RAND Journal of Economics
Vol. 47, No. 2, Summer 2016
pp. 394–432

An empirical study of observational learning

Peter W. Newberry*

This article provides an empirical examination of observational learning. Using data from an online market for music, I find that observational learning benefits consumers, producers of high-quality music, and the online platform. I also study the role of pricing as a friction to the learning process by comparing outcomes under demand-based pricing to counterfactual pricing schemes. I find that employing a fixed price (the industry standard) can hamper learning by reducing the incentive to experiment, resulting in less consumer surplus, but more expected revenue for the platform.

1. Introduction

■ The way in which consumers discover new products is an important issue in many of today's markets. To help with this discovery, markets often feature opportunities for observational learning, or learning about the quality of a product through the purchase decisions of peers. Popularity (i.e., sales) rankings in online markets are an especially prevalent example. The primary goal of this article is to empirically study the role of observational learning in a real-world online market. Specifically, I estimate the effect of observational learning on three market-level outcomes: the probability of success for a high-quality product, the expected consumer welfare, and the expected revenue. The uniqueness of this exercise lies in the fact that I estimate the impact of this type of learning, rather than test for its existence. Therefore, I am able to quantify the benefit of learning to firms, consumers, and producers of high-quality products.

An additional goal is to study how different pricing mechanisms affect the learning process. Price is an important determinant of consumer experimentation, meaning the way a firm prices its product affects the efficiency of learning. To analyze this, I compare the three outcomes across different pricing mechanisms. The results of this exercise provides policy makers and/or firms with information about what mechanism is "best" in a learning environment.

* The Pennsylvania State University; pwnewberry@psu.edu.
I gratefully acknowledge the help and support of my dissertation committee: J.-F. Houde, Ken Hendricks, Alan Sorensen, and Daniel Quint. I also benefited from comments and suggestions from Chris Adams, Russ Cooper, Amit Gandhi, Paul Grieco, Mark Roberts, Joel Waldfogel, and JuanJuan Zhang. I thank Chad Syverson and two anonymous referees for providing comments which substantially improved the article. I am also grateful for comments from audiences at Colby College, the University of Pennsylvania, the Wharton School, the Midwest Economics Association 2012 meetings in Chicago, IL, the IIOC 2012 meetings in Washington, DC, the Econometric Society 2013 summer meetings in Los Angeles, CA, the NBER Summer Institute 2013, Economics of Information Technology and Digitization Workshop in Cambridge, MA, and the Quantitative Marketing and Economics Conference 2013 in Chicago, IL.

I focus on an online market for music named Amie Street Music. Amie Street provides an ideal setting in which to study observational learning. The reason is that the learning process is quite transparent. Specifically, Amie Street employs a "demand-based pricing" scheme. With this, a song is free when it is first posted, and then the price increases with every download up to 98¢. This scheme not only encourages consumers to experiment with new music, it also serves as a signal of how many times a song has been purchased. Additionally, like most online music stores, users can listen to a sample of a song before they buy it (hereafter, "listen to") at no monetary cost. Unique to Amie Street, however, is the fact that a consumer observes how many of her peers have listened to the sample. Therefore, a consumer has access to two pieces of information: (i) the price, which is a signal of the number of purchases, and (ii) the number of listens.

The transparency of these signals allows me to formulate a tractable structural model of individual observational learning in this market. Specifically, I assume a consumer arrives at a song and observes the price and the number of listens. Because preferences include a common quality component (either high or low), the consumer can use this public information to form a belief about her own utility. Based on this belief, she decides to either ignore the song or listen to it. If she listens, all uncertainty disappears, and the purchase decision is straightforward. If she ignores the song, she exits the market.

Hendricks, Sorensen, and Wiseman (2012) show that in this type of environment it is possible for a high-quality product to fail. In the model described above, this would happen if the first few individuals who arrive at the song have unusual tastes for it, provide "inaccurate" information to the market, resulting in others ignoring it. The more often this occurs, the higher the probability the high-quality song fails.[1]

To quantify the effect of observational learning on this outcome, along with surplus and revenue, I first estimate the parameters of the learning model. To do this, I iterate over the individual consumer-decision process until all songs have either failed or succeeded. This iteration produces the long-run distribution of the number of listens and the price. With this joint distribution and the distribution of the outcomes observed in the data, I form a likelihood function. I then find the parameters which maximize the likelihood of the data.

To assess the impact of observational learning, I use the estimated parameters to compare the outcomes on Amie Street to a counterfactual environment in which consumers have no information. Under this assumption, consumers are unaware that prices are tied to demand and base their listening decisions only on the prior (i.e., the overall quality of music on Amie Street). The comparison of these environments allows me to quantify the value of observational learning. Results suggest that observational learning leads to an increase in total welfare by 1.35%. However, it only reduces the number of failed high-quality songs by about 1% (one song).

I also investigate how the choice of pricing scheme can impact outcomes in this market. Different pricing schemes act as frictions to learning if they discourage experimentation, limiting the amount of information that is passed to the market. First, I compare demand-based pricing to a fixed price. Although the results of this exercise serve as a general analysis of pricing in these types of markets, they are of particular interest in the music industry because of the tradition of a fixed price of 99¢ per song (e.g., iTunes).[2] I find that the counterfactual environment with a fixed price of 98¢ has a higher percentage of failed high-quality songs (3.6%), lower expected consumer surplus (25%), and higher expected revenue (142%), compared to the Amie Street environment.[3] The transfer from consumers to the platform in this environment leads to an overall welfare loss of about 3%. These results indicate that the firm can increase revenues by fixing the price, but this comes at the expense of consumers and producers of high-quality music. Next, I compare

[1] This is analogous to a "bad herd" discussed in Hendricks, Sorensen, and Wiseman (2012). However, due to assumptions made to fit the data, the interpretation of bad herds and how and why they happen differs from Hendricks, Sorensen, and Wiseman (2012).

[2] In 2009, iTunes moved to a three-tiered pricing system with prices of $0.69, $0.99, and $1.29.

[3] I use 98 instead of 99 because this is the maximum price reached on Amie Street. I believe this is a good approximation of the 99¢ price point.

the observed structure of the demand-based pricing scheme on Amie Street to two alternative demand-based schemes and find that the mechanism on Amie Street generates at least as much welfare as the others. Finally, I compare two revenue enhancing policies: (i) increasing the amount of information and (ii) fixing the price. I find that having a fixed price is the most efficient way to increase revenue, in that it hurts the consumers the least.

This article contributes to several different strands of the literature. Classic theoretical observational learning articles (e.g., Banerjee, 1992; Bhikchandani, Hirshleifer, and Welch, 1992; Smith and Sørensen, 2000) show that a decision maker, when observing the decisions of her peers, may ignore any private information she has and base her decision purely on what others have done. This can lead to herd behavior: individuals making a decision purely because others made that decision. Hendricks, Sorensen, and Wiseman (2012) show that in a model for a search good with observational learning, it is possible that beliefs converge to a point where a high-quality product gets ignored in the long run. I contribute to the above articles by empirically studying these long-run effects of observational learning.

There are many articles in the empirical literature which study how consumers learn about new products. Erdem and Keane (1996), Ackerberg (2003), and Crawford and Shum (2005) study consumer learning through experimentation, whereas Chevalier and Mayzlin (2006), Sorensen (2006), and Luca (2011) analyze the impact of peer recommendations on the decisions of consumers. In this article, I focus on observational learning as the avenue for product discovery. Empirical articles studying this type of learning are sparse. In two examples, Cai, Chen, and Fang (2009) and Zhang (2010) test for the presence of this type of learning in the restaurant industry and donated kidney market, respectively. Knight and Schiff (2010) tests for observational learning but does so using voting data from presidential primaries. I contribute to this literature by estimating the effect of observational learning, rather than testing whether it is occurring.

In a project closely related to the current study, Salganik, Dodds, and Watts (2006) show descriptive evidence of the long-run effect of social learning in an experimental music market. I quantify these effects by estimating a structural model. Additionally, I measure the effect of different pricing mechanisms in a market with observational learning. Although Bose et al. (2006, 2008) examine this topic theoretically, I am not aware of any empirical articles which do the same.

Articles which examine different aspects of the music industry include Hendricks and Sorensen (2009), which studies skewness due to information problems, Olberholzer-Gee and Strumpf (2007), which looks at the effect of piracy, and Shiller and Waldfogel (2011) and Waldfogel (2012), which examine pricing and the effect of digitization, respectively. I contribute to this literature by studying observational learning and pricing in a unique real-world market for digital music.

Finally, I estimate the learning model using only the long-run outcomes of products. To my knowledge, this is new to the empirical learning literature.

The remainder of the article is organized as follows. Section 2 discusses some of the institutional details of Amie Street Music. Section 3 describes the structural model of observational learning, whereas Section 4 provides a more detailed description of the data. In Sections 5 and 6, I present the estimation procedure and the results. Section 7 concludes.

2. Amie Street Music

■ To motivate the model, I introduce some of the institution details of Amie Street Music. On Amie Street, users can download individual songs from many different artists and genres.[4] Although a few recognizable artists appear on Amie Street, it is primarily focused on "indie" music, implying that most of the artists who appear on the site are unknown, or known to very

[4] The site was active from 2006 to September 2010, when it was bought by Amazon.com, and subsequently shut down.

few, when they first post their songs.[5] Another reason many of the artists may be unknown is the fact that *any* user can post his or her music without it being prescreened by the website.

Amie Street Music hoped to become "the place to discover new music" by encouraging consumers to learn about new music on the site through features such as demand-based pricing.[6] Although some consumers may arrive knowing exactly what they want to buy, it is assumed that most tend to use the site as an avenue for discovery and don't have prior information regarding an individual artist or song.

One of the features that makes Amie Street unique is the use of a demand-based pricing scheme. All songs start free when they are first posted, and they stay free until the song has been downloaded 13 times. After the 13th download, the price jumps to 13¢, and then increases by 1¢ for each additional download until it reaches 98¢ (at 98 downloads). Once it reaches 98¢, it remains at that price for the duration of its life. Amie Street advertises this pricing policy on their home page and many other places throughout the site, meaning consumers presumably know that the price of a song is tied to the number of purchases. This pricing scheme has two effects: it encourages people to experiment with new and/or unpopular music, and it provides a consumer with a signal of how many of her peers have purchased the song.

Similar to other online music stores, Amie Street allows consumers to listen to a one-minute and 30-second sample of a song at no monetary cost. Unique to Amie Street, however, is the fact that consumers can observe the number of times a song has been sampled, providing a "hit rate," or the percentage of times consumers purchase a song after listening to it. Along with a song's hit rate, a consumer observes other information such as the artist name, the album name, the song name, the length of the song, the download quality, and the release date of the album.

There are many ways a consumer may come across a given song. The Amie Street home page features a few artists but also has various filters by which a consumer can sort music, including from A to Z, by release date, by price, and by "buzz index."[7] In Appendix A, I show three screen shots of Amie Street: (i) the home page, (ii) song list filtered by genre (folk/country) and sorted by popularity, and (iii) an album's home page. On the second and third screen shots, the black arrow to the left of the song title is the link to listen to the sample of the song, and the information to the right of the song are the number of listens, price, and recommendations, in that order.[8] The information about when the album is released, the length of the song, and the download quality will appear when a user clicks on the album name and arrives at the album page.

The simplicity and transparency of the features on Amie Street allow me to specify a tractable structural model of consumer learning and decision making in this market. In other markets associated with learning, such as restaurants, it may not be as clear what information individuals observe and how they interpret this information.

3. Model

■ Below, I specify a model of how a consumer arrives at a song, forms beliefs about her preferences for that song, and then uses those beliefs to make her listening and purchasing decisions. The model is a modified version of the learning model found in Hendricks, Sorensen, and Wiseman (2012).

□ **Consumer preferences.** Consumer i's utility for a song is given by:

$$w_i = X + u_i - p_i,$$

[5] The term "indie" refers to artists who are not a part of the major record labels: Sony, EMI, Universal, and Warner. Amie Street did, eventually, offer music from Sony's label but did not use the dynamic pricing schedule.

[6] This slogan was shown on the front page of the Amie Street Website.

[7] Buzz index is a measure of how much action a song has received recently.

[8] In this study, I ignore the effect of recommendations. I present evidence that recommendations do not affect the learning process in Section 4.

where X is the quality of the song shared among all consumers, u_i is a idiosyncratic preference term, and p_i is the price offered to consumer i based on the demand-based pricing scheme.[9] The quality of the song is either L (low-quality song) or H (high-quality song), where $H > L$. The prior probability that a song is H is:

$$Pr(X = H) = \lambda.$$

This is the only information that consumers have about a new song when it is first posted, implying that consumers arrive with no prior knowledge of the quality for any specific song, but a correct prediction of the overall quality of music on the site.

The consumer heterogeneity term, u_i, is distributed $N(0, \gamma_u)$, and is assumed to be independent of the quality of the song. The normal distribution assumption means it is possible for consumers to have negative preferences, which may occur because of the time it takes to download a song, the space it takes up on a computer or MP3 player, or because it is simply unpleasant to listen to.[10]

Finally, p_i is the price that consumer i will pay if she chooses to buy the song, which is determined by the demand-based pricing scheme. When consumer i arrives at the song, she does not know $X + u_i$ but observes p_i.

□ **Arrival.** Before outlining the decision-making process, it is necessary to describe the arrival process. I assume that a song belongs to one of two groups: "living" songs or "dead" songs.[11] A living song receives a steady stream of consumers via a constant arrival process, whereas a dead song does not receive any consumers. We can think of a dead song as one which has "fallen off the map" by reaching the bottom of one of filters or being removed from the site altogether. All songs which have not yet fallen off the map are equally visible to a given consumer, meaning the endogeneity of arrival is captured only by the process by which songs die. This assumption can be justified by the fact that in any of the filters, a consumer would see a number of songs on her computer screen at once.

Further, consumers arrive in exogenous order to a living song and are subscripted by their arrival order. In other words, $i - 1$ consumers have made their decision about a song when i arrives. The exogeneity of arrival implies that preferences and/or learning behavior do not depend on when a consumer arrives. The primary reason for this assumption is tractability. However, the fact that the artists on Amie Street are mostly small and unknown implies that there is not likely a lot of targeted marketing occurring upon the posting of a new song or album. Therefore, there is little reason to believe that consumers arriving in the early stages of a particular song's life are fundamentally different from those who arrive later. One also might worry that there is strategic waiting for some consumers: they wait until learning has been completed. However, as long as arrival is exogenous during the learning process (i.e., before these consumers arrive), then it does not change the model. There is evidence that the learning process happens quite quickly (often within one week), meaning it is reasonable to believe that learning is complete before the waiting consumers arrive.

In the outline of the model below, I assume that a consumer knows her order of arrival, or i. Below, I point out that relaxing this assumption does not change consumer behavior, meaning that this assumption is unnecessary. However, making this assumption at this point allows for a more straightforward description of the model and makes the description of the estimation procedure in Section 5 easier to follow.

□ **Decision-making process.** When consumer i arrives at a **living** song, she observes the following information: song title, artist name, album name, album artwork, release date, current

[9] Because songs are assumed to be completely independent, I omit any song subscript.

[10] See your teenage daughter's music collection for examples.

[11] Not to be confused with songs by the incomparable Grateful Dead.

price, and the aggregate number of listens, From the current price (p), the aggregate number of listens (l), and her arrival order (i), consumer i forms her belief about the quality of the song: $\mu_i = \mu(l, p, i) = Pr(X = H|(l, p, i))$. The details of how she forms this belief can be found in the next section. Because of the heterogeneity in preferences, the information contained in (l, p, i) tells consumer i nothing about u_i. However, it is assumed that the consumer receives a private signal (S_i), which is a noisy signal of her idiosyncratic preferences:

$$S_i = u_i + \epsilon_i \qquad \epsilon \sim N(0, \gamma_\epsilon).$$

The private signal gives the consumer an idea about how much more or less she likes a song relative to the average consumer and is assumed to be generated by observing the song name, the artist name, the album name, the album artwork, or any other information that may be visible.[12] This signal provides no information about X.[13]

With μ_i and S_i in hand, consumer i can calculate her expected benefit of listening:

$$E[W|S_i, \mu_i] = Pr(w_i \geq 0|S_i, \mu_i)E[w_i|w_i \geq 0, S_i, \mu_i]. \tag{1}$$

The first element of this expression is the probability that she will purchase the song after she listens, whereas the second is the expected utility she would receive if she listens and subsequently purchases. The consumer will choose to listen if the value in equation (1) exceeds the cost of listening, assumed to be c. Because it is “free” to listen to the sample, the cost of listening is assumed to be the opportunity cost of the time it takes a consumer to learn her preferences.

