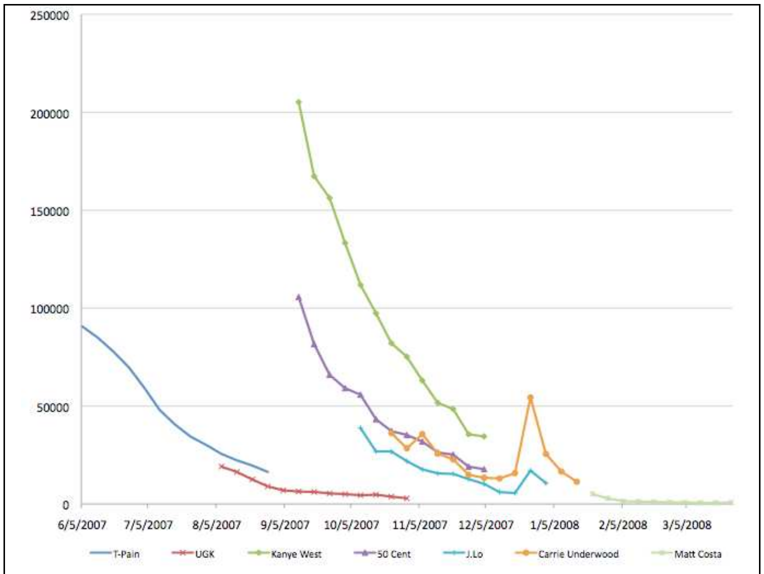

Figure 2. Timeline of song sales for selected artists

Sampling and the music industry

In the music industry, online sampling has become an increasingly important way for users to try out new music. New technologies enable listeners to sample songs online through embedded music players on musicians' websites or at social networks such as MySpace or last.fm. These sites allow users to listen to available songs on demand. As music is a hedonic good whose valuation depends primarily on the user's experience of the product rather than on specific attributes (Moe and Fader 2001), these online sampling mechanisms provide an alternative avenue for users to better evaluate music before making a purchasing decision.

Before the rise of online sampling, consumers typically learned about new music through radio or video broadcasts. Still today, record labels are known to make large payments in order to influence and promote radio broadcasts of their music (Coase 1979, Montgomery and Moe 2000, Liebowitz 2004). However, because of these large marketing costs, labels typically choose to promote in this way only those acts that have a large potential audience (Duchêne and Waelbroeck 2006). These large promotional costs have been a significant barrier to entry for lesser-known artists. However, the Internet and online sampling have enabled these up-and-coming artists to reach wide audiences at a modest cost without the backing of a major label. As a result, online sampling may have greater benefits for these newer artists (Gopal et al. 2006). Because established artists already carry with them a high degree of recognition, as well as significant promotional support from the music labels, buyers may therefore face less uncertainty regarding the valuation of music from these well-known musicians. In contrast, because online sampling encourages users to try out music they might not have otherwise purchased, the online availability of music for lesser-known acts may help users overcome this valuation uncertainty. Indeed, user sampling of songs on MySpace has been instrumental in launching the career of several emerging artists, including Lily Allen, Colbie Caillat, and Taylor Swift.

In this study, we examine the influence of online sampling of songs at MySpace on music sales. MySpace has become the preeminent destination for music, featuring over 8 million artist pages. We track the number of times a song is played at artists' MySpace Music pages. Because songs are played upon user demand, this metric gives us a direct measure of overall listener interest in a particular song. This is in contrast to radio, where listeners can choose only which station to listen to and have limited influence over the songs played at any given time. We also include weekly measures of radio play to assess the impact of this traditional sampling mechanism in addition to online activity. However, the number of times a song is played on the radio only gives a rough indication of the potential audience exposure to the song, as opposed to genuine interest in the song. The impact of radio broadcasts is likely to differ from online sampling, as radio airplay is heavily influenced by the promotional activities of the music labels and does not directly reflect the level of user interest (Bhattacharjee et al. 2006). We assess the impact of both sampling mechanisms on music sales. Moreover, we examine the effects of sampling of songs directly on song sales, in contrast with earlier work that has used only album-level data.

Prior literature

The results of previous research on the impact of radio listening on albums sales has been mixed. Using aggregate-level data on album sales, Liebowitz (2007) exploits the variation in the average listener time spent listening to radio across cities to identify a negative impact of radio play on industry-wide music sales. Using similar data, later work by Dertouzos (2008) reports that radio play significantly increases music sales, attributing the difference in results to earlier failure to properly account for market-level factors. While the debate on the effect of radio play on overall music sales continues, these studies uses aggregate data to identify broad industry-wide trends. There have been no definitive findings on the effects of radio play of specific album titles on sales. This may be due in part to the difficulties in accurately identifying the impact due to issues of endogeneity and simultaneity when estimating a model at the title level. In the case of radio play, it is likely that both album sales and radio play are both influenced by artist popularity. Because we cannot directly observe artist popularity, correlation between this title-specific unobserved characteristic and radio play will cause estimation by OLS to be biased. We must also consider the issue of simultaneity. Consumers may decide to purchase an album after hearing it on the radio. On the other hand, radio stations may decide to play songs that are popular and have enjoyed high sales. Failure to account for this feedback loop between radio play and sales may yield misleading results.

Research examining the effects of online activity such as online WOM, illegal downloading, and sampling on sales at the title-specific level are plagued with similar endogeneity problems. Recent studies have employed various strategies to address the estimation issues raised by the specific nature of their data. Chevalier and Mayzlin (2006) consider the effects of user reviews on book sales for two retailers, Amazon and Barnes and Noble. Because they have cross-section data at very few points in time, they exploit the variation in volume of user reviews across sites over time to identify the effect of online WOM. They note that standard OLS estimates will be biased due to unobserved individual book-site characteristics. Using a second cross-section at a later date, they eliminate this book-site fixed effect by evaluating the difference over time the difference in activity across sites. This "difference-in-differences" approach addresses the issue of unobserved heterogeneity that is constant over time. However, due to data limitations, it is unable to address issues of time-varying heterogeneity and simultaneity.

Duan et al. (2008) address the simultaneity problem in an examination of the dynamic effects of online WOM on box office revenues in the movie industry. They specifically model the feedback mechanism between WOM and sales with a system of interdependent dynamic equations. Estimation of this system of simultaneous equations by 3SLS alleviates the simultaneity issue. To address time-invariant heterogeneity, they include movie fixed-effects. This study also fails to adequately account for time-varying heterogeneity and the decay pattern and timing of sales. Because of the set-up of the simultaneous equations, the inclusion of lagged regressors captures to some extent the decay pattern, but this treatment constrains the decline to be constant over time. The other issue relates to their handling of the timing of movie releases for their sample. Using panel data, they include a variable for the number of days since release. While the number of days of release will account for the differences to some degree, the estimates are likely to be biased by the specific titles released in a given week.

There has yet to be an empirical analysis of online song sampling and sales in the music industry, although the literature examining the impact of illegal downloading of music on album sales is similar both in motivation and in the estimation issues faced. Past work on piracy provides evidence that users download music in order to learn about music that they might not have other wise purchased and to discover and experiment new genres or artists (Gopal et al. 2006; Peitz and Waelbroeck 2005, 2006; Rob and Waldfogel 2006). The overall effect of piracy on sales depends on whether the substitution effect, in which a downloaded copy displaces the sale of a legitimate album, or sampling effect dominates, where users obtain files in order to learn about music that they later purchase. While most studies find a negative impact of downloading on sales, Oberholzer-Gee and Strumpf (2007) fail to find a significant effect of piracy on album sales. They explicitly address estimation issues of unobserved heterogeneity by including album fixed effects. They also note that since artist popularity is likely to be varying over time, the presence of time-varying unobserved heterogeneity will bias OLS estimates. To account for this, they rely on the number of German children on vacation as an instrumental variable to identify the impact of downloading. Although the validity of this particular instrument has come into question (see Liebowitz 2007, 2008), the use of valid instrumental variables will account for both types of heterogeneity and simultaneity. However, this study also fails to account for the difference in release date and the declining pattern of sales across albums in their specification. They observe download and sales activity over a 17-week period, without accounting for the time since release of an album. While they include week fixed effects, this only captures any factors that occur during a calendar week, and is common across albums. Unless all the albums in their sample are released on the same date, this specification does not capture the decay pattern in sales.