Taking advantage of the assumptions of the model, the expected benefit of listening can be rewritten as:

$$\begin{aligned} E[W|S_i, \mu_i] = {} & \mu_i F_\epsilon(H + S_i - p_i)(H + S_i - p_i - E[\epsilon|H + S_i - p_i \geq \epsilon_i]) \\ & + (1 - \mu_i)F_\epsilon(L + S_i - p_i)(L + S_i - p_i - E[\epsilon|L + S_i - p_i \geq \epsilon_i]). \end{aligned} \tag{2}$$

The first line of equation (2) is i’s belief the song is high quality multiplied by i’s expected utility of listening to an H song. The second line is the equivalent, but for L songs. Because this expression is strictly increasing in S_i, I define the cutoff value of the private signal, $\bar{S}(\mu_i)$, as the S_i that makes the expected benefit of listening exactly equal to c. If a consumer receives a private signal greater than or equal to $\bar{S}(\mu_i)$, she will listen, learn w_i, and purchase if $w_i \geq 0$. If she receives a private signal lower than $\bar{S}(\mu_i)$, she ignores the song and exits the market.

This process has important assumptions besides the exogeneity of arrival discussed in the previous section. The first is that a consumer will never purchase a song without listening to it first. Although this was not required by the website, it simplifies the model, and there is evidence of this assumption in the data. Specifically, the distribution of the ratio of the number of purchases to the number of listens is concentrated toward the bottom, and there are no songs with a ratio above one (see Table 1). However, I note that I have estimated the model relaxing this assumption. In this version of the model, a consumer forms two cutoff values of the private signal: one which equates the utility of listening to the utility of exiting and one which equates utility of listening to the utility of purchasing without listening. She then decides to exit, listen, or purchase without listening based on the value of the private signal and its relationship to the cutoffs. The estimates of this version do not vary from the model presented above, implying that listening without purchasing does not happen very often. See Section 5 for further discussion.

The second assumption is that a consumer becomes perfectly informed of her preferences by listening to the sample of a song. This is justified by the fact that the sample is one minute and 30 seconds and the average length of a song is around four minutes. It is reasonable to believe that consumers can learn how much they like a given song from listening to nearly 40% of it.[14]

[12] One could also assume that the private signal is generated from information received offline. However, it cannot be the case that individuals learn anything about X outside of the Amie Street market.

[13] I have estimated an approximation of a model in which the signal is about $X + u$ with little change in the results.

[14] I note that if I relax this assumption, then the model is still identified, but then the estimated preferences are the expected utility of the song rather than the true utility.

TABLE 1 Purchase-to-Listen Ratio Distribution

rat Bin	Percent	Cumulative
[0,0.10]	3.0	3.0%
(0.10,0.20]	17.0	20.%
(0.20,0.30]	19.5	39.5%
(0.30,0.40]	12.5	52.0%
(0.40,0.50]	16.5	68.5%
(0.50,0.60]	4.8	73.3%
(0.60,0.70]	8.6	81.9%
(0.70,0.80]	7.1	89.0%
(0.80,0.90]	5.5	94.5%
(0.90,1.00]	5.5	100.0%

Notes: The table displays distribution of the purchase-to-listen ratio.

Finally, there is no competition among songs. That is, consumers are making decisions about these songs on an individual basis and not choosing among a set of differentiated products.

Before consumer i makes her listening decision, she forms her belief about the quality of the song, μ_i, using the process described below.

□ **Learning process.** For expositional reasons, let:

$$\pi((l, p, i)|X) \equiv Pr((l, p, i)|X)$$
$$\alpha(\mu_i) \equiv Pr(S_i \geq \bar{S}(\mu_i))$$

and

$$\beta(X, \mu_i) \equiv Pr(w_i \geq 0|X, S_i \geq \bar{S}(\mu_i)).$$

The first term is the probability a song of quality X has reached price p and number of listens l when consumer i arrives, whereas the second term is the probability consumer i listens, given she observes (l, p, i). The third term is the probability consumer i purchases a song of quality X, given she listens. The initial values of the problem are $\pi((0, 0, 1)|H) = \pi((0, 0, 1)|L) = 1$ and $\mu(0, 0, 1) = \lambda$.

To simplify the presentation of the model, I assume that the consumer observes the number of purchases exactly, or that p always equals the number of purchases. This will be relaxed in estimation to match the details of the demand-based pricing scheme.[15]

Upon arrival, consumer i observes (l, p, i) and forms her belief using Bayes' rule:

$$\mu(l, p, i) = \lambda \frac{\pi((l, p, i)|H)}{\lambda\pi((l, p, i)|H) + (1-\lambda)\pi((l, p, i)|L)}. \tag{3}$$

This shows that as long as a consumer can calculate the probability of observing (l, p, i) given the song is H or L, then she can form a belief about X. These probabilities are calculated by iterating over the possible decisions of consumers $i' = \{1, \ldots, i-1\}$ that could lead to this listen and purchase combination. Specially, given the initial values, the consumer iterates over the following updating equation:

$$\begin{aligned}\pi((l, p, i)|X) = {} & \pi((l, p, i-1)|X)(1 - \alpha(\mu(l, p, i-1)) \\ & + \pi((l-1, p, i-1)|X)\alpha(\mu(l-1, p, i-1)(1 - \beta(X, \mu(l-1, p, i-1))) \\ & + \pi((l-1, p-1, i-1)|X)\alpha(\mu(l-1, p-1, i-1) \\ & \times \beta(X, \mu(l-1, p-1, i-1)),\end{aligned} \tag{4}$$

[15] In Appendix B, I show how aggregation can be used to modify the model to account for the demand-based pricing scheme.

from $i' = \{1, \ldots, i-1\}$ to calculate $\pi((l, p, i)|H)$, $\pi((l, p, i)|L)$, and $\mu(l, p, i)$ for all realizations of (l, p, i). The first term of the updating equation is the probability consumer $i-1$ observed the exact same price and number of listens and decided to ignore the song. The second term is the probability $i-1$ observed one less listen and listened to the song but didn't purchase it. The third term is the probability $i-1$ observed one less listen and one less purchase, and then listened to and purchased the song. Note that to use this updating equation, the consumer must use equation (3) to calculate μ'_i for each i'. This is because correct beliefs are only formed if consumer i knows what consumer i' believed when she made her listening decision.

Intuitively, the belief is formed by first calculating the likelihood of observing a given combination of listens and purchases for each quality. That is, the consumer asks: what is the probability a song ended up at (l, p), given that it is an H or L song? To do this, the consumer must iterate over the likelihood of all the possible decisions of consumers before her that would have led to (l, p). If a song is high quality, it is more likely to have more purchases per listen along this path. If a song is low quality, then the opposite is true. The consumer then uses equation (3) to calculate her belief. Therefore, songs with a higher number of purchases, conditional on the number of listens, are going to have a higher belief.

In addition, it is important to note that, because of the randomness in tastes and signals, the more listens there are, the more informative purchases become. This is similar to much of the learning literature: the more signals a consumer receives, the closer the belief gets to the true value. For example, the belief for a song with two listens and one purchase will be very different from the belief for a song with 100 listens and 50 purchases, with the former being more weighted toward the prior.

For a concrete example of the learning process, assume consumer $i = 21$ arrives and observes $(15, 13, 21)$. She knows that five people didn't listen to the song at all, two people listened but didn't purchase, and 13 people listened and purchased. However, she does not know the path. She starts by calculating the probability consumer $i' = 1$ listened and purchased given X:

$$(\alpha(\lambda = \mu_1), \beta(H, \lambda = \mu_1), \beta(L, \lambda = \mu_1)).$$

She then moves to consumer 2 and uses equations (3) and (4) to calculate $\pi((l, p, 2)|X)$ and hence, $\mu((l, p, 2))$ and $(\alpha(\mu((l, p, 2))), \beta(H, \mu((l, p, 2))), \beta(L, \mu((l, p, 2)))$ for $(l, p) \in \{(0, 0), (1, 0), (1, 1)\}$ and $X \in \{H, L\}$. Next, she moves to consumer 3 and performs the same calculations, only with an expanding set of (l, p) combinations. She does this until $i = 21$, which will result in, among other things, $\pi((15, 13, 21)|H)$, $\pi((15, 13, 21)|L)$, and $\mu(15, 13, 21)$.

For the interested reader, I provide a full example of the learning process with calculations for consumer $i = 3$ in Appendix B.

Note that the path matters because it is important what consumer i' observed and believed when she made her decision. A model that assumes "naive" learning (what is the probability one of the consumers didn't listen) versus this "sophisticated" learning (what is the probability consumer $i' = 2$ didn't listen) will have different implications.

Similar to much of the learning literature, this inherently assumes that consumers perform a very complex updating procedure. Admittedly, this is a strong assumption. Another way to think about it, however, is that this procedure is an approximation of the consumer's decision rule: she may not explicitly make all of these calculations, but she does consider the fact that (l, p) implies certain behavior from $i' < i$ and knows that this behavior is a function of the model's parameters. Although there are certainly other decision rules which may be possible (e.g., "rule-of-thumb" learning), I provide evidence in Section 4 that the learning process described above is a good approximation of consumer behavior on Amie Street.

□ **Updating without knowledge of i.** Above, I assume that a consumer knows her place in line, i. However, the consumer behavior under this assumption is equivalent to the behavior when this assumption is relaxed. To see this, notice that the private signal, S_i, does not provide a consumer any information about the common component of utility, X. Therefore, the act of ignoring a song does not provide any information about the quality of a song. This implies that, to calculate her belief, a consumer does not need to know her place in line, she only needs to know that she has arrived after the lth consumer. For a simple example, suppose a consumer arrives and observes ($l = 15$, $p = 13$). The belief is the same whether the consumer is the 100th to arrive or the 21st. The fact that 85 people didn't listen to the song does not provide any additional information about the true quality, it is purely due to their individual tastes and private signals. What does provide information is that 13 of the 15 people who listened, purchased the song as well.

More specifically, the assumption that the private signal only gives information about one's own tastes implies that:

$$\mu(l, p, i) = \mu(l, p, i') \quad \forall i' \neq i, l < \min\{i, i'\}, p \leq l.$$

Therefore, to calculate any belief, the consumer i' can iterate on the updating procedure presented earlier up until $i = l$. From this, she will be able to calculate $\mu(l, p)$. Note that though this simplifies the updating for the consumer substantially, it does not necessarily simplify the estimation procedure. There is a short discussion of this in Section 5.

□ **Long-run implications.** In this article, I am mainly concerned with the long-run outcomes of the learning model, as opposed to the transition to these outcomes. Therefore, with the additional assumption introduced below, I define what a long-run outcome is and then describe how a song reaches this outcome. I then describe the "life" of a song and how the learning model above may affect the probability of certain outcomes.

□ **Long-run outcomes.** I assume a song has reached its long-run outcome if it has either died somewhere below the 98¢ threshold or it has survived all the way to 98¢. The former will be classified as a "failure," whereas the latter will be classified as a "success." Although these seem to be loosely defined terms, the data indicate that the 98¢ threshold is a very important landmark.[16] Therefore, the long-run outcome of a song is either the (l, p) at which it died, or $p = 98$. Note that I will use failure and death interchangeably for the remainder of the article.

The way the model is presented above, no song will ever fail because there is always a private signal which would induce someone to listen. Therefore, I introduce an additional parameter, β_d, which is an exogenous probability of death for a song at each $(l, p < 98, i)$. That is, any song that has yet to reach the 98¢ threshold is hit with this exogenous death rate at each i. One could think of this as the probability in each time period, or before each consumer arrives, that a song will "fall off the map."[17] I assume that once a song reaches 98¢, its death rate falls to some rate $\beta_{98} < \beta_d$. This assumption is not necessary to estimate the model, but it is apparent in the data that songs which make it to 98¢ are very different from songs that don't, in terms of their listening rate.[18]

Including the parameter β_d implies that this is not a standard herding model: it isn't convergence of beliefs which leads to a song's death. An alternative assumption, which is more in line with the standard models, is to define an upper bound on the private signal, $\tilde{S}$. The primary reason the death rate assumption is preferred is because it allows for a better fit of the data. Specifically, there are songs with very good signals which die (see Table A6 for examples). It is difficult for the model to predict these outcomes without some random rate of death. I have explored estimating

[16] Songs that make it to 98¢ have many more listens than songs that do not. These songs also get listened to months after they are released on Amie Street. See Figure 2.

[17] The death rate can also be thought of as a way of endogenizing the sorting mechanisms of Amie Street. If songs are mostly sorted by newest added, then it is reasonable to believe that the likelihood a song dies is related to the time it has been on the website. The death rate captures this relationship.

[18] For an easy example, if β were the same for all songs, songs would die at a similar rate at 97 purchases and 99 purchases. This doesn't seem to be the case because of the listening rate of 98¢ songs: the songs which die at higher prices have a significantly lower number of listens compared to songs which make it to 98¢.

the more standard model and though the qualitative results do not differ a great deal, the model has a much more difficult time predicting the heterogeneity in observed outcomes.

In addition, showing convergence under the standard assumption proves to be a challenge, making the death rate more appealing.[19] Finally, a model without the death rate assumes that consumers will continue to arrive at the song indefinitely, even though none of them will listen to it. It seems more reasonable to assume that songs, specifically songs with less than 98 purchases, will eventually "fall off the map."

In the estimation and results sections, I discuss a few additional reasons why including the death rate is preferred to an upper bound on the private signal.

□ **The life (and death) of a song.** In this section, I discuss why songs die and how that relates to observational learning. The probability that a song of quality H dies at $(l, p < 98)$ by the time consumer I arrives is:

$$\Pi_d(l, p < 98, I|H) = \sum_{i=0}^{I} \beta_d(1-\beta_d)^{i-1}(\pi(l, p, i|H)), \tag{5}$$

where $(1-\beta_d)^{i-1}$ is the probability a song survives all the way to i, $\pi(l, p, i|H)$ the probability it reached listens l and price p at consumer i, and β_d is the probability that the song dies. The overall probability a song dies at (l, p) by the time consumer I arrives is then the sum of this over all consumers up until I. If we take the limit as I goes to infinity, sum over all possible $(l, p < 98)$, then this will result in the overall probability that a high-quality song dies at some price below 98¢:

$$\Pi_d(H) = \sum_{(l,p<98)} \sum_{i=0}^{\infty} \beta_d(1-\beta_d)^{i-1}(\pi(l, p, i|H)). \tag{6}$$

The important feature of this equation is the fact that the "longer," or the more consumers, the song stays under 98¢, the higher the probability the song dies. For an example, suppose a song would be listened to and purchased by all consumers with probability 1. The probability that this song dies along the way is:

$$\Pi_d(H) = \sum_{(l,p<98)} \left(\sum_{i=0}^{98} \beta_d(1-\beta_d)^{i-1}(\pi(l, p, i|H)) + \sum_{i=99}^{\infty} 0 \right), \tag{7}$$

as it would only be below 98¢ for 98 consumers. Now assume that the song would be listened to and purchased by all but 10 consumers in the absence of dying. These 10 consumers would either ignore the song, or listen to it without buying it. The probability this song dies is:

$$\Pi_d(H) = \sum_{(l,p<98)} \left(\sum_{i=0}^{108} \beta_d(1-\beta_d)^{i-1}(\pi(l, p, i|H)) + \sum_{i=109}^{\infty} 0 \right). \tag{8}$$

The sum in equation (8) is greater than the sum in equation (7) purely because of the fact that there will be some mass on the price distribution under 98¢ past the $i = 98$th consumer. This example shows that within this model, the more consumers ignore a song, the more likely it is that it will fail. This is also true the more consumers listen to but do not purchase, the song. More importantly though, this shows the effect of uncertainty about X on the probability a high-quality song dies. Uncertainty will lead consumers to ignore a high-quality song because of the risk of wasting time on a low-quality song, thereby increasing the probability of death.

To emphasize the role of learning in a song's death, consider the simple numerical example presented in Appendix B. Suppose the true quality of the song is H and that $i = 1$ listened to

[19] Hendricks, Sorensen, and Wiseman (2012) show convergence in their model, but this modified version created to mirror the Amie Street market does not necessarily have the same properties.

but did not purchase the song. The probability of this happening to a high-quality song is 0.29, meaning that this consumer must have tastes that don't align with most of the other shoppers on Amie Street. The second consumer then arrives, observes $(l, p) = (1, 0)$, and forms her belief, $\mu(1, 0, 2) = 0.21$. From this, it may be the case that $i = 2$ decides not to listen (this happens with 0.16 probability). The act of ignoring the song increases the probability that the song will die before it reaches 98¢. In this case, the song received a "bad draw" of the first consumer. More generally, if a song receives a "bad draw" of early consumers (i.e., individuals with abnormal tastes), then the probability of death increases.

4. Data

■ In this section, I introduce the data used to estimate the model. I begin by describing the data and then provide some reduced-form evidence of learning in this market.