While these studies address the issues of unobserved heterogeneity and simultaneity to varying degrees, our approach aims to address all of these estimation issues. There are several significant differences between this paper and prior studies. The use of panel data allows us to control for individual song-specific fixed effects. The difference- and system- GMM estimator controls for the endogeneity of our regressors by first-differencing to remove time-invariant fixed effects and uses appropriate instruments to account for any remaining endogeneity and simultaneity issues. Finding valid exogenous instruments is always a challenge, but particularly so for panel data analysis at the title-level, since most external instruments that are correlated with the endogenous regressors specific to each title will likely be correlated with the dependent variable as well. While not strictly exogenous, our assumption that our regressors are predetermined enables us to use lagged values as valid instruments. Finally, we capture the dynamic nature of the pattern of sales with several techniques. First, we anchor our data set to the date of release so that week 1 represents the week of release, week 2, the week after release, and so on. We include week indicator variables to capture the decay over time. In addition, for several specifications we also include lagged sales as a regressor in order to capture the persistence in sales over time.

Data

Our data set contains weekly measures of MySpace plays, radio plays, and song sales for a sample of over 40 different musical artists from both major and independent music labels. These artists were randomly chosen from a roster of artists who released an album near the last quarter of 2007, beginning as early from June 2007 into January 2008. Weekly data for each artist were collected for 13 weeks following each album release date. Data on song sales was obtained from Nielsen SoundScan, a market research firm that tracks music sales at the point of sale for both brick and mortar and online retailers. Individual digital music tracks are sold through online retailers. In addition, weekly data on radio airplay exposure were collected from Nielsen Broadcast Data System, which uses a digital pattern recognition system to track airplay for most radio markets and video cable channels.

Data on MySpace song plays was collected weekly from each artist's dedicated MySpace page. On each of the MySpace artist pages, there is an embedded music player that features three to five songs, selected by the artist or the artists' record label or management. The music player tracks the number of times the each song is sampled. These metrics were scraped each week. Our resulting unbalanced panel data set encompasses 43 artists and 134 songs over 13 weeks. Descriptive statistics and a correlation matrix may be found in Table 1. There is a wide range for sales and song plays at MySpace and on the radio; some songs sell as few as a single unit, while there are others that sell more than a quarter-million units. Because of this high degree of variability, we apply a logarithmic transformation to our variables for estimation.

Table 1. Descriptive Statistics

	Obs.	Mean	SD	Min	Max	1	2	3
Song sales	7802	3036.51	13940.68	1	277145	1.00		
MySpace plays	3092	36267.98	102230.5	0	1825960	0.50	1.00	
Radio plays	4940	729.07	2338.13	0	23057	0.76	0.46	1.00

We also consider several artist-specific characteristics that may affect song sales or MySpace sampling: genre, using the three broad musical categories of Country, Urban, and Pop/Rock; whether or not the album is a debut album for the artist; and whether or not the album has been featured on the Billboard charts.

Model specification

We examine the effects of user sampling of songs on MySpace and radio plays on sales of song i at week t with the following specification:

$$\ln(SongSales_{it}) = \beta_0 + \beta_1 \ln(MyspacePlays_{it}) + \beta_2 \ln(RadioPlays_{it}) + \gamma_t + u_{it}$$
$$\text{where } u_{it} = \alpha_i + \varepsilon_{it} \tag{1}$$

The subscript t indexes the week following release. We include week indicators, γ_t, to capture the decaying path of sales over time. A key concern in estimating the effects of MySpace plays on song sales is the potential presence of unobserved song-specific characteristics. We explicate the structure of the error term in order to better evaluate possible sources of endogeneity. The error term u_{it} contains the following effects: (1) the unobserved, time-invariant, individual song-specific effect, α_i, and (2) a stochastic error term, ε_{it}. The time-invariant fixed effect captures any song or artist characteristic that does not change over time. For instance, such factors may include musical genre, debut artist, independent label. The correlation between MySpace activity and this unobserved characteristic, α_i, will cause traditional OLS estimates to be biased. There are several strategies for addressing this unobservable song-specific fixed effect. One method is to include indicator variables for each song fixed effect, as in the Least Squares Dummy Variable regression. Other methods exploit the time dimension within panel data to transform the data in order to eliminate the song-specific fixed effect: the first difference transformation, which measures the change from week-to-week in the variables; and the fixed effects (or within) transformation, which time-demeans the variables by song i across time t observations. Both methods effectively remove the song-specific fixed effect, which alleviates the correlation between the regressors and error term. Estimates obtained from OLS will now be unbiased and consistent.

While these methods control for some aspects of popularity, there are likely to be some unobserved characteristics that do change over time. For example, song popularity may time-varying if the song is featured on a television program in a particular week, giving a boost to both MySpace plays and song sales. Because this type of popularity is changing from week to week, it is not captured by the individual fixed effects α_i, and resides in the error term, ε_{it}. Again, the correlation between this unobservable heterogeneity and our regressors will cause OLS estimates to be biased and inconsistent. One strategy to address this endogeneity is to use an instrumental variable. An appropriate instrumental variable must fulfill two assumptions: (1) it must be uncorrelated with the error term, and (2) it must be correlated with the endogenous regressors, conditional upon the exogeneous variables. However, in a panel context, it is difficult to find an exogenous instrument which varies both across time and across each time-section. Any such variable that varies across cross-section is likely to be correlated with any potential unobserved heterogeneity, such as number of artist fans.

To address these concerns, we implement a difference- and system GMM- approach developed by Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998). These estimators are designed for "small *T*, large *N*" panel data sets with few time periods relative to the number of individuals. This method also allows for independent variables that are not strictly exogenous and controls for heteroskedasticity and autocorrelation within individuals. In addition to controlling for issues of unobserved heterogeneity and simultaneity, these estimators also allow for consistent estimation of dynamic panel models.

The difference GMM estimator controls for the endogeneity of our regressors by first-differencing to remove the time-invariant fixed effect and then uses appropriate instruments to account for any remaining endogeneity and simultaneity issues. After first-differencing equation (1), our specification becomes:

$$\Delta \ln(SongSales_{it}) = \beta_1 \Delta \ln(MyspacePlays_{it}) + \beta_2 \Delta \ln(RadioPlays_{it}) + \Delta\gamma_t + \Delta\varepsilon_{it} \tag{2}$$

Our selection of instruments relies on the assumption of sequential exogeneity. In the present context, we assume that our regressors are weakly exogenous, so that $E[x_{is}\varepsilon_{it}]=0$ for $s=1, 2,\ldots, t-1$. This assumption is equivalent to saying that our independent variables are predetermined and are not strictly exogenous; that is, the error term in time t does not affect *past realizations* of our regressors in times $t-1$ and earlier. This assumption allows for the error term ε_{it} to be contemporaneously correlated with our regressors in time t, X_{it}, so that $E[x_{it}\varepsilon_{it}]\neq 0$. Returning to our previous example, a song may receive a boost in MySpace sampling and sales after exposure on a television program; however, such an event is not likely to affect user behavior in the weeks *prior* to its occurrence. As seen in equation (2), our regressors include $\Delta x_{it}=x_{it}-x_{it-1}$. Given this assumption of weak exogeneity, we are able to use any lagged values of our regressors, $x_{i1},\ldots,x_{it-2}$ as instruments as they are uncorrelated with the error term in time t.

Lagged variables may be used as instruments for our endogenous regressors in a traditional IV approach for our static specification in equation (2) (Anderson and Hsiao 1981). However, using lagged variables as instruments results in the loss of observations available for estimation. This may be problematic in a small T scenario.

Estimation efficiency may be improved by using the difference-GMM approach. This method generates a set of moment conditions to account for these missing observations. The GMM approach differs from IV by creating an expanded instrument matrix and weights the resulting moment conditions in inverse proportion to their variance-covariance matrix.

To model the persistent nature of music sales, we next introduce lagged values of the dependent variable as a regressor. Our final model specification then becomes:

$$\begin{aligned}\Delta \ln(SongSales_{it}) = \rho_0 \Delta \ln(SongSales_{it-1}) + \beta_1 \Delta \ln(MyspacePlays_{it}) \\ +\beta_2 \Delta \ln(RadioPlays_{it}) + \Delta\gamma_t + \Delta\varepsilon_{it}\end{aligned} \tag{3}$$

The presence of a lagged dependent variable introduces an estimation bias into our model, because the lagged dependent variable and the fixed effect, α_i, within the error term are correlated (Nickell 1981). Standard panel data techniques of fixed effects or first-differencing cannot overcome this bias. However, maintaining our assumption of weak exogeneity, our explanatory variables now include $\Delta y_{it-1}=y_{it-1}-y_{it-2}$ and we can use lags of *t-3* and earlier to instrument for the lagged dependent variable, in addition to the lagged values of *t-2* and earlier for the other endogenous regressors.