□ **Data description.** The data used in this study were collected by scraping the Amie Street website. For a subset of songs, I collected price and listen information each Tuesday morning for the four-month period from March to July 2010. In March 2010, I used Amie Street's "newest added" filter to begin scraping price and listen information for the last 1000 folk/country artists who added their music to the site. I subsequently scraped the information for any artists who were added during the sample period (about 300 in total). I therefore have a left-truncated history for the songs of 1000 artists and a "complete" history for the songs of the 300 artists. This is important because some of the songs from the 1000 artists had reached their long-run outcome before I started scraping the information, meaning I don't observe their transition. Also, because many of these songs reach their long-run outcome within a week of being posted, and the data are collected weekly, I rarely observe the transition to the long-run outcome for the 300 new artists. These facts together lead me to avoid using any of the transition data for estimation. There is more discussion of this choice in the identification and the estimation sections of the article.

Because of the restrictions of the model, I attempt to create a subsample of the data with the following characteristics:

1. Songs are homogeneous.
2. Songs are new to the consumers.
3. Songs have reached one of the two long-run outcomes.

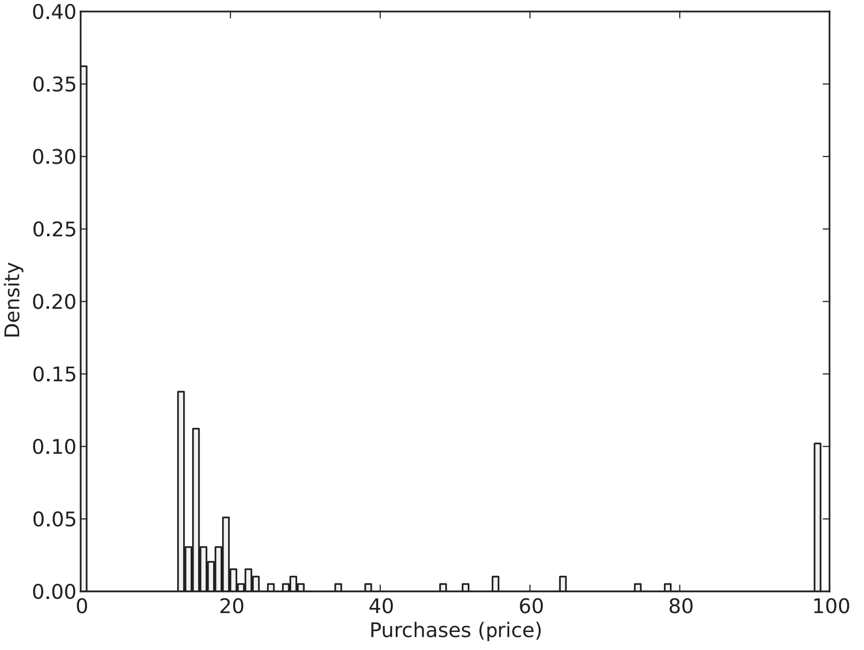
For restriction 1, I limit the data to artists who list folk/country as one of their primary genres.[20] This is an attempt to keep the heterogeneity in the songs somewhat limited. The estimation of the model relies on the assumption that the only exogenous heterogeneity which is present across songs is the value H or L. Under this assumption, the distribution of long-run outcomes of a song, given it is H or L, is a function of consumer heterogeneity and not any other unobserved component of a song. See Section 5 and Appendix A for discussions of song heterogeneity.

For restriction 2, the fact that most artists on Amie Street are small and unknown implies that consumers likely don't have any prior information about a particular song. Further, evidenced by its motto and by its marketing techniques, Amie Street was designed to be a place where consumers can come and discover new music, rather than arrive knowing exactly what they want. To further ensure that restriction 2 holds, I eliminate any artists who have an album that was released before 2006, because this implies that they may have been known to consumers before Amie Street even launched. For artists who have more than one album posted on Amie Street, I keep only the album with the earliest release date, as this is the album for which there is likely the most uncertainty about the quality of the music. Finally, each album usually contains multiple songs, so it may be important to take into account the fact that consumers could learn about an

[20] Artists can list as many as four genres.

FIGURE 1

LONG-RUN PURCHASE (PRICE) DISTRIBUTION



Notes: The figure displays the distribution of long-run prices for the 196 folk songs in the data.

artist through one song and then buy an entire album. It is evident in the data that consumers are listening to the sample of the first track to learn their preferences for that artist and/or album, so I include only the first track listed on a given album.[21] Including only the first track eliminates the need to model the learning that occurs between single songs on a given album or by a given artist. Therefore, though the data are technically at the song level, it can also be thought of as artist-level data. Note that I assume there is no learning across songs by different artists.

For restriction 3, the data are limited to songs which have reached one of two outcomes by the end of the sample period: (i) it has reached the maximum price of 98¢ or (ii) it has not reached 98¢ and has died. I define a song to be dead if the price at the end of the sample period (July 2010) is the same as the price when the site was shut down (September 2010). The sample size totals 196 songs from 196 artists. I refer to these data as the "primary data" for the remainder of the article. A more comprehensive data set I call the "full data" is discussed in Appendix A.

As discussed above, the reason why I focus on long-run outcomes is that many of these 196 have little variation in the observables across time. For example, only 50% of the songs have variation in listens and only 13% have variation in price (purchases).

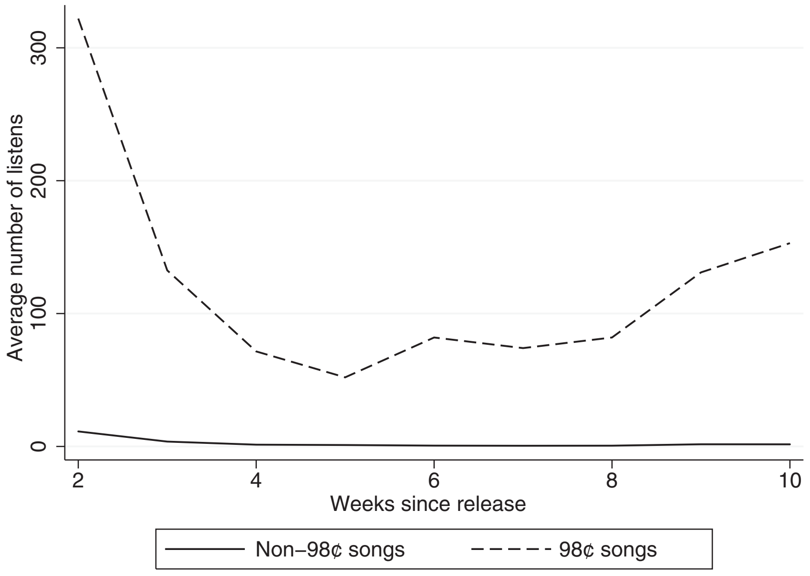
The data are separated into three categories: (i) the long-run joint distribution of listens and price for songs which died between 13¢ and 98¢, (ii) the number of songs which died at 0¢, and (iii) the number of songs which reached 98¢. Note that I do not use information about the number of listens for 0¢ or 98¢ songs. The reasons for this are discussed in the estimation section.

Figure 1 shows the price (purchases) distribution for the sample. The first thing to notice is that only about 10% of the songs make it to the 98¢ threshold, even though it only takes 98 downloads to get there. Also, many of the songs fail at very low prices (below 30¢), whereas

[21] The first track listed on an album has on average 198 listens, the second has 133, the third has 126, the fourth has 105, and so on.

FIGURE 2

LISTENS FOR 98¢ AND NON-98¢ SONGS



Notes: The figure displays the average number of listens for songs depending on how long they have been on Amie Street and whether they are a 98¢ song.

TABLE 2 Listen Distribution for Unsuccessful Songs

Buys (Price)	Mean Listens	Standard Deviation Listens
[13,97]	46	53
[13,30]	39	34
[30,97]	178	109

Notes: The table displays the distribution of listens at different price levels.

there is a scattering of songs which reach a higher price before failure. In fact, over 30% of the songs don't even make it out of the free range.

The fact that there are very few songs which die between 30¢ and 98¢ provides some preliminary evidence that learning is occurring: songs which make it above 30¢ continue to get listened to and bought.[22] Finally, note that there are some large spikes in the data at 13¢ and 15¢. The 13¢ spike is likely due to the fact that songs are free until 13 people have downloaded them, but the 15¢ spike is difficult to explain. Overall, the price distribution shows the binary nature of the data: songs likely either die at low prices or make it to 98¢.

To further investigate the data, Table 2 shows the average and standard deviation for the number of listens for different price levels. The first row displays the distribution of listens for all the songs which failed below 98¢, whereas the second and third row split the data into songs that failed above and below 30¢. For all failed songs, it takes on average about 46 listens before a song dies. This number is 39 for songs which die under 30¢ and 178 for songs which die above 30¢. Intuitively, this increase in the conditional mean makes sense, as it is assumed that people listen to songs before they purchase them, but the jump is quite large. The small purchase-to-listen ratio

[22] This leads to the question of whether artists will try to game the system by buying their own music. On the site, a song can only be bought once by each username, alleviating this concern to some extent. The fact that consumers are able to listen before they buy may alleviate these concerns as well: a low-quality song may not benefit from buying their own song because this will only lead to more listens and not necessarily more buys.

for these songs implies that there is a lot of consumer heterogeneity in preferences and private signals: for a song to make it that high with such a low purchase-to-listen ratio, it must be the case that tastes for music varies greatly. Table 2 also displays the standard deviation of listens for dead songs. All three rows show that the spread in the number of listens it takes a song to die is quite large, implying that songs are dying with either many people listening without buying or just nobody listening.

□ **Learning on Amie Street.** To provide evidence that the learning model is a good approximation of consumer behavior, I test for properties that are intuitively implied by the model. In doing so, I also test the assumption that recommendations do not affect the consumer's learning process. The properties I test for are 1: the probability of listening is increasing in the purchase-to-listen ratio, and 2: the effect of the purchase-to-listen ratio increases as the number of listens increases. It is out of the scope of this project to prove these properties analytically, but I show evidence that they hold via simulation in Appendix A. The first of these properties comes from the fact that, all else equal, a higher purchase-to-listen ratio implies a higher belief, which leads to a higher rate of listening. The second property comes from the fact that more information is better: the purchase-to-listen ratio is more informative if more people have listened.

To test for these, I utilize the dynamic portion of the data. This allows me to examine behavior at the song level at different values of price, listens, and recommendations. As pointed out in the earlier section, there is not a lot of variation in some key variables across time for any specific song. Specifically, for these tests, 50% of the data have variation in listens, 13% have variation in price (purchases), and less than 5% have variation in recommendations. Therefore, in discussing the results, I focus on a regression which utilizes variation across songs (i.e., without a song-level fixed effect) and present a regression which utilizes variation within songs (i.e., with a fixed effect) as a robustness check. I note that if the model is correct, then songs are identical *ex ante*, and the regression without song fixed effects is enough to identify the properties above.

The regression model is given by:

$$\Delta listens_{jt} = \alpha_0 + \alpha_1 recs_{jt} + \alpha_2 rat_{jt} + \alpha_3 price_{jt} + \alpha_4 order_{jt} + D_t + \epsilon_{jt}.$$

The dependent variable, $\Delta listens_{jt}$, is the number of listens for song j during week t. The independent variables are measured at the beginning of week t. Specifically, $recs_{jt}$ is the number of recommendations at the beginning of week t for song j, $price_{jt}$ is the price, and $order_{jt}$ is a measure of how long since the song has been posted on Amie Street. The primary learning signal, rat_{jt}, is the ratio of the number of purchases to the number of listens. Finally, D_t is a week dummy variable.

Notice that this regression models the change in the number of listens in a given week as a function of the independent variables at the beginning of that week. Therefore, any dynamics that occur during the week are not accounted for, leaving the magnitude of the estimated coefficients difficult to interpret. However, this model can still be used to test for learning using the estimated sign of the coefficient.

The error term ϵ is assumed to be i.i.d. across time and songs. This accounts for any unobserved demand shocks for a song during a given week, which could include concerts, radio play, etc. The independence assumption implies that these shocks only affect the number of listens in that week and will not carry over to future weeks, nor be correlated to any activity in previous weeks. Endogeneity of the price, recommendations, and/or purchase-to-listen ratio may be a concern, as bands which have concerts and play on the radio may be more likely to be bands which receive a lot of attention on Amie Street. Again, the fact that most of these artists are new and unknown should alleviate these concerns. However, because I am not able to account for possible correlation between the error term and these observables, it is reasonable to interpret the results as correlational evidence consistent with the learning model.

Before presenting the results, it is important to mention that these tests rely on the assumption that consumer arrival is either constant across weeks or that it is accounted for through the *order*

TABLE 3 Learning on Amie Street (1)

Variable Dep. Var.	(1) $\Delta list$	(2) $\Delta list$	(3) $\Delta list$	(4) $\Delta list$
rat	8.2*** (2.2)	132.4*** (8.5)	7.1*** (2.1)	189.5*** (9.8)
recs	0.3 (0.3)	19.6*** (6.2)	0.4 (0.3)	–
price	−0.3 (0.1)	−7.9*** (0.7)	−0.0 (0.0)	−2.0 (1.2)
order	−0.0 (0.0)	0.0 (0.0)	−0.0 (0.0)	0.0 (0.0)
Fixed Effects	N	Y	N	Y
R^2	0.17	0.48	0.02	0.26
Obs.	1469	1469	1411	1411
Songs	110	110	105	105

*** $p < 0.01$, ** $p < 0.05$.

Notes: The table displays the results of a regression of weekly listens on the value of the independent variable at the beginning of the week. A week dummy is included in all specifications. *rat* is the ratio of purchases to listens and *order* is a measure of how long the song has been on Amie Street. Data are limited to prices between 13¢ and 97¢.

TABLE 4 Learning on Amie Street (2)

listen Range	$\hat{\alpha}_2$	Obs.
$listen \leq 50$	10.4*** (2.9)	860
$50 < listen \leq 75$	14.6** (6.4)	269
$75 < listen \leq 100$	30.6 (16.9)	94
$100 < listen$	23.7 (17.5)	246

*** $p < 0.01$, ** $p < 0.05$.

Notes: The table displays the results of a regression of weekly listens on the value of the independent variable at the beginning of the week. The same regressors as in Table 2 are included in each regression. Only the value of the coefficient on *rat* is displayed. Data are limited to prices between 13¢ and 97¢ cents. Displayed are the coefficients on interaction variables between *rat* and dummy variables indicating the number of listens.

and/or D_t. Under this assumption, I can attribute differences in the number of listens across weeks to changes in *rat*, *recs*, or *price*, rather than changes in the arrival process. This is violated if songs which have a high price, for example, are more visible on Amie Street. The fact that there are many different sorting mechanisms (price, recommendations, age, etc.) alleviates this concern to an extent. Also, because this is merely a check if learning is occurring, I proceed assuming that the arrival process is accounted for through the observables.

I do not observe the true number of purchases for songs at 0¢ and 98¢, so I restrict the data to songs which are between 13¢ and 97¢. Results from this analysis can be found in Tables 3 and 4. In Table 3, I show the results from two different regressions to perform test 1: one with song-level fixed effects and one without. These are specifications 1 and 2, respectively.

Results from specifications 1 and 2 show a positive and significant coefficient on the purchase-to-listen ratio.[23] The price coefficient is negative as predicted but is only significant when including a song fixed effect. The coefficient on the measure of time on the site, *order*, is not statistically significant in either regression. These results match up fairly well with the property 1. That is, consumers are taking price (purchases) and the number of listens as a signal of song quality: listening increases with a better ratio.

In addition, I find the effect of recommendations to be insignificant in specification 1 and positive in specification 2. Very few of these songs actually receive recommendations: only 25% have any recommendations and only 16% have more than one. Further, only 5% of songs have any variation in recommendations across time. To check if these 5% have a significant affect on

[23] The differences between specifications 1 and 2 are possible evidence that song heterogeneity is important in this market. I show evidence that this may not be the case in Appendix A. Additionally, the fixed effects regressions rely on very little variation at the song level, so the results must be taken with that in mind.

the other coefficients, I run the same regressions but dropping these songs. As can be seen in specifications 3 and 4 in Table 3, the results do not change significantly when dropping these songs. Although ignoring recommendations makes the model more tractable, there may be a worry that this is an important determinant in the listening decision. This is a valid concern. However, I take the limited amount of variation in recommendations along with results of specification 1 as justification for ignoring recommendations in the model.

Finally, to perform test 2, I run the same regression as specification 1 above, but I include interaction terms between the purchase-to-listen ratio and dummy variables for different ranges of the number of listens. The coefficients on these interactions indicate the effect of the learning signal in these ranges. The model predicts that the effect of the purchase-to-listen ratio should increase as the number of listens increases. This is because with more information, it is less likely that the signal came from pure randomness. The results in Table 4 are in line with this behavior for the most part: the effect of the signal grows as the number of listens increases. Although the coefficients in the high ranges are not significant at the 5% level, they are larger than the coefficients in the low ranges. Further, most of the observations fall in the lower ranges, meaning that the prediction holds for much of the sample.

Overall, these results suggest that learning is occurring and that the specific model of learning which I propose is, at the very least, a good approximation of behavior in this market.

□ **Other learning models.** In specifying the model, I assume that consumers are undertaking a very complicated procedure to formulate their belief. As discussed earlier, it may be more appropriate to think of the Bayesian learning model as an approximation of consumer behavior. However, it may be the case that even this approximation is misspecified. For example, consumers may use a "rule-of-thumb" to make their decision.[24] It is important to decipher between the learning model proposed and these other common forms of learning. Although I cannot test for every other possible learning rule that could be in the data, I believe that the tests in Section 4 provide some evidence that consumers are not using a straightforward "rule-of-thumb" based on the observable information.

I believe the most plausible rule-of-thumb is to listen to a song if the price is higher than a given threshold and ignore the number of listens. Another possibility is that consumers have a threshold of listens and ignore the price. The results presented in Section 4 suggest that it is the *combination* of price and listens which affect behavior, rather than either of these variables by itself. Another possible rule-of-thumb is to listen as long as the purchase-to-listen ratio is above a threshold, unconditional on the amount of information available. The results in Table 4 indicate that consumers are more responsive to the learning signal when there is more information available, again suggesting that consumers are learning according to an updating procedure.

5. Estimation

■ The vector of parameters to be estimated is:

$$\theta = [H, L, \gamma_u, \gamma_\epsilon, c, \lambda, \beta_d],$$

where H and L are the average utility for a high-quality song and a low-quality song, respectively, γ_u is the standard deviation of preferences, γ_ϵ is the standard deviation of the private signal, c is the cost of listening, λ is the prior, and β_d is the death rate. Below, I describe the procedure to estimate θ.

□ **Procedure[25].** Given a vector of parameters, $\tilde{\theta}$, I calculate the probability that a song has died at (l, p, i), $\Pi_d(l, p, i|\tilde{\theta})$, the probability a song dies at 0¢, $\Pi_0(i|\tilde{\theta})$, and the probability

[24] See Ellison and Fudenberg (1993) for a theoretical analysis of this type of learning process.

[25] See Appendix B for additional details about the estimation procedure.

a song has made it to to 98¢, $\Pi_{98}(i|\tilde{\theta})$, using the same updating procedure introduced in the model section. That is, I iterate over the probability of decisions by consumers $i' < i$ to produce $\pi(l, p, i|H, \tilde{\theta})$ and $\pi(l, p, i|L, \tilde{\theta})$ for any (l, p, i). Using equation (5), I form:

$$\Pi_d(l, p, i|\tilde{\theta}) = \lambda \Pi_d(l, p, i|H, \tilde{\theta}) + (1 - \lambda)\Pi_d(l, p, i|L, \tilde{\theta})$$

for all l and $12 < p < 98$. Also, I form:

$$\Pi_0(i|\tilde{\theta}) = \sum_l (\lambda \pi(l, p = 0, i|H, \tilde{\theta}) + (1 - \lambda)\pi(l, p = 0, i|L, \tilde{\theta})),$$

and

$$\Pi_{98}(i|\tilde{\theta}) = \sum_l (\lambda \pi(l, p = 98, i|H, \tilde{\theta}) + (1 - \lambda)\pi(l, p = 98, i|L, \tilde{\theta})).$$

Because the songs in the data are assumed to have reached one of the two specified long-run outcomes, I need to iterate to the i such that a song has either surely died below 98¢ or made it to 98¢. This is what I define as convergence of the model. Convergence will occur at the i such that:

$$\sum_{(l, 12<p<98)} \Pi_d(l, p, i|\tilde{\theta}) + \Pi_{98}(i|\tilde{\theta}) + \Pi_0(i|\tilde{\theta}) = 1.$$

An equivalent representation of convergence is the i such that there are no longer any songs below 98¢ that are living, or:

$$\Pi_l(l, p, i|\tilde{\theta}) = (1 - \beta_d)^{i-1}(\lambda \pi(l, p, i|H, \tilde{\theta}) + (1 - \lambda)\pi(l, p, i|L, \tilde{\theta})) = 0 \quad \forall\ (l, p < 98).$$

With this representation, it is straightforward to show that the model surely converges as i approaches ∞. In practice, however, I define convergence of the model as the i such that:

$$\max_{(l,p,i)} \Pi_l(l, p, i|\tilde{\theta}) < \varepsilon,$$

which I call I.[26] Once the model converges, I form the log-likelihood function:

$$\ell(\tilde{\theta}) = \sum_{(l,p)\in(l^*, 12<p^*<98, I)} \log(\Pi_d(l, p, I|\tilde{\theta})) + \sum_{k=1}^{K_{98}} \log(\Pi_{98}(I|\tilde{\theta})) + \sum_{k=1}^{K_0} \log(\Pi_0(I|\tilde{\theta}))$$

where $(l^*, 12 < p^* < 98)$, K_{98}, and K_0 are data. K_{98} and K_0 are the number of songs that reached 98¢ and 0¢. Recall that the data are the joint distribution of (l, p) for songs which died between 13¢ and 98¢ and the number of songs that made it to 0¢ or 98¢. I do not use any information about the listens for 98¢ songs because the data indicate that these songs are still alive, and for the model to predict the number of listens for living songs, I would need to know the unobserved i. This is an advantage of using only data which have converged: knowledge of i by the econometrician is unnecessary. Further, listen information for 98¢ songs would likely be most helpful in identifying β_{98}, which I am not estimating. I also do not include the listen information for the 0¢ songs because I cannot exactly match the number of listens to the number of purchases for these observations.

To find the global maximum of the objective function, I use the ARS (accelerated random search) algorithm in Appel, Labarre, and Radulovic (2003) to find various points where the objective function is high. I then use these as the starting points in a Newton method.

Note that the econometrician is not able to take advantage of the fact that the updating procedure for consumers does not require knowledge of i, as discussed in Section 3. This is

[26] For the results presented, I set ε to 0.01. I have also experimented with other values between 0.001 and 0.01, with little change in the results. Each calculation of the objective function takes somewhere between 20 seconds and five minutes, depending on the values of the parameters and the convergence criteria.

because the number of times a song has been ignored is an important determinant of the long-run outcome and hence, affects the calculation of the likelihood. A consumer, on the other hand, can form her belief without knowing how many of her peers ignored the song.

□ **Identification.** Intuitively, identification of the parameters comes from the fact that each parameter changes the consumer's decision process in a different way. That is, as the price increases according to the mechanism of the website, a change in a parameter will change the listening and purchasing behavior of consumers along the path to the song's final outcome. Further, each parameter has a different effect on the reaction to these changes in price, which implies that the final outcome (i.e., price and listens) varies with the parameter values. For instance, a higher value of H increases the probability the average consumer will buy a high-quality song at a given price, whereas a higher value of c will lower the probability people listen. Below, I provide some heuristic arguments for each parameter, but first I discuss the key assumptions which facilitate identification of the model.

The first key assumption is that the quality of songs are identical *ex ante*, meaning the variation in outcomes of songs, conditional on true quality, comes from the taste shocks (u) and the private signals (S). If this assumption is violated, then the model would incorrectly attribute the differences in the outcomes for these songs to the learning process rather than the differences in expected quality.

To solidify this intuition, it may be helpful to think of this in the reduced form. Suppose we run the regression found in Section 4 without the song fixed effect. The concern is that there is something about the song which is unobserved to the econometrician but is correlated with the learning signal. The most obvious example of this would be if there is unobserved quality of the song which affects the behavior of the previously arriving consumers and thus, the learning signal, and also affects people's listening and purchasing behavior in this time period. This is a common difficulty in identifying the effect of social learning signals. The most straightforward solution to this problem is to add a song-level fixed effect, which is what I do in specification 2. I am essentially doing the same thing in the structural model by assuming that all songs are the same *ex ante*.

However, there are a few possible approaches to solve this problem without making this assumption. First, I could use the dynamic portion of the data and include a song-level effect. As discussed above, the dynamic information in the data is very limited, which precludes me from being able to identify these individual song effects. The second is to add some more song-level or artists-level observables in the specification of song quality, such as the length of the song and/or the subgenre of the artist. This approach requires the econometrician to solve the model for each of the possible values of the observables. The reason I do not take this approach is that adding this feature would increase the computational time substantially. Finally, I could take an instrumental variables approach by finding an excluded variable which affects the learning process but is not correlated with unobserved quality of a song. One such example would be if Amie Street changed the way it displayed the learning information and/or changed the amount of observable information. In fact, a change like this actually occurred in September 2010: Amie Street removed the number of listens from the website. This structural change could act as an instrument which changes the learning process but does not change the true quality of a song. Because my data are scraped directly from the website, this change also meant that I no longer had access to the listen information and precludes me from using this shock as an instrument. Therefore, I move forward under the assumption that all songs in the folk genre are identical *ex ante*. See Appendix A for some validation of this assumption.

An additional assumption which deserves discussion is the fact that, according to the model, consumers must listen to a song before they purchase it. Although I do not believe it is essential for the identification of the model, it does provide a clean way to separately estimate the preference parameters from the learning parameters. Specifically, assuming that anyone who purchases knows their tastes for a song implies that I observe a decision affected by the learning parameters (listening) and a decision not affected by the learning parameters (purchasing). However, I have estimated the model relaxing this assumption and the results do not change. The reason is because it appears that purchasing without listening isn't common: if consumers were purchasing without listening with a positive probability, then there would be a positive probability that a song ends up with a purchase-to-listen ratio greater than one. Because this does not happen in the data (see Table 1), the estimates converge to points where consumers are not behaving this way, at least not very often. To examine this closer, I calculate the probability that multiple consumers choose to purchase without listening under the estimated parameters of the model, and find that it is very small.[27] This implies that it is unlikely that this behavior affects the results a great deal.

I now move to a more detailed heuristic discussion of identification of each parameter. If true song quality were observable, then identification is straightforward. Specifically, the preference parameters (H, L, and γ_u) are identified by the location and shape of the price distribution for H and L songs. The learning parameters (γ_ϵ and c) are identified by the location and shape of the listen distribution for H and L songs. The prior, λ, is identified by comparing the number of H songs to the number of L songs. Finally, β_d is identified by the H songs which have very good signals, but still die.

Because I do not have any exogenous measure of song quality, I cannot systematically separate the high-quality from the low-quality songs. However, I argue that the parameters can be identified as long as there are two distinguishable distributions in the data. I refer to the price distribution found in Figure 1. There is evidence that songs fall into four categories: songs which die at a price under 25¢, songs which die between 25¢ and 40¢, songs which die above 40¢, and songs which make it to 98¢.[28] Identification relies on the assumption that the distribution of songs which die under 25¢ are mostly low-quality songs and the distribution of songs which either die above 40¢ or make it to 98¢ are mostly high-quality songs. Songs which fall between 25¢ and 40¢ could be either high quality or low quality.

Therefore, per the above discussion, the L parameter is identified by the location of the price distribution at the low end, and more specifically, the number of songs which die at 0¢. The H parameter is identified by the number of songs which make it to 98¢ and the location of the price distribution of songs at the high end. The heterogeneity in preferences is identified by the spread of the price distribution for songs at the low end. The listening cost is identified by the distribution of the conditional mean of listens, whereas the standard deviation of the private signal is identified by the conditional variance of listens. The prior is identified by the ratio of the number of songs at the high end of the price distribution to the low end and the death rate is identified by the number of songs which die with a very good signal (i.e., high ratio of purchases-to-listens).

To provide strength to these heuristic arguments, I perform the sensitivity analysis proposed in Gentzkow and Shapiro (2014) in Appendix B. For the most part, the results of this analysis line up with the discussion above.

There are a few further identification notes that deserve discussion. First, the assumption of a two-state world is crucial. However, one may think that there is more heterogeneity at the song level. To test for this, I perform various tests of the equivalence of the price distribution across different observables. These tests mostly indicate that the price at which a song dies is not dependent on the number of albums the artist has or the genre it belongs to. I take this as evidence that the heterogeneity in the price distribution is due to consumer preferences rather than differences in songs and/or artists. Results of these tests can be found in Appendix A.

Second, though I assume that preferences have infinite support, the fact that prices fall between 0 and 98 means that I cannot exactly estimate the entire distribution. Therefore, the estimate of preferences at unobserved prices comes from the mass of people who have utility

---

[27] An upper bound for the probability that two consumers choose to purchase without listening can be calculated by assuming that the belief is equal to 1 and determining the probability that consumers purchase without listening at $p = 13$ and $p = 14$. This is equal to 0.0451. For three consumers, this becomes 0.0096, and for four consumers, 0.0020.

[28] The distribution of the full data is similar. See discussion in Appendix A.

TABLE 5 Estimates

H	L	γ_u	γ_ϵ	c	λ	β
\$1.37	−\$3.43	\$4.11	\$8.56	\$0.37	0.40	0.010
(0.24)	(0.09)	(0.08)	(0.99)	(0.01)	(0.06)	(0.001)

Notes: Standard errors calculated numerically in parentheses.

lower than 0 or higher than 98. For example, there are many songs which people do not buy even when they are free (see the distribution at price of 0). The assumption of normally distributed preferences implies that L may be negative, and possibly very negative. However, all this is really indicating is that there is a large group of people who do not value the song above 0.

Finally, though most empirical learning articles use individual consumer-level data, I argue above that I am able to identify the model using only aggregate data. Although the transition data may be helpful for more precise identification, in many cases these types of data may not be available. Additionally, using only the converged data allows me to avoid any assumptions about the unobserved arrival process of consumers: all I need to know is that there have been enough consumer arrivals for the songs to have converged.

6. Results

■ In this section, I present the estimates of the parameters, along with the results of the counterfactual analysis.

□ **Estimates.** The estimates are found in Table 5 with standard errors in parentheses. The value of a high-quality song is \$1.37, and the value of a low-quality song is −\$3.43. The heterogeneity in preferences is large, \$4.11, which is reasonable in a music market. The negative value for L is interesting, but because I do not observe decisions being made at negative prices, this, along with the estimated heterogeneity in preferences, is more informative of the percentage of consumers who would not download the song for free: about 80%. This result comes from the large number of songs which fail at 0¢ in the data.

The high value of γ_u also implies that there is a large overlap in preferences for high- and low-quality songs. This means that observing the purchases of others may not be all that informative of one's own tastes. We see this in the data, as there are some songs which make it to a high price with a low purchase-to-listen ratio: the average number of listens for songs which die above 30 is 178 (see Table 2).

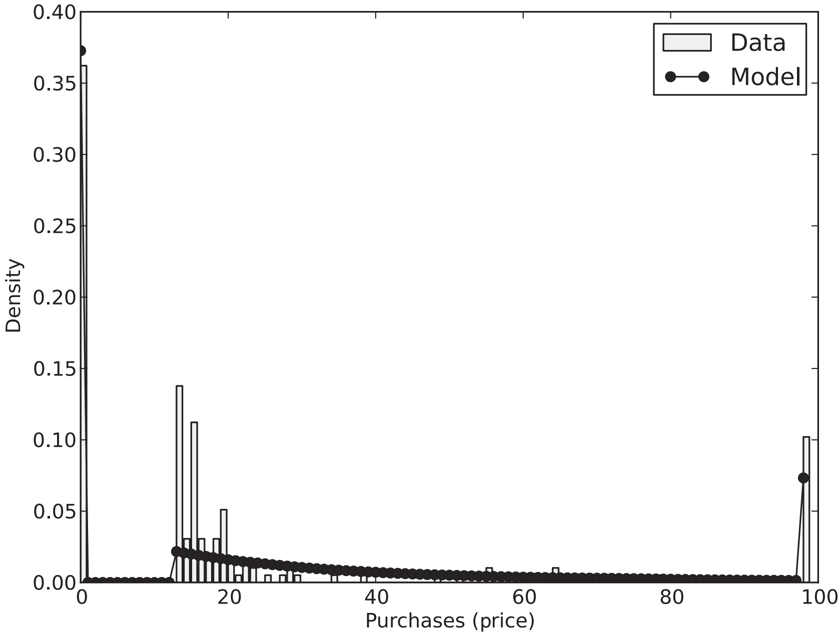
The estimates of λ and β_d are 0.40 and 0.01, respectively, indicating that 40% of the music on Amie Street is high quality and each song has a 1% chance of failing at each consumer arrival. The cost of listening to the one-and-a-half minute sample is about \$0.37, which translates to \$15 an hour. Finally, γ_ϵ's estimate is large in magnitude, implying that the private signal is not very informative. We can see this in the data with the large amount of heterogeneity in the conditional listen distributions shown in Table 2.