When the dependent variable exhibits persistence over time, the untransformed levels may be weak instruments for the transformed variables, which may result in finite sample bias and weak accuracy. To address this the system GMM estimator proposed by Arellano and Bover (1995) and Blundell and Bond (1998) creates a system that uses the estimating equation both in first differences, as in the difference GMM, and in levels. In addition to using lagged variables in levels as instruments for the first-differenced regressors, the system GMM also employs lagged variables in differences for the regressors in levels, which effectively transforms the instruments to make them exogenous to the fixed effects. The use of the system GMM estimator requires that the dependent variable follow a convergent process, so that $|\rho_0|<0$. Finally, by including the estimating equation in levels, system GMM also allows for consideration of time-invariant regressors, which otherwise are eliminated in a fixed-effects or first-differenced transformation. In addition to MySpace and radio song plays, we consider several factors that may affect the level of song sales: genre, artist debut, and appearance on the Billboard music charts.

Results

We present the results of our difference- and system-GMM estimations of equations (2) and (3) in Table 2. As a point of comparison, we also estimate equation (1) by applying the fixed effects transformation, and present the results from this estimation in the first column of Table 2. In this basic fixed-effects model, we see that there is a distinct impact for both radio and MySpace plays on song sales; both coefficients are significant and relatively close in magnitude (0.186 and 0.163, respectively). The intercept in this model represents the average level of base sales not including the impact of MySpace and radio plays, and is relatively large and significant (exp(5.379)). In addition, the coefficient estimations on the week indicators are all significantly negative and increase in magnitude for weeks further away from the release date. This is consistent with the exponential decay pattern of sales and illustrates the steep decline following release. By capturing the pattern of sales with week dummy variables, this model essentially concludes that there is a net decrease in sales each week as given by the difference between the average level of base sales and the coefficient of the corresponding week dummy.

While the fixed effects estimation can address factors that remain constant over time, such as genre, for any other unobservable characteristics that are time varying will cause these fixed-effects estimators to be biased, which leads us to consider the uses of the difference and system GMM estimator. We first discuss the results of the static model in equation (2), which are presented in columns 2-4 in Table 2, and for the dynamic model in equation (3), which are presented in columns 5-7. Each specification is estimated using the two-step GMM estimator. The robust standard errors are corrected for any heteroskedasticity and serial correlation using Windmeijer's finite sample correction. As the consistency of the GMM estimator is sensitive the number of instruments, we control this by specifying which lags to consider, as noted at the bottom of the table. We restrict the number of instruments to be less than the number of groups (*N*) in our panel (Roodman 2006). In order for our instruments to be valid, they must be uncorrelated with the error term. The GMM estimator requires that there is no serial correlation in error terms. The presence of serial correlation could result in correlation between the error term and the lagged variables used as instrument, thus indicating the need to restrict the instrument set to yet deeper lags of the regressors. We test for the absence of serial correlation with the Arellano-Bond m2 statistic. In order to determine the absence of first-order serial correlation in the equation in levels, the Arellano-Bond statistic tests for second-order serial correlation, AR(2), in the first-differenced residuals. Under the null hypothesis, there is no autocorrelation present in the differenced residuals. Finally, we evaluate the validity of our instruments by consulting Hansen's test of overidentifying restrictions. Under the null hypothesis, our instruments are exogeneous and uncorrelated with the error term. As can be seen in Table 2, for all GMM specifications, our estimation tests fail to reject both the Arellano-Bond and Hansen tests, indicating that there is no serial correlation in the residuals and that our instruments are exogenous.

For the static model, the coefficient estimates for MySpace plays across the difference and system GMM estimators are generally similar, with estimates ranging from 0.208 to 0.251. Clearly, as compared to the fixed-effects model (coefficient estimate 0.613), this seems to overstate the effect of MySpace plays on song sales. Similarly, the impact of radio play is found to be much higher in the system GMM estimations, where it now has a positive and significant coefficient estimate, ranging from 0.434 to 0.466, as compared to the fixed effects estimate of 0.186. The effects of the time-invariant indicators for genre, debut, or Billboard status are not found to be significant in the system GMM estimation in column 4; however, the signs of the coefficient estimates suggest that songs in the Urban and Country genres do not sell as well as the base genre of Pop/Rock; an artist's debut album generally experiences lower sales volume, while albums featured on the Billboard charts enjoy greater sales.

The large increase in the coefficient estimates between the fixed effects and the GMM estimators is surprising, as we expect the fixed effects estimator to already be biased upwards in the presence of unobserved heterogeneity. If our instruments are weak, this can create an even greater bias than that of the OLS or fixed effects estimators. The case of weak instruments is also likely to occur in the panel setting when the dependent variable is persistent, in which case we must consider the dynamic model presented in equation (3).