I compare how well the model fits the data in Figure 3. This shows the estimated price distribution (dark circles) along with the true distribution (light bars). The model fits the overall shape of the price data fairly well. One place where the model misses is predicting the number of songs which die at 13¢ and 15¢. The model predicts a smooth distribution, making it impossible to predict these jumps. Considering I am assuming a model of only two qualities, I find the fit of the price distribution impressive.

The model also does well in predicting the number of listens conditional on the price. Table 6 shows that the average and the standard deviation of listens under 30¢ match the data closely. The mean and the standard deviation are underestimated for songs which die above 30¢, but this is likely due to the sparse data in these regions.

FIGURE 3

PURCHASE (PRICE) DISTRIBUTION FIT



Notes: The figure displays the long-run price distribution in the data (bars) and the estimated long-run price distribution (dots).

TABLE 6 Model Fit. Listen Distribution for Unsuccessful Songs

Buys (Price)	Mean Listens	Standard Deviation Listens
[13,97]	85 (46)	81 (53)
[13,30]	44 (39)	36 (37)
[30,97]	137 (178)	78 (110)

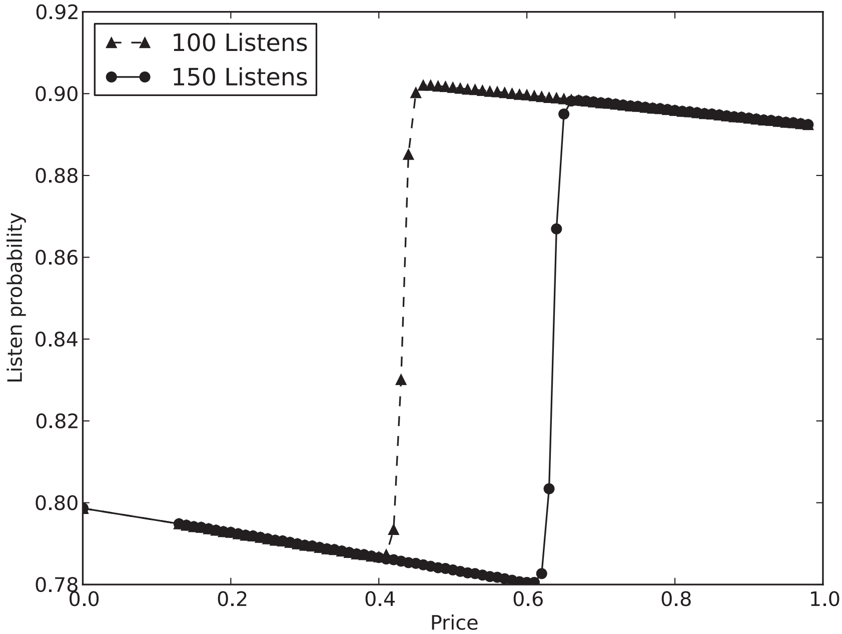
Notes: The table displays the conditional distribution of listens at different price levels. Estimates displayed, data in parentheses.

Although I addressed the question of the existence of learning in a reduced-form way in Section 4, I now discuss it in the context of the results of the model. Figure 4 displays the probability the 200th consumer listens to a high-quality song given she observes 100 listens and the price on the X-axis. The plot for 150 listens is also displayed. The figure indicates that the learning information has no impact for prices between 0¢ and 40¢, as the probability of listening is strictly decreasing in that range. In other words, the increase in price is affecting the listen probability through downward sloping demand and not through learning. Once the song reaches 40¢, however, learning happens very quickly and the listen probability jumps to above 0.90. The same pattern can be seen for 150 listens, though at a higher price. The point of displaying this graph is to show that learning impacts listening behavior only after songs have reached a certain price.

I compare this feature of the model to the data in Figure 5. This displays the estimates of the effect of price on the propensity to listen using a regression similar to the one found in Section 4 but including dummy variables for different price levels. The figure indicates that the price has little to no effect on listening behavior at lower levels (under 50¢). Further, this effect is negative for some prices under 50¢, matching what is seen in Figure 4. After a song reaches 50¢, however, the price begins to have a positive effect on the listening rate.

FIGURE 4

ESTIMATED LISTEN PROBABILITY



Notes: The figure displays the estimated listen probability for a high-quality song at the 200th consumer over different price levels. The triangles are when there are 100 listens and the dots are when there are 150 listens.

One could imagine an exercise where we aggregate plots like Figure 4 across the different number of listens and consumers. In doing this, I believe we would likely see something very similar to Figure 5. I take this as evidence that I am estimating the learning process without using any information about individual learning.

□ **The effects of observational learning.** Using the parameter estimates, I calculate three outcomes of interest: (i) the probability a high-quality song fails, (ii) the expected consumer surplus, and (iii) the expected revenue. The first outcome is a measure of the efficiency of learning, but also serves as a measure of the benefit to producers of high-quality music.

The second outcome measures the benefit to consumers and is calculated by summing the expected surplus for each consumer arrival until a song converges. I then multiply this by the total number of songs in the market to get the overall *ex ante* consumer surplus for the 196 songs. The third measures the short-run benefit to the online platform and is calculated by summing the expected revenue for each consumer arrival. Like the surplus calculation, I multiply this value times the total number of songs. It is a short-run benefit because it does not take into consideration that consumers who have a good experience on Amie Street may return to purchase more music later. Therefore, the amount of consumer surplus can be considered an additional benefit for the online platform.

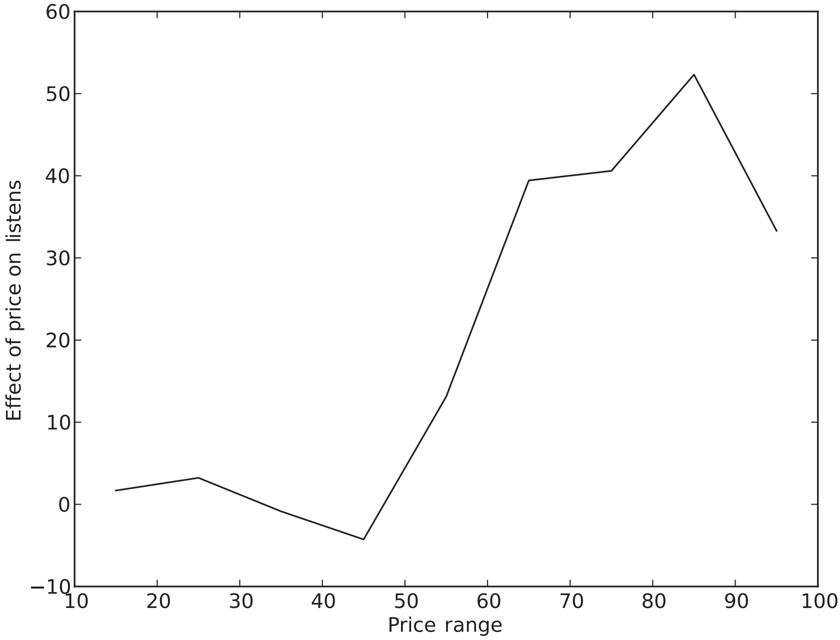
Table 7 displays the results for the estimated model.[29] Results indicate that 82% of high-quality songs fail on Amie Street.[30] This is about 63 total songs. Applying this same percentage

[29] The expected consumer surplus and revenue are calculated assuming $\beta_d = \beta_{98}$. This implies that the estimated impacts of observational learning on consumer surplus and revenue are lower bounds.

[30] In Appendix A, I investigate the existence of high-quality songs failing by using Facebook "likes" of artists in the data as a measure of song quality. The results of this exercise are mixed but indicate that there are some successful artists who have songs which fail. Table A6 provides some artists who may fall into that category.

FIGURE 5

THE EFFECT OF PRICE ON THE PROPENSITY TO LISTEN



Notes: The figure displays the estimated coefficients of a regression of the change in listens on different price levels (along with other covariates mentioned in the second subsection of Section 4). This is intended to display the effect of different prices on the propensity to listen. The dynamics in the data are used for this exercise.

TABLE 7 Learning Outcomes

Row	Informational Environment	Probability of Failure for *H* Songs	Total Consumer Surplus	Total Revenue	Total Welfare	Total Cost or Benefit
(1)	AS	81.8%	\$15,145	\$2,289	\$17,435	–
(2)	NI	0.91%	−1.20%	−2.35%	−1.35%	−\$236
(3)	PI	−0.15%	0.29%	0.46%	0.31%	\$54

Notes: Table displays the outcomes for AS in row (1), the percentage difference between AS and no information (NI) in row (2), and the difference between AS and perfect information (PI) in row (3).

to the full data implies that about 689 high-quality songs fail. The expected consumer surplus is about \$77 per song, or \$15,198 for the entire market[31] and about 17¢ per consumer.[32] Finally, the expected revenue is \$2289 for this subset of songs. Applying these results to the full data implies a total expected revenue of over \$24,528.

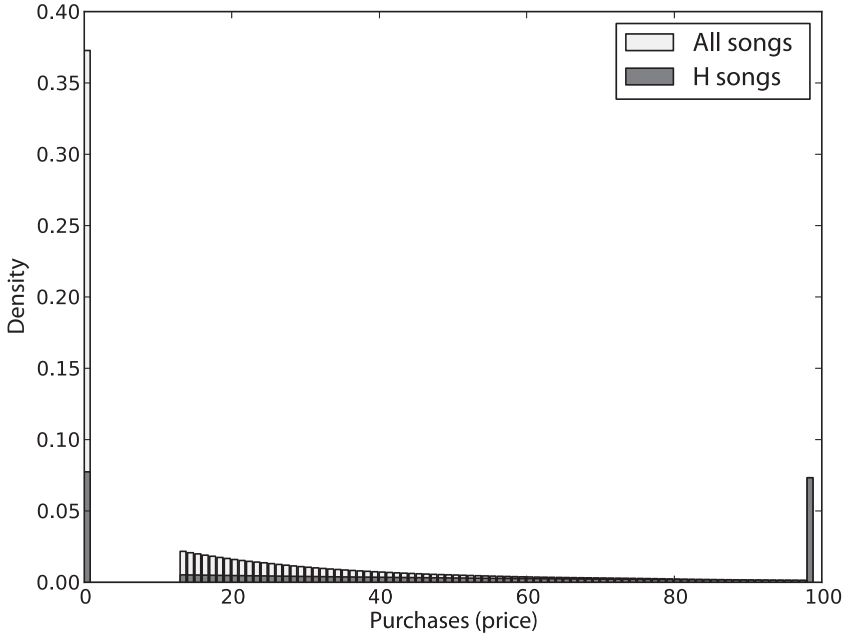
For a closer look at song failure on Amie Street (AS), I present Figure 6. The plot displays the estimated price distribution for all songs (white) and high-quality songs (gray), separately. In the region below 40¢, all songs are being listened to with high probability, resulting in a flat high-quality distribution. These songs are failing purely because of bad luck (i.e., the death rate). Once a song reaches 40¢, the uncertainty and the high price kick in, causing some individuals to

[31] This calculation relies on the assumption of normally distributed preferences. Other distributional assumptions may lead to different levels of consumer surplus, but comparisons across different environments are still informative of the counterfactual effects.

[32] The market converges in about 445 consumers.

FIGURE 6

ESTIMATED PURCHASE (PRICE) DISTRIBUTION SEPARATED BY QUALITY



Notes: The figure displays the estimated long-run price distribution for all songs (light) and H songs (dark), separately.

ignore high-quality songs. This can be seen in the "bump" in the gray distribution between 40¢ to 60¢. At prices beyond 60¢, consumers are certain songs are of high quality, leading again to a flat distribution. Therefore, the bump represents the additional high-quality songs which fail due to uncertainty in the AS market.

Note that this bump is also crucial in identifying a learning from a nonlearning environment. In a nonlearning environment, the listen and purchase rate will be constant for high-quality songs, resulting in a flat distribution over the entire price range. In the data, we see evidence of a bump between 50¢ and 70¢ (see Figure 1).

To estimate the impact of observational learning, I compare the outcomes under AS to the outcomes when observational learning is absent. In this "no-information" (NI) environment, I assume consumers only know the prior, or the overall quality of music posted on Amie Street. To isolate the effect of the additional information on AS, I hold the pricing mechanism constant (i.e., demand-based pricing). This counterfactual examines a scenario where the platform employs a demand-based pricing scheme, but consumers are not aware of it.

Row (2) of Table 7 displays the comparison between the Amie Street (AS) market and NI. The results indicate that an additional 0.91% songs fail under NI. This translates to less than one song in the primary data and eight in the full data. The reason the results are not very stark is because there exists a large amount of heterogeneity in preferences and private signals, and there is a relatively small cost of listening. This leads to a high listening rate at all prices and in all environments, regardless of song quality. Therefore, the additional information on AS does not do much to affect the outcomes of high-quality songs. However, the information on Amie Street impacts consumer surplus (1.20%) and expected revenue (2.35%) to a greater degree. This can be attributed to the fact that individuals listen more often to songs they don't purchase and/or ignore songs they would purchase under NI. To be more specific, there is 92¢ of lost consumer surplus per song under NI because individuals are wasting their time listening to bad songs. This translates to about \$180 of lost surplus for the entire market. Finally, revenue for Amie Street decreases by about \$54 when consumers do not observe learning information.

The result that more information leads to higher revenue for Amie Street is similar in spirit to the "linkage principle" in the auction literature (see Milgrom and Weber, 1982). In simple terms, the linkage principle states that it is in the seller's best interest to reveal information about the item up for auction, as asymmetric information between bidders leads to less aggressive bidding and, hence, lower expected revenue. Similarly, when consumers are uncertain of the quality of a song, they will be less likely to take action because they are reluctant to waste time on a bad song, resulting in lower revenue. This inaction due to uncertainty is analogous to bid shading in response to the "winner's curse" in the auction setting.

The final column of Table 7 indicates that the overall benefit of observational learning in this market is about \$236. Expanding this for the full data implies a benefit of around \$2528. Because the cost of providing this information only involves making consumers aware of the demand-based pricing scheme, it is reasonable to conclude that this is an overall welfare-improving policy.

I also estimate the benefits of providing additional information to the market by comparing the outcomes on AS to a perfect information environment (PI). Under perfect information, it is assumed that consumers know X but do not know u_i. Once again, I keep the pricing scheme constant to isolate the effect of information. This counterfactual examines a situation in which the platform is able to fully inform consumers of song quality. Although perfect information is unrealistic, Amie Street could provide more information about a given artist or song. One example of this is a rating system similar to what is seen on iTunes.

Results are presented in row (3) of Table 7. The AS environment is quite close to perfect information, as all the outcomes are within 0.5%. AS has a higher failure rate of high-quality songs, which is 0.15% higher than PI. This translates to less than one song in the primary data and about one in the full data.

Expected consumer surplus and expected revenue are slightly lower on AS compared to PI. The final row indicates that the uncertainty on Amie Street costs the market about \$54 total. In the full data, the cost is about \$579. Although more information benefits the market, it is likely that the costs of providing this information outweigh these benefits.

Overall, these results suggest that the online platform, the consumers, and the producers of high-quality songs all benefit from observational learning (i.e., $AS > NI$). Additionally, providing even more information to the market (i.e., PI) does not significantly improve welfare.

□ **Fixed price.** Above, I show that improving the learning environment can increase consumer surplus and expected revenue. Another way to adjust the learning environment is through the price. The reason is that the price of a song affects the listening behavior, or the incentives to experiment. For example, a lower price increases the propensity to listen through its effect on the conditional probability of purchasing.[33] This increase in listening leads to more information being transmitted to the market, which leads to a more efficient learning environment. Therefore, the "learning effect" of a lower price results in an increase in consumer surplus. Surplus is further increased through the "price effect": consumers who are purchasing the song are paying less.

A lower price also increases platform revenue through the learning effect: more people are listening to and purchasing high-quality songs. However, this comes with lost revenue due to the price effect: the platform is receiving less for each purchase. Hence, to determine the best pricing policy, the platform and/or the planner would have to weigh these positive and negative effects.

The demand-based pricing scheme (DBP) employed by Amie Street is a clever way to try to achieve an efficient learning environment without affecting revenue a great deal. Individuals are

[33] This is not necessarily true under a model in which a consumer can purchase without listening, as a lower price may lead to less information being transmitted.