In the dynamic model, lagged song sales are included as regressor. Comparing coefficient estimates across the difference and system GMM estimations, the coefficient estimate on lagged song sales differs greatly between the difference GMM, at 0.21, and the two system estimations, which have positive and significant coefficient estimates of 0.719 and 0.746. The coefficient estimates on the lagged dependent variable satisfy the stationary condition and indicate that current week sales can be explained by sales in the previous week. However, the difference GMM estimator will exhibit biased coefficients due to weak instruments when the dependent variable is persistent, which may explain the difference in coefficient estimates across the difference and system estimators.

Looking to our estimation results of the system GMM estimator in columns 6 and 7, the coefficient estimates for MySpace plays range from 0.12 to 0.13, which are more in line with our previous fixed effect estimate. Similarly, the effect of radio play on song sales is estimated to be about 0.16-0.18, which is also comparable to the prior fixed effect estimates. However, in the presence of unobserved time-varying heterogeneity, the fixed effects estimates will be biased. Our system GMM estimator addresses the potential endogeneity of regressors, while also allowing for the inclusion of lagged dependent variables.

The coefficient estimates on the week indicators are negative, but much smaller in magnitude than in our first specification. The constant term in both system GMM estimations are drastically less than the fixed effects estimate of 5.379, with estimates of 0.269 and 0.432. These differences between the fixed effects model and the dynamic system GMM specification may be explained by the persistent nature of sales. In the fixed effects model, we do not consider the effect of lagged sales on current week sales. The resulting estimated constant term thus represents the average level of sales, and the week indicators are increasingly negative in order to capture the steep decline in the sales pattern. By accounting for the dynamic and persistent nature of sales in the system GMM, much of the current week sales can be explained by sales in the previous week, and less of the remaining variation needs to be explained by the constant and indicator variables.

Conclusion

Overall, our results indicate that there is a positive and significant effect of user sampling of songs at MySpace on sales, above and beyond the impact of radio exposure. Moreover, our estimations results highlight the need to consider both time-varying unobserved heterogeneity and the dynamic nature of sales. In the present context, lagged values of the dependent variable are significant in explaining current week's sales. Such dynamic models in a small T, large N panel framework require the use of the GMM estimator, which alleviates the bias that emerges with the inclusion of lagged dependent variables as regressors, as well as addressing any endogeneity issues with our remaining explanatory variables. Our study is one of the first to empirically investigate the influence of user sampling of songs on sales. We identify the effects of online and traditional methods of sampling. While radio exposure continues to be an important predictor of song sales, online sampling of songs has a nearly equivalent effect on sales.

These findings point to new opportunities for musicians and record labels. Because the level of online sampling is a direct reflection of user interest in songs and artists, music companies might strategically use this information to make more informed promotional and marketing decisions. Our results also suggest that record labels may want to consider increasing the availability of music online for sampling as a way to increase user interest in their products.

Our primary objective in this paper was to address estimation concerns of unobserved heterogeneity and simultaneiety and discuss the use of the difference- and system-GMM estimators. While our findings are preliminary, this paper has several limitations that present opportunities for future research. We focus here on song sales, but future work might explore the effects of sampling on album sales. It would be particularly interesting to investigate whether online sampling of music has different effects on sales of physical CDs and digital albums. Future studies could also examine the effects of sampling for new, up-and-coming musical acts versus established, popular artists.

Another avenue of research might investigate the social network aspects of MySpace and whether this has an influence on the level of sampling or sales. For instance, do MySpace friends listen to and purchase the same songs? Do popular users influence their network friends? So far we have only considered a general level of activity at MySpace, the level of user plays of a song, and there are many exciting features to explore within social networks. In addition, we consider data from only one site that features music for sampling. While MySpace is one of the more popular music sites, there are a number of other websites, such as last.fm and imeem, which also allow users to stream music on demand. Future work might consider collecting data over time from a broader range of websites.

Researchers examining the relationship between online activity and sales should explicate the nature of endogeneity present in the proposed relationship in order to select the most suitable estimation strategy. We decompose the error term to reveal both time-constant and time-varying unobserved heterogeneity and discuss issues of simultaneity and persistence as possible sources of endogeneity. We implement a difference- and system-GMM estimator to addresses these estimation issues. A key advantage of this method is the ability to create instruments from within the data, although external instruments may also be included. While it allows for consistent estimation of a dynamic panel model, this method may be used for static models and offers gains in estimation efficiency over traditional IV using lagged values as instruments.