TABLE 8 Fixed Price Outcomes

Row	Fixed Price	Probability of Failure for H Songs	Total Consumer Surplus	Total Revenue	Total Welfare	Total Cost or Benefit
(1)	DBP	81.8%	\$15,145	\$2,289	\$17,435	–
(2)	\$0.98	3.57%	−25.03%	141.92%	−3.10%	−\$541
(3)	\$0.90	3.10%	−22.08%	127.64%	−2.42%	−\$422
(4)	\$0.80	2.09%	−18.21%	109.69%	−1.41%	−\$246
(5)	\$0.70	1.41%	−14.31%	86.90%	−1.01%	−\$177
(6)	\$0.60	0.64%	−10.62%	65.97%	−0.25%	−\$43
(7)	\$0.50	0.08%	−6.08%	40.44%	0.03%	\$5
(8)	\$0.40	−0.55%	−1.74%	14.85%	0.43%	\$76
(9)	\$0.30	−1.35%	2.77%	−12.24%	0.80%	\$139
(10)	\$0.20	−2.03%	7.08%	−40.28%	0.86%	\$151
(11)	\$0.00	−3.21%	16.11%	−100.00%	0.86%	\$150

Notes: Table displays outcomes for demand-based pricing (DBP) in row (1) and the percentage difference between DBP and various fixed prices in rows (2) through (11). In the fixed price counterfactuals, the learning environment is kept constant.

encouraged to experiment with new music at low prices, implying that more information will be transmitted to the market. However, the price increases as learning occurs, increasing the revenue opportunities for the platform. In this section, I examine the effectiveness of this pricing scheme.

To do this, I compare the outcomes on AS to outcomes under a fixed price, or a price that remains the same throughout the life of the song. The primary reason I focus on this policy is because it was the traditional pricing scheme of the industry during the early 2000s. To isolate the effect of price, I keep the learning environment identical to AS but vary the pricing mechanism. This counterfactual explores a situation in which consumers still observe the number of purchases (and listens), but price is no longer tied to demand.

Table 8 presents the difference between DBP and various fixed prices. Not surprisingly, the learning environment becomes more efficient the lower the fixed price. This again, is due to the increased propensity to experiment at low prices. The results also indicate that consumer surplus is decreasing and expected revenue is increasing in price. Therefore, for expected revenue, the price effect outweighs the learning effect: the additional purchases of high-quality songs due to the increased learning efficiency are not enough to make up for the lost revenue from charging a low price. Total welfare is decreasing in the price, implying that the negative effects of a high price for the consumers outweigh the positive effects on the artists/platform.

I focus the reader's attention on the 98¢ price point. The results of this counterfactual show how Amie Street would have looked had it adopted the fixed price popular to the industry.[34] At this price, an additional 3.57% of high-quality songs fail. This translates to about three additional songs failing in the primary data and 29 in the full data. There is a decrease in consumer welfare of 25% and an increase in revenue of 142%. Overall, the decrease in consumer surplus outweighs the gain in revenue, leading to a welfare loss of over \$540.

Although revenue is the highest at this 98¢ price point, it is not obvious that this is the optimal policy for the platform. Each firm must weigh the benefits of various pricing schemes to consumers, producers of high-quality songs, and the platform. For example, a website like Amie Street, which encourages discovery of new music, may rely on a more efficient learning environment to bring consumers back to the website in the future. In addition, high-quality artists may be more likely to post their music on a site which has a better learning environment. iTunes, on the other hand, may be less concerned about returning customers and high-quality artists

[34] The price on iTunes was actually 99¢. Because AS does not ever reach this price, I use 98¢ as an approximation.

TABLE 9 Other Pricing Policies

Row	Informational Environment	Probability of Failure for H Songs	Total Consumer Surplus	Total Revenue	Total Welfare	Total Cost or Benefit
(1)	DBP	81.8%	\$15,145	\$2,289	\$17,435	–
(2)	Jump (15)	2.69%	−12.87%	72.40%	−1.67%	−\$291
(3)	Jump (25)	2.15%	−8.12%	45.26%	−1.11%	−\$193
(4)	Jump (50)	1.03%	−2.30%	12.56%	−0.34%	−\$60
(5)	Jump (75)	0.26%	−0.38%	2.12%	−0.05%	−\$9
(6)	Step (2)	1.77%	−6.64%	39.01%	−0.65%	−\$113
(7)	Step (5)	2.64%	−12.60%	71.54%	−1.55%	−\$270
(8)	Step (10)	2.70%	−13.46%	75.82%	−1.74%	−\$303

Notes: Table displays the outcomes for AS in row (1), the percentage difference between AS, and various other pricing schemes in rows 2–8. A jump policy is when the price follows the demand-based pricing system up to the number of downloads indicated in parentheses and then jumps to 98¢. The step policy changes the step in the price function and increases the price by the given number of cents in parentheses for each download.

because of their large presence in the market. This is one possible explanation for the different pricing policies on AS and iTunes.

There are a few other price points of interest. First, a fixed price of slightly less than 50¢ achieves the same efficiency of learning as PI. Second, a price between 50¢ and 60¢ results in welfare equal to that of PI. Finally, a price of 0 leads to the second highest amount of welfare: it benefits the market by \$150. Although it is not a perfect comparison, this may imply that piracy platforms such as Napster or BitTorrent are welfare improving as they allow for learning but keep songs free.

□ **Other pricing policies.** I examine other possible demand-based pricing policies which may increase revenue and/or consumer surplus. Although I cannot search over the entire set of policies, I examine two of which are reasonable for this marketplace. First, I allow the price to make a discrete jump to 98¢ after a certain number of downloads (hereafter, the “jump price”). Second, I allow the increase in price for each additional purchase to vary. Both of these mechanisms encourage experimentation early in a song’s life without sacrificing as much revenue as the DBP mechanism. In other words, prices are still kept low in the beginning stages to facilitate learning, but then they increase at a faster rate compared to DBP. Note that I again keep the learning environment constant in these counterfactual scenarios.

Table 9 indicates that total revenue is decreasing and total consumer surplus is increasing in the jump price. One interesting exercise is to compare this to a 98¢ fixed price, which is equivalent to a jump price of 0. Results show that encouraging the first 15 consumers to experiment through a lower price leads to a 1% reduction in the number of high-quality songs failing and increases consumer surplus by more than \$1800. However, the platform realizes lower revenues due to the fact that the first 15 people are downloading the song at low prices instead of 98¢.

Rows 6 through 8 display the results from adjusting the relationship between price and the number of downloads. The change in price for each download, or step, is indicated in parentheses in column 2. The probability that a high-quality song fails is increasing in the step, due to the fact that increasing the gradient of price increases leads to less experimentation. Not surprisingly, consumer surplus decreases and revenue increases as the step increases. Increasing the step by just 1¢ results in a decrease in total surplus of over \$100 as more high-quality songs are failing and consumers are paying more for the songs that they purchase. However, this 1¢ increase has a large impact on revenue, increasing it by nearly 40%.

Overall, the results of this exercise suggest that the Amie Street mechanism compares favorably, in terms of total welfare, to the jump price mechanism and other pricing gradients. As

TABLE 10 Comparing Policies

Row	Policy	Probability of Failure for H Songs	Total Consumer Surplus	Total Revenue	Total Welfare
(1)	PI	81.7%	$15,189	$2,300	$17,489
(2)	Fix (35)	80.8%	$15,189	$2,351	$17,540

Notes: The table displays the outcomes for PI (1), a fixed price of 35¢ (2).

discussed above, if the platform puts value on consumer welfare along with the short-run revenue, then it appears that the DBP mechanism does at least as well as these alternatives.

□ **Comparing policies.** In the previous sections, I presented the results of two possible revenue-enhancing policies for the platform: (i) increasing the amount of information and (ii) employing a fixed price. In this section, I compare these policies to determine which one is the most efficient. That is, keeping consumer surplus constant, which policy increases expected revenue by the most? To do this, I locate the fixed price that leads to the same amount of consumer surplus as the PI environment and then compare expected revenue across these scenarios. In this exercise, I exclude jump pricing and adjusting the step as options because the only way to increase consumer surplus to PI levels is to have unreasonable values for the jump price (above 98¢) and step (below 1¢).

Results are presented in Table 10. A fixed price of about 35¢ leads to expected revenue which is nearly $51 higher than the perfect information environment with the same amount of consumer surplus. This suggests that the platform can increase revenues and harm consumers the least by implementing a fixed price, rather than providing more information to the market. It is also likely that this policy would be less costly to implement.

7. Conclusion

■ I studied the impacts of observational learning in an online market for music. I was able to estimate and compare three different long-run, market-level outcomes under different learning and pricing environments. These comparisons provided a measure of the value of observational learning in the market. The data allowed me to study these long-run outcomes because of the transparency of the learning environment. I found that observational learning on Amie Street Music is valuable to consumers, producers of high-quality music, and the online platform. I also explored how different pricing schemes may hamper the learning process and found that the scheme traditionally used by the industry may lead to worse outcomes for the consumer, but better outcomes for the firm.

Although this project studied one market in particular, I believe the results can be applied to online markets with similar learning features. The market for Apps and eBooks are two prevalent examples. In general, the results provide evidence that (i) observational learning can be a valuable tool in an online market and (ii) the pricing scheme can have a significant impact on the learning process and platform revenues. Possible future work on this issue includes studying the long-run impact of other forms of learning which may be prevalent, such as word-of-mouth.

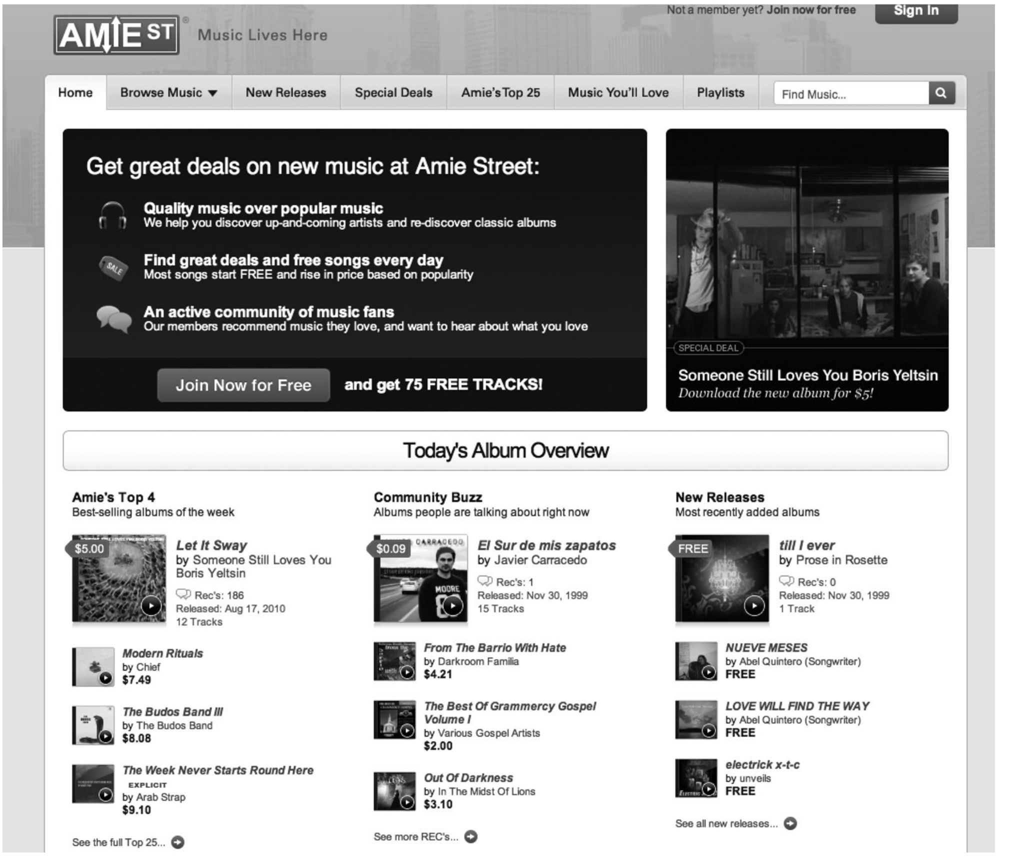
Appendix A

In appendix A, I provide a more detailed description of the data and present the results of various robustness checks.

□ **Screen shots.**

FIGURE A1

SCREENSHOT OF THE AMIE STREET HOME PAGE



□ **Alternative data.** Before the website was shut down, I downloaded information for nearly 300,000 songs which were posted on Amie Street. I applied some of the same restrictions that I did with the primary data set: first song on the first album for a given artist within the folk/country genre. I also restricted it to songs which were posted on the website at least six months before it was shut down. That way, I am hoping to only include songs which have reached their long-run outcome. After this, there are 2138 songs.

The primary reason I do not utilize these data is the fact that I do not observe the number of listens (they were removed a few weeks before). In addition, because I do not observe the weekly data, I cannot confirm that these songs have reached their long-run outcome. The price distribution of these data can be found in Figure A4.

There are a few differences with the primary data to point out. First, fewer songs fail at a 0 price. This is made up for by more mass on price points which had 0 mass in the primary data. Despite these differences, the general pattern of the data matches well. Specifically, there is a decrease of the distribution from 13¢ until about 40¢. Then, there is somewhat of a change: songs are more evenly distribution thereafter. Although the "bump" isn't as obvious in these data, it could be argued that it occurs near 70¢. There also could be a bump near 90¢.

Because of these similarities, I believe the primary data is a good approximation of what we would see with a more comprehensive data set. However, I estimate the model adding these data. I do this by computing the likelihood of the price distribution separately from the likelihood of the conditional listen distribution. Basically, this is assuming that the conditional distribution in the restricted data is what we would observe in the full data set. The remaining conditional

FIGURE A2

SCREENSHOT OF FOLK/COUNTRY SONGS FILTERED BY POPULARITY

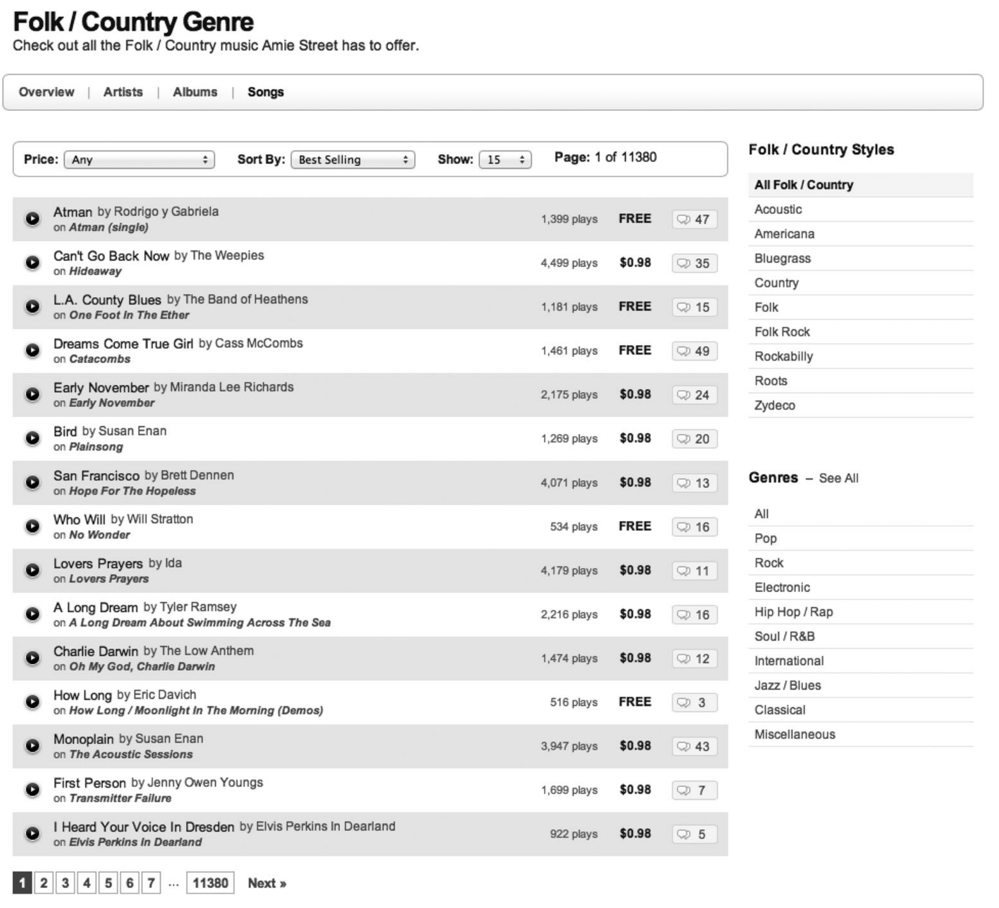


TABLE A1 Full Data Estimates

H	L	γ_u	γ_ϵ	c	λ	β
\$1.38	−\$3.65	\$4.72	\$7.94	\$0.32	0.27	0.007
(0.25)	(0.09)	(0.03)	(0.29)	(0.01)	(0.02)	(0.000)

Notes: Standard errors calculated numerically in parentheses.

listen distributions are treated as missing. The results of this exercise are very similar to the results presented (see Table A1). The main differences are in the value of the low-quality song. This can be attributed to the fewer number of songs failing at 0¢.

□ **Song heterogeneity.** One concern is that there is song heterogeneity which I am not accounting for, and this heterogeneity is what determines the learning process for songs. For instance, if an artist is already known outside of Amie Street, it may be the case that this song is likely to receive more attention on Amie Street. I address this concern in this section. Specifically, I test whether the price distribution for songs is dependent on two observables: genre and the number of albums.

A song appearing within the "folk/country" genre on Amie Street means that the artist listed "folk/country" as one of the four possible genres it belongs to. Therefore, an artist's primary genre may not be folk/country. I therefore test whether the artists which list their primary genre as "rock" have a different learning process in the Amie Street market. I choose rock because this is the genre that appears most in the data. If "rock" artists are more popular outside of Amie

FIGURE A3

SCREENSHOT OF AN ALBUM PAGE

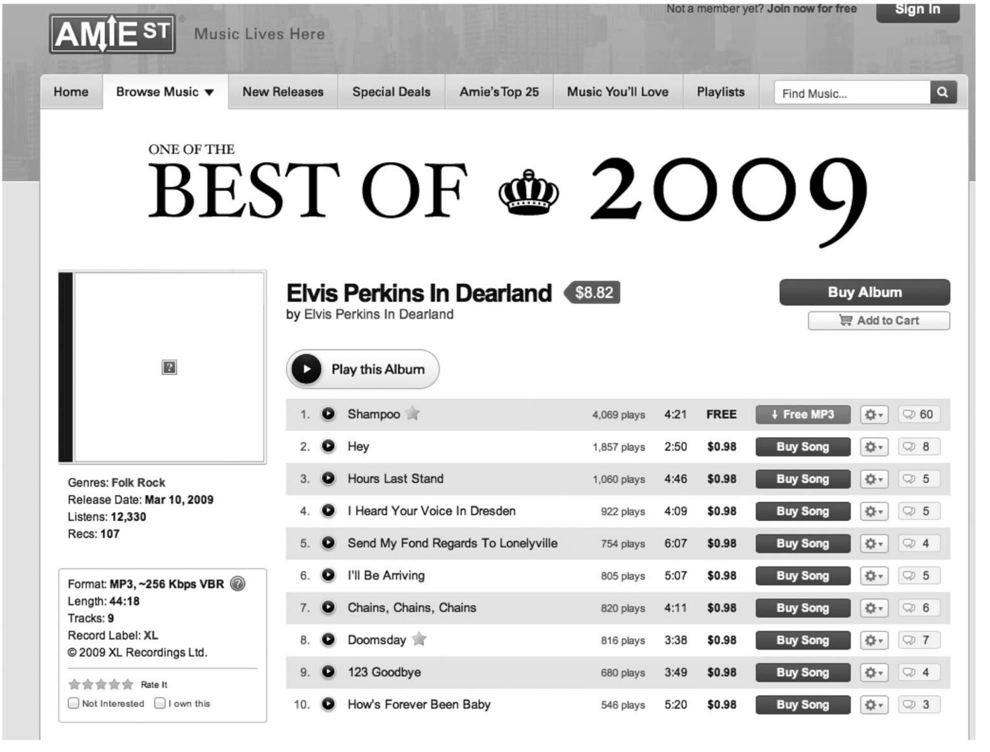
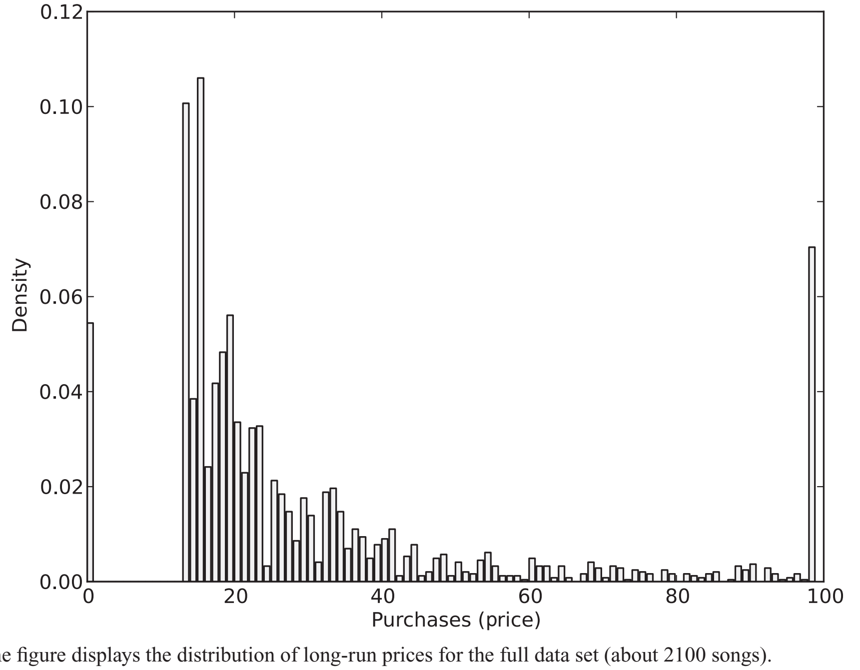


FIGURE A4

PURCHASE (PRICE) DISTRIBUTION FOR FULL DATA



Notes: The figure displays the distribution of long-run prices for the full data set (about 2100 songs).

TABLE A2 Heterogeneity in Learning Process

Variable Dep. Var.	(1) $\Delta list$	(2) $\Delta list$	(3) $\Delta list$	(4) $\Delta list$
rat	8.7*** (2.7)	5.9*** (9.4)	8.2*** (3.1)	4.0*** (2.0)
price	0.2 (0.1)	−0.3 (0.5)	0.01 (0.7)	−0.02 (0.04)
order	0.02 (0.1)	−0.00 (0.0)	−0.00 (0.0)	−0.00 (0.0)
Subset	Rock	Oth. Genre	1 Album	>1 Album
R^2	0.16	0.26	0.05	0.78
Obs.	270	1,141	802	595
Songs	20	83	58	43

$^{***}p < 0.01$, $^{**}p < 0.05$.

Notes: The table displays the results of regressions of weekly listens/buys on the value of the independent variable at the beginning of the week. A week dummy is included in all specifications. *rat* is the ratio of purchases to listens and *order* is a measure of how long the song has been on Amie Street. Data is limited to prices between 13¢ and 97¢. Specification 1 includes only songs by "rock" artists, (2) includes songs by all other artists, (3) includes songs by artists which have only one album on Amie Street, and (4) includes songs by artists who have more than one album on Amie Street.

TABLE A3 Heterogeneity in Price Distribution

	(1)	(2)	(3)	(4)
p-val.	0.70	0.67	0.54	0.06
Data	Primary	Primary	Full	Full
Test	Genre	Albums	Genre	Albums

Notes: The table displays the p-value for a Kolmogorov-Smirnov test of equal price distributions. Test (1) uses the primary data set and separates distributions by genre and (2) separates them by albums. Tests (3) and (4) do the same but use the full data set.

Street because they are more likely to be played on the radio, for example, then they may also be more popular on Amie Street. This would lead to a different learning process for "rock" artists, which violates the assumption that all artists are the same, *ex ante*.

Additionally, I test whether the artists with more than one album posted have a different learning process on Amie Street. Although I restrict the data to only include the first song off an artist's first album, it doesn't mean that the artists have only one album.

I perform these tests in two primary ways. First, I run the same regression that appears in Section 4 but separate the data into the different groups discussed above. The idea is that if the learning process is different, then the results of this regression should look different. The estimates are presented in Table A2. Specifications 1 and 2 separate the data into rock artists and other artists, whereas specifications 3 and 4 separate the data by the number of albums. Although there are some differences in the estimates, the key result from this exercise is that learning appears to be occurring across all different subsets. That is, the coefficient on *rat* is positive and significant for all specifications.

Second, I compare the final price distribution across these different subsets of data. If the learning process is heterogeneous based on these observables, then the final price distribution will be different across these subsets. Therefore, I perform Kolmogorov-Smirnov tests of equal price distributions across genre and number of albums. I perform these tests using the primary data set and the full data set discussed above.

Results are presented in Table A3. In all tests, I cannot reject that the price distribution is the same based on the observables (at the 5% level).

Although I cannot rule out that song heterogeneity plays a role in the data generating process, I take these tests as evidence that it is not the driving force in the outcomes I observe in the data.

□ **Testing properties of the model.** I provide evidence of two intuitive properties of the model via simulation. Specifically, I show that (property 1) the likelihood of listening is increasing in the purchase-to-listen ratio, and (property 2) the effect of the purchase-to-listen ratio is increasing in the number of listens, all else equal. To do this, I use equation (4) to iterate over consumer decisions until $i = \bar{i}$, where $\bar{i} \in \{100, 150, 200, 250, 300\}$. At these points, I calculate the probability of listening to a song at every possible l and p. This is given by:

$$\alpha(\mu(l, p, \bar{i})).$$

TABLE A4 Model Simulation

listen Range	Estimate
$listen \leq 50$	0.19^{***} (0.00)
$50 < listen \leq 75$	0.21^{***} (0.00)
$75 < listen \leq 100$	0.22^{***} (0.00)
$100 < listen$	0.25^{***} (0.00)

$^{***}p < 0.01$, $^{*}p < 0.05$.

Notes: The table displays the results of a regression of the listen probability on interactions between rat and dummy variables indicating the number of listens, along with *price*, and $\bar{i}$. The data are simulated from the model assuming the estimated values of the parameters. Data are limited to prices between 13¢ and 97¢. Displayed are the coefficients on the interaction variables.

TABLE A5 Facebook Fan Page Likes by Purchases

Buys (Price)	FB Likes Median
[13, ∞]	25
[98, ∞]	9221
[13, 97]	0
[13, 29]	0
[30, 97]	3424

Notes: I set FB likes to 0 if the artist did not have a Facebook fan page, which was approximately 45% of them. The table displays the median FB likes for different price ranges of folk songs on Amie Street.

Therefore, the "data" to test properties 1 and 2 are α, *price*, and the purchase-to-listen ratio (rat) for all combinations of l and p at each $\bar{i}$. I create the data assuming the estimated parameters found in Table 5.

To test property 1, I regress α on rat, along with the *price* and $\bar{i}$. This is similar to the regressions run in Table 3. The coefficient on the purchase-to-listen ratio is 0.19 and is significant at the 1% level, indicating that people are more likely to listen to a song the higher the purchase-to-listen ratio.

To test property 2, I regress α on interactions between rat and dummy variables based on the number of listens, along with *price* and $\bar{i}$. This is similar to the regression run in Table 4. Results are found in Table A4. The coefficient on the purchase-to-listen ratio is increasing in the number of listens, which suggests that property 2 holds.

□ **High-quality songs failing.** I provide some evidence that the outcomes of the estimation are reasonable.

I do not have an exogenous measure of quality in my data, so I cannot directly say which songs were high quality and failed. Instead, I use Facebook (FB) fan page "likes" as a somewhat exogenous measure of artist quality. This is not perfect but allows me to loosely test how many high-quality songs there are in the data which failed. For this to be valid, I assume that the number of FB likes collected in November 2012 is directly related to the quality of a song by an artist: the more FB likes, the higher the likelihood that a song by that artist is high quality. Table A5 presents the median FB likes for the songs in my data. If the artists did not have a FB fan page, I set their number of likes to 0. There is a clear correlation between making it to 98¢ and the amount of FB likes an artist has. Songs which made it to a high price before dying also have significantly more likes (i.e., are more likely to be high quality), which is one of the predictions of the model. Although this provides some evidence that high-quality songs are dying, a more accurate test of the model would be to see how well it did in predicting the number of failed high-quality songs. The model predicts this to be 63.

I perform two exercises to analyze the accuracy of this number. First, I take the lower bound of FB likes for songs which made it to 98¢ (632), assume that any artist who has at least this many FB likes is high quality, and count how many of these high-quality songs failed (28). Here, the model overpredicts the failure of high-quality songs.

In the second approach, I find the FB likes of the 63rd "best" failed song (72) and assume this is the lower bound of FB likes for high-quality artists. I then compare this to the FB likes of the lowest 98¢ song: 632. Again, there is an overprediction of failure. The primary reason for this is the fact that the data are very limited between 30¢ and 80¢. However, the model predicts that there will be a continuous distribution of songs in this region. In addition, these are likely to be high quality. Therefore, I believe if the data were more substantive (e.g., the full data set), the results of the model would match the Facebook data more closely.

At the very least, this exercise shows that there are likely some high-quality songs which failed on Amie Street. Table A6 lists some possible examples of these artists. The first six artists have a relatively high purchase-to-listen ratio, implying that they are high-quality songs which were randomly hit with the death rate. The purchase-to-listen ratio of the remaining four implies that they might be high-quality songs, which failed because of learning. If this is the case, the

TABLE A6 Failed High-Quality Artists

Artist	Listens	Buys	FB Likes
Corey Tremaine	29	14	649
David Story	20	15	1,170
Megan Callahan	20	17	697
Jared Mees and the Grown Children	25	21	2,003
Danny Barnes	35	34	523
Hope Sandoval and the Warm Inventions	62	48	94,381
Daniel Ward-Murphy	227	51	2,014
Sofia Tavlik	327	64	4,834
Ben Nichols	345	74	720
Amelia Curran	270	78	5,369

Notes: The table displays examples of some artists who have many FB likes but were not successful on Amie Street.

model underpredicts the occurrence of this outcome. Recall that the model predicted less than one song dying because of learning.

Appendix B

In appendix B, I provide further details about the estimation of the model and discuss identification.

□ **Example of learning process.** I provide an example of the learning process for the first three arriving consumers to solidify the intuition of how the model works. For the purposes of this exercise, I assume that price and listens are observable at prices below 13¢. Additionally, I assume that the parameters are equal to the estimated parameters found in Table 5.

Consumer $i = 1$ arrives with her belief equal to the prior, $\mu_1 = \lambda$. Based on this belief, she forms her cutoff value of the private signal, $\bar{S}(\lambda) = -10.47$, by finding the S that equates equation (2) to c. If she receives a private signal, $S = u_i + \epsilon_i$, higher than this value, she listens. If she does not, she ignores the song and exits the market. Once she listens, she learns her exact utility and will choose to purchase if $X + u_i - p_i = X + u_i \geq 0$.

Consumer $i = 2$ arrives and observes one of three possible states: $\{(l = 0, p = 0), (l = 1, p = 0), (l = 1, p = 1)\}$. She can calculate the probability of observing each of these states conditional on the song quality. Note that she only needs to calculate the probability of the *observed* state, but I demonstrate the calculation at *all* states to provide a complete outline of the updating process.

Specifically, suppose $X = H = 1.37$. Then, the probability consumer 1 arrived, formed her belief (λ), and ignored the song is given by:

$$\begin{aligned} 1 - \alpha(\lambda) &= 1 - Pr(S_i \geq -10.47) \\ &= F_{u+\epsilon}(-10.47) \\ &= 0.13. \end{aligned}$$

Note that this probability is the same if the song is low quality. Additionally, she can calculate the probability consumer 1 listened but didn't purchase:

$$\begin{aligned} \alpha(\lambda)(1 - \beta(X = 1.37, \lambda)) &= Pr(S_i \geq -10.47)(1 - Pr(w_i \geq 0 | X = 1.37, S_i \geq -10.47)) \\ &= 0.87 * 0.32 \\ &= 0.29. \end{aligned}$$

The conditional probability of purchase (the second term) is calculated by simulation. The probability that consumer 1 listened and purchased is:

$$\begin{aligned} \alpha(\lambda)(\beta(X = 1.37, \lambda)) &= Pr(S_i \geq -10.47)(Pr(w_i \geq 0 | X = 1.37, S_i \geq -10.47)) \\ &= 0.87 * 0.68 \\ &= 0.59. \end{aligned}$$

The equivalent probabilities can be calculated for $X = L = -3.43$:

$$\begin{aligned} \alpha(\lambda)(1 - \beta(X = -3.43, \lambda)) &= Pr(S_i \geq -10.47)(1 - Pr(w_i \geq 0 | X = -3.43, S_i \geq -10.47)) \\ &= 0.87 * 0.79 \\ &= 0.69, \end{aligned}$$

and,

$$\begin{aligned}\alpha(\lambda)(\beta(X=1.37,\lambda)) &= Pr(S_i \geq -10.47)(Pr(w_i \geq 0|X=1.37, S_i \geq -10.47))\\ &= 0.87 * 0.21\\ &= 0.18.\end{aligned}$$

Given these values, the probability that consumer i observes each state, conditional on the song being high quality, is given by the following matrix:

$$\pi(l,p,2|H) = \begin{pmatrix} 0.13 & \\ 0.29 & 0.59 \end{pmatrix}.$$

Conditional on the song being low quality, the matrix is:

$$\pi(l,p,2|L) = \begin{pmatrix} 0.13 & \\ 0.69 & 0.18 \end{pmatrix},$$

where l represents the index in the first dimension of the matrix (plus 1) and p represents the index in the second dimension (plus 1). Therefore, the belief of consumer 2 is given by equation (3):

$$\mu(l,p,2) = \lambda\frac{\pi((l,p,2)|H)}{\lambda\pi((l,p,2)|H) + (1-\lambda)\pi((l,p,2)|L)} = \begin{pmatrix} 0.40 & \\ 0.21 & 0.68 \end{pmatrix}.$$

That is, if consumer 2 observes ($l=0, p=0$), then her belief is $\mu_2 = 0.40$. If she observes ($l=1, p=0$), then her belief is $\mu_2 = 0.21$. Finally, if she observes ($l=1, p=1$), then her belief is $\mu_2 = 0.68$. With this belief, consumer 2 forms her cutoff value for the private signal:

$$\bar{S}(\mu(l,p,2)) = \begin{pmatrix} -10.47 & \\ -9.42 & -11.68 \end{pmatrix}.$$

If she receives a private signal higher than this, she will listen, otherwise, she will ignore the song. Once she listens, she observes her true utility and purchases if $X + u_2 - p_2 \geq 0$.