References

Anderson, T., and Hsiao, C. "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association* (76:375) 1981, pp 598-606.

Arellano, M., and Bond, S. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies* (58:2) 1991, pp 277-297.

Arellano, M., and Bover, O. "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics* (68:1) 1995, pp 29-52.

Bhattacharjee, S., Gopal, R., Lertwachara, K., and Marsden, J. "Whatever happened to payola? An empirical analysis of online music sharing," *Decision Support Systems* (42:1) 2006, pp 104-120.

Bhattacharjee, S., Gopal, R.D., Lertwachara, K., Marsden, J.R., and Telang, R. "The Effect of Digital Sharing Technologies on Music Markets: A Survival Analysis of Albums on Ranking Charts," *Management Science* (53:9), 2007, pp 1359-1374.

Blundell, R., and Bond, S. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics* (87:1) 1998, pp 115-143.

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

Coase, R. "Payola in Radio and Television Broadcasting," *Journal of Law and Economics* (22:2) 1979, pp 269-328.

Dertouzos, J. "Radio Airplay and the Record Industry: An Economic Analysis," *National Association of Broadcasters*, 2008

Duan, W., Gu, B., and Whinston, A. "The dynamics of online word-of-mouth and product sales: An empirical investigation of the movie industry," *Journal of Retailing* (84:2) 2008, pp 233-242.

Duchêne, A., and Waelbroeck, P. "The legal and technological battle in the music industry: Information-push versus information-pull technologies," *International Review of Law & Economics* (26:4) 2006, pp 565-580.

Gopal, R., Bhattacharjee, S., and Sanders, G. "Do Artists Benefit from Online Music Sharing?," *The Journal of Business* (79:3) 2006, pp 1503-1533.

Hendricks, K., and Sorensen, A. "Information and the Skewness of Music Sales," *Journal of Political Economy* (117:2) 2009, pp 324-369.

Liebowitz, S.J. "The Elusive Symbiosis: The Impact of Radio on the Record Industry," *Review of Economic Research on Copyright Issues* (1:1) 2004, pp 93-118.

Liebowitz, S.J. "File Sharing: Creative Destruction or Just Plain Destruction?," *The Journal of Law and Economics* (49:1) 2006, pp 1-28.

Liebowitz, S.J. "How Reliable is the Oberholzer-Gee and Strumpf Paper on File-Sharing?" *SSRN eLibrary*, 2007.

Liebowitz, S.J. "Sequel to Liebowitz's Comment on the Oberholzer-Gee and Strumpf Paper on Filesharing," *SSRN eLibrary*, 2008b.

Liebowitz, S.J. " Testing File Sharing's Impact on Music Album Sales in Cities," *Management Science* (54:4) 2008a, p 852.

Liu, Y. "Word of mouth for movies: Its dynamics and impact on box office revenue," *Journal of Marketing* (70:3) 2006, pp 74-89.

Moe, W., and Fader, P. "Modeling hedonic portfolio products: A joint segmentation analysis of music compact disc sales," *Journal of Marketing Research* (38:3) 2001, pp 376-385.

Montgomery, A., and Moe, W. "Should Music Labels Pay for Radio Airplay? Investigating the Relationship Between Album Sales and Radio Airplay," Working Paper, 2002.

Nickell, S. "Biases in dynamic models with fixed effects," *Econometrica: Journal of the Econometric Society*) 1981, pp 1417-1426.

Oberholzer-Gee, F., and Strumpf, K. "The effect of file sharing on record sales: An empirical analysis," *Journal of Political Economy* (115:1) 2007, pp 1-42.

Peitz, M., and Waelbroeck, P. "An economist's guide to digital music," *CESifo Economic Studies* (51:2-3) 2005, pp 359-428.

Peitz, M., and Waelbroeck, P. "Why the music industry may gain from free downloading: The role of sampling," *International Journal of Industrial Organization* (24:5) 2006, pp 907-913.

Rob, R., and Waldfogel, J. "Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students*," *The Journal of Law and Economics* (49:1) 2006, pp 29-62.

Roodman, D. "How to do xtabond2: An Introduction to Difference and System GMM in Stata," *Center for Global Development*, 2006.