Consumer 3 then arrives and observes one of six possible states: $\{(l=0, p=0), (l=1, p=0), (l=1, p=1), (l=2, p=0), (l=2, p=1), (l=2, p=2)\}$. She must calculate the probability of observing the given state, conditional on song quality. To do this, she calculates the probability of each action consumer 2 could have taken to lead to the observed state. First, the probability consumer 2 arrived, was in state ($l=0, p=0$), and ignored the song is given by:

$$\begin{aligned}1-\alpha(\mu(0,0)) &= 1 - Pr(S_i \geq -10.47)\\ &= F_{u+\epsilon}(-10.47)\\ &= 0.13.\end{aligned}$$

The probability consumer 2 listened but didn't purchase in this state:

$$\begin{aligned}\alpha(\mu(0,0))(1-\beta(X=1.37,\mu(0,0))) &= Pr(S_i \geq -10.47)(1 - Pr(w_i \geq 0|X=1.37, S_i \geq -10.47))\\ &= 0.87 * 0.32\\ &= 0.29.\end{aligned}$$

Finally, the probability that consumer 2 listened and purchased is:

$$\begin{aligned}\alpha(\mu(0,0))(\beta(X=1.37,\mu(0,0))) &= Pr(S_i \geq -10.47)(Pr(w_i \geq 0|X=1.37, S_i \geq -10.47))\\ &= 0.87 * 0.68\\ &= 0.59.\end{aligned}$$

These same probabilities in state ($l=1, p=0$) are:

$$\begin{aligned}1-\alpha(\mu(1,0)) &= 1 - Pr(S_i \geq -9.42)\\ &= F_{u+\epsilon}(-9.42)\\ &= 0.16,\end{aligned}$$

$$\begin{aligned}\alpha(\mu(1,0))(1-\beta(X=1.37,\mu(1,0))) &= Pr(S_i \geq -9.42)(1 - Pr(w_i \geq 0|X=1.37, S_i \geq -9.42))\\ &= 0.84 * 0.31\\ &= 0.26,\end{aligned}$$

and

$$\begin{aligned}\alpha(\mu(1,0))(\beta(X=1.37,\mu(1,0))) &= Pr(S_i \geq -9.42)(Pr(w_i \geq 0|X=1.37, S_i \geq -9.42))\\ &= 0.84 * 0.69\\ &= 0.58.\end{aligned}$$

Finally, the probabilities in state ($l = 1$, $p = 1$) are:

$$\begin{aligned}1-\alpha(\mu(1,1)) &= 1 - Pr(S_i \geq -11.68)\\ &= F_{u+\epsilon}(-11.68)\\ &= 0.11,\end{aligned}$$

$$\begin{aligned}\alpha(\mu(1,1))(1-\beta(X=1.37,\mu(1,1))) &= Pr(S_i \geq -11.68)(1 - Pr(w_i \geq 0|X=1.37, S_i \geq -11.68))\\ &= 0.89 * 0.32\\ &= 0.28,\end{aligned}$$

and

$$\begin{aligned}\alpha(\mu(1,1))(\beta(X=1.37,\mu(1,1))) &= Pr(S_i \geq -11.68)(Pr(w_i \geq 0|X=1.37, S_i \geq -11.68))\\ &= 0.89 * 0.68\\ &= 0.61.\end{aligned}$$

These are also calculated for $X = L$. With this, consumer 3 can form the likelihood of observing each of the six possible states, conditional on song quality:

$$\pi(l,p,3|H) = \begin{pmatrix} 0.02 & & \\ 0.08 & 0.14 & \\ 0.07 & 0.33 & 0.36 \end{pmatrix},$$

and if the song is low quality:

$$\pi(l,p,3|L) = \begin{pmatrix} 0.02 & & \\ 0.20 & 0.05 & \\ 0.44 & 0.26 & 0.03 \end{pmatrix}.$$

These numbers are formed using the updating procedure in equation (4). Here is an example:

$$\begin{aligned}\pi(1,0,3|H) &= \pi(1,0,2|H)(1-\alpha(\mu_2)) + \pi(0,0,2|H)\alpha(\mu_2)(1 - Pr(w_i \geq 0|X, S_i \geq \bar{S}(\mu_2)))\\ &= 0.29 * 0.16 + 0.13 * 0.29\\ &= 0.084.\end{aligned}$$

From this, beliefs can be calculated at each state:

$$\mu(l,p,3) = \lambda\frac{\pi((l,p,3)|H)}{\lambda\pi((l,p,3)|H) + (1-\lambda)\pi((l,p,3)|L)} = \begin{pmatrix} 0.40 & & \\ 0.21 & 0.68 & \\ 0.10 & 0.46 & 0.87 \end{pmatrix}.$$

Note that the first two rows are identical to the beliefs that consumer 2 held. This is because the private signal is only informative about u as opposed to $X + u$. Therefore, the fact that people ignore a song does not provide any information to other consumers and hence, this model is equivalent to one in which the consumer does not know her arrival order i. See Section 3 for a discussion.

With this belief, consumer 3 forms her cutoff value for the private signal, and so on. This process continues until the model converges.

□ **Calculating cutoff *S*.** I present the procedure to calculate the cutoff level of the private signal $\bar{S}(\mu_i)$. With the parametric assumptions, there is no closed form solution for this. Therefore, I use the bisection method to calculate the S_i that solves:

$$0 = E[W|S_i, \mu_i] - c.$$

I calculate this for every possible (l, p, i) combination.

□ **Calculating the expected value of listening.** The probability a consumer buys given she listens, or:

$$Pr(w_i \geq 0|X, S_i \geq \bar{S}(\mu_i)) = \beta(X, \mu_i),$$

does not have a closed form solution. To approximate this, I simulate 1000 values for u_i and ϵ_i and calculate the percentage of consumers who would purchase, given they listen.

□ **0 probabilities.** The computer predicts a 0 probability for some of the data points because the numbers get "too" small. Theoretically, the model should never predict a 0 probability for $\Pi_d(l, p, I|\tilde{\theta})$, $\Pi_0(I|\tilde{\theta})$, or $\Pi_{98}(I|\tilde{\theta})$ in the long run. The 0 probabilities are a problem when I try to calculate the log-likelihood function. To correct for this, I add 1×10^{-150} to each likelihood the model predicts and then recalculate probabilities so they sum to one.

□ **Learning under DBP.** I presented the updating procedure in the body of the article assuming that the number of purchases is observable at all times. The demand-based pricing scheme, however, restricts the observable number of purchases to when they are between 13 and 97. The way to modify the learning process to include this feature is straightforward. Simply, the consumer can calculate $\pi((l, 0, i|X)$ by summing over the probability of all possible numbers of purchases below 13:

$$\pi((l, 0, i|X) = \sum_{p \in (0,12)} \pi((l, p, i)|X).$$

Similarly,

$$\pi((l, 98, i|X) = \sum_{p \geq 98} \pi((l, p, i)|X).$$

With these, the belief at any $(l, 0, i)$ or $(l, 98, i)$ can be calculated using Bayes' rule.

□ **Identification tests.** I perform the sensitivity analysis found in Gentzkow and Shapiro (2014), which proposes a simple test to determine the variation in the data which identifies each parameter. I take the example in Section 4.2 of Gentzkow and Shapiro (2014) as a guide and calculate the scaled score of each observation with respect to each of the parameters. I then regress the score on the moments in the data which I believe identify the parameters. The higher the value of the estimated coefficient, the more power that moment has in identifying the given parameter.

Per the discussion in Section 5, identification comes from moments in separate regions of the price distribution. Therefore, I pool the data by these regions, and then run separate regressions for each region. For example, I regress the score for observations which died between 13¢ and 25¢, g_i, on the value of the moments of the observations in that region:

- The average price of songs which died between 15¢ and 25¢: $\tilde{m}_{1i} = (p_i - \bar{p})$
- The variance in price of songs which died between 15¢ and 25¢: $\tilde{m}_{2i} = (p_i - \bar{p})^2$
- The average listens of songs which died between 15¢ and 25¢: $\tilde{m}_{3i} = (l_i - \bar{l})$
- The variance in listens of songs which died between 15¢ and 25¢: $\tilde{m}_{4i} = (l_i - \bar{l})^2$
- The covariance in price and listens of songs which died between 15¢ and 25¢: $\tilde{m}_{5i} = (p_i - \bar{p})(l_i - \bar{l})$

The results of this regression and the regressions for the other regions are displayed in Figure B1. The rows of the heat map present the coefficients of the specific moment, with the price region in parentheses. Each column represents a parameter.

For the most part, these results are in line with the identification discussion. The value of the high-quality song, H, is identified by the distribution of prices for songs above 40¢ and, more specifically, the number of songs which reach 98¢. The latter seems to be more important as the coefficient on the mean price for prices between 41 and 97 is close to zero, and the coefficient for prices between 41¢ and 98¢ is close to one. Therefore, the 98¢ songs must lead to a significant amount of variation in the score. Results are similar for the low-quality songs: the coefficient on the mean of the songs between 13¢ and 25¢ is low, whereas the coefficient when including 0¢ songs is around 0.7.

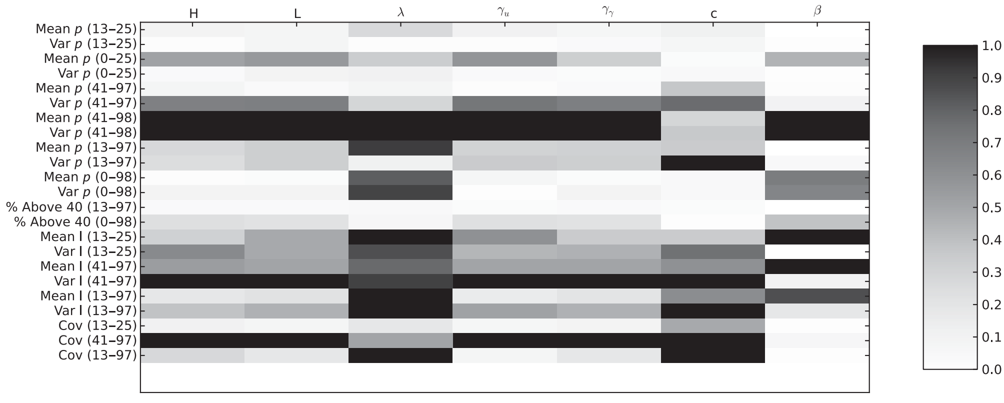
Additionally, it appears that the mean price of songs between between 0¢ and 25¢ helps to identify the variation in preferences, σ_u. This is not exactly in line with the identification discussion but makes sense given the truncation of the price distribution at 0: a higher amount of variation leads to a higher average price. Further, the mean number of listens for songs between 13 and 25 seems to have some identifying power for σ_u.

The figure also indicates that the learning parameters are identified by the listen distribution. Specifically, the listening cost, c, seems to have an impact on average and the variance of listen distribution. On the other hand, σ_γ impacts mostly the variance.

The prior, λ, appears to affect many moments, with the mean and variance of price for all songs being the most sensitive. This makes sense, as the prior not only affects consumer behavior, but is also assumed to be the true distribution of quality in the data. The death rate has the most impact on the mean of the listen distribution without affecting the variance a great deal. The reason for this is the fact that an increased death rate implies songs will not live as long, and therefore, will receive fewer listens.

FIGURE B1

STANDARDIZED SENSITIVITY FOR ESTIMATED PARAMETERS



Notes: The figure displays the estimated standardized sensitivity using the methods proposed in Gentzkow and Shapiro (2014). Some of the moments are separated into two separate price bins: songs which die under 25¢ and songs which die above 40¢. The darkest regions of the heat map are coefficients which are greater than or equal to 1. l indicates listens and p indicates price.

I note that the results of this analysis also provide evidence that the parameters are separately identified from one another. Specifically, if two columns are identical, then it implies that the parameters respond equally to the same variation in the data. Although some are similar (e.g., L and σ_u), there appears to be at least some differences in all columns.

For more details regarding these sensitivity tests, see Gentzkow and Shapiro (2014).

References

ACKERBERG, D.A. "Advertising, Learning and Consumer Choice in Experience Goods Markets: An Empirical Examination." *International Economic Review*, Vol. 44(3) (2003), pp. 1007–1040.

APPEL, M., LABARRE, R., AND RADULOVIC, D. "On Accelerated Random Search." *SIAM Journal of Optimization*, Vol. 14(3) (2003), pp. 708–731.

BANERJEE, A. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics*, Vol. 107(3) (1992), pp. 797–817.

BHIKCHANDANI, S., HIRSHLEIFER, D., AND WELCH, I. "A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades." *Journal of Political Economy*, Vol. 100(5) (1992), pp. 992–1026.

BOSE, S., OROSEL, G., OTTAVIANI, M., AND VESTERLUND, L. "Dynamic Monopoly Pricing and Herding." *The RAND Journal of Economics*, Vol. 37(4) (2006), pp. 910–928.

———. "Monopoly Pricing in the Binary Herding Model." *Economic Theory*, Vol. 37(2) (2008), pp. 203–241.

CAI, H., CHEN, Y., AND FANG, H. "Observational Learning: Evidence from a Randomized Field Experiment." *American Economic Review*, Vol. 99(3) (2009), pp. 864–882.

CHEVALIER, J.A. AND MAYZLIN, D. "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research*, Vol. 43(3) (2006), pp. 345–354.

CRAWFORD, G.S. AND SHUM, M. "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica*, Vol. 73(4) (2005), pp. 1137–1173.

ELLISON, G. AND FUDENBERG, D. "Rules of Thumb for Social Learning." *Journal of Political Economy*, Vol. 101(4) (1993), pp. 612–643.

ERDEM, T. AND KEANE, M.P. "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science*, Vol. 15(1) (1996), pp. 1–20.

GENTZKOW, M. AND SHAPIRO, J.M. "Measuring the Sensitivity of Parameter Estimates to Sample Statistics." NBER Working Paper no. w20673, 2014.

HENDRICKS, K. AND SORENSEN, A. "Information and the Skewness of Music Sales." *Journal of Political Economy*, Vol. 117(2) (2009), pp. 324–369.

HENDRICKS, K., SORENSEN, A., AND WISEMAN, T. "Observational Learning and Demand for Search Goods." *American Economic Journal: Microeconomics*, Vol. 4(1) (2012), pp. 1–31.

KNIGHT, B. AND SCHIFF, N. "Momentum and Social Learning in Presidential Primaries." *Journal of Political Economy*, Vol. 118(6) (2010), pp. 1110–1150.

LUCA, M. "Reviews, Reputation, and Revenue: The Case of Yelp.com." Harvard Business School Working Paper No. 12-016, 2011.

MILGROM, P. AND WEBER, R. "A Theory of Auctions and Competitive Bidding." *Econometrica*, Vol. 50(5) (1982), pp. 1089–1122.

OLBERHOLZER-GEE, F. AND STRUMPF, K. "The Effect of File Sharing on Record Sales: An Empirical Analysis." *Journal of Political Economy*, Vol. 115(1) (2007), pp. 1–42.

SALGANIK, M.J., DODDS, P.S., AND WATTS, D.J. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science Magazine*, Vol. 311 (2006), pp. 854–856.

SHILLER, B. AND WALDFOGEL, J. "Music for a Song: An Empirical Look at Uniform Pricing and Its Alternatives." *Journal of Industrial Economics*, Vol. 59(4) (2011), pp. 630–660.

SMITH, L. AND SØRENSEN, P. "Pathological Outcomes of Observational Learning." *Econometrica*, Vol. 68(2) (2000), pp. 371–398.

SORENSEN, A. "Social Learning and Health Plan Choice." *RAND Journal of Economics*, Vol. 37(4) (2006), pp. 929–945.

WALDFOGEL, J. "Copyright Protection, Technological Change, and the Quality of New Products: Evidence from Recorded Music since Napster." *Journal of Law and Economics*, Vol. 55(4) (2012), pp. 715–740.

ZHANG, J. "The Sound of Silence: Observational Learning in The U.S. Kidney Market." *Marketing Science*, Vol. 29(2) (2010), pp. 315–335.