An Empirical Study of Observational Learning

Peter W. Newberry*  
The Pennsylvania State University  
June 12, 2013

Abstract

In this paper, I estimate the effect of observational learning in an online market for music. I find that observational learning is a valuable tool for the producers of high quality products and for consumers, but not necessarily for the firm running the market. I also study the role of pricing as a friction to the learning process by comparing outcomes under a demand-based pricing scheme to the counterfactual outcomes under a fixed price. I find that a price of 99 cents per song (the traditional price in the industry) hampers learning by reducing the incentive to experiment.

Keywords: Observational Learning, Pricing, Online Markets, Recorded Music, Search Goods

JEL Codes: L15, L82, M31, D83

*Contact information: pwnewberry@psu.edu. I am grateful to J-F Houde, Ken Hendricks and Alan Sorensen for their support and advice on this project. I also am thankful for comments and suggestions from Paul Grieco, Dan Quint, Chris Adams, Mark Roberts, Russ Cooper, Amit Gandhi and numerous other seminar and conference participants. All remaining errors are my own.
1 Introduction

When a consumer is uncertain about a new product, she may rely on information provided by her peers as a signal of the unknown quality. Opportunities for this social learning exist in many markets today, with product reviews and/or popularity rankings being especially prevalent. The latter is an example of observational learning: learning from observing the purchase decisions of your peers. In this paper, I estimate the effect of observational learning on three market-level outcomes: the probability of success for a high quality product (i.e., discovery), the expected consumer welfare, and the expected revenue. This provides market participants and policy makers with an overall measure of the value of observational learning. The uniqueness of this exercise lies in the fact that I empirically investigate the implications of observational learning, rather than test for its existence. In addition, I study the effect of policies that serve as frictions to the learning process. Specifically, I examine how different pricing schemes may hamper learning and lead to undesirable market outcomes for consumers and producers of high quality products.

A reason why the impact of observational learning is generally difficult to study empirically is that the researcher must observe when the effect of learning is complete, or when a product has either been discovered or failed. ‘Convergence’ of this nature is both difficult to define and rarely observed in the data. The structure of the online music market I study, called Amie Street Music, allows me to define success versus failure in a very straightforward way: songs which make it to a demand threshold are successful, while the others are not. Also, the data indicate that convergence to one of these outcomes not only occurs, but occurs very quickly: within one week, most products have either succeeded or failed.\footnote{I formally define success or failure in the model section.}

A second reason why this topic is difficult to study is that there is rarely a market where the learning process of consumers is transparent, making it difficult to specify a structural model of consumer decision making. For example, in the restaurant industry a researcher likely does not observe the information each consumer received and how they interpreted it. The observational learning information provided on Amie Street Music, on the other hand, is available to all consumers and has a straightforward interpretation for both the consumer and the researcher. Specifically, Amie Street Music 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 cents. This scheme not only encourages consumers to experiment with new music at cheap prices, 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 observational learning information: 1) the price, which is a signal of the number of purchases and 2) the number of listens.

The transparency of these signals allows me to formulate a tractable structural model of obser-
vational learning in this market. The key feature of the model is the fact that preferences include a quality component (either high or low) which is common to all individuals. Therefore, a consumer can learn something about her own utility by observing the actions of others and forming a belief about whether the song is high or low quality based on this information.

The model has two important features which differentiate it from standard herding models (e.g., Banerjee (1992), Bikhchandani et al. (1992), and Welch (1992)). The first is the fact that a consumer observes the aggregate decisions of all the consumers which arrived before her, rather than the actions of each individual. This means that the updating procedure is not a martingale (Hendricks et al., 2012) and, therefore, is computationally intensive: one must integrate over all the possible paths a song could have taken to get to a given price and listen combination. It becomes especially difficult if the number of consumer arrivals is large. The second is that the learning information is important for a consumer’s ‘search’ decision instead of the purchase decision. The model assumes that a consumer arrives to the market, observes the price and listens, and then decides to either ignore the song or listen to the sample based on this information. If she listens, all uncertainty disappears, and the purchase decision is straightforward. With this set up, observational learning impacts the decision to listen to the sample, but does not affect the decision to buy the song conditional on listening.

Rather than analyze this process at the individual level, I focus on how observational learning affects the probability of long-run success or failure of products. I define success as reaching the 98 cent threshold, while failure is when a song’s demand dies off below 98 cents. Given the assumptions of the model, songs which are ignored more often have a higher likelihood of failing, meaning that the short-run learning process is an important determinant in a song’s long-run outcome. The existence of heterogeneity in both preferences and private signals leads to the possibility of a high quality song failing. For example, a consumer may arrive to a song and ignore it because of a negative signal (i.e., low belief). However, the low belief might result from a group of individuals who have unique tastes for a high quality song and decided not to purchase after listening to it. The more often this occurs, the higher the probability the high quality song fails. The primary goal of this paper is to measure the impact of observational learning on the likelihood of this bad outcome, along with consumer surplus and revenue.

To accomplish this, I estimate the parameters of the learning model using data scraped from the Amie Street site. By iterating over consumers until the model has converged, I am able to predict the probability of a song failing at a given price and listen combination and the probability of a song reaching 98 cents. This process produces the long-run joint distribution of listens and price, which, along with the observed outcomes for failed and successful songs, forms the likelihood

---

2Note that I estimate the model without any exogenous measure of song quality.
3Newberry (2013) argues that the search decision is often the one which is impacted by social learning, especially in online markets.
4This is analogous to a ‘bad herd’ discussed in (Hendricks et al., 2012). However, due to assumptions made to fit the data, the interpretation of bad herds and how and why they happen differs from Hendricks et al. (2012).
function. I then find the parameters which maximize the likelihood of the data. 

With the estimated parameters, I calculate three long-run outcomes: 1) the probability a high quality song fails, 2) the expected consumer welfare, and 3) the expected revenue for the market. These outcomes serve as a measure of market value to producers of high quality products, consumers and the website as a whole, respectively. In addition, outcome (1) is an estimate of how efficient the learning is, as fewer high quality songs failing implies more efficient learning. In order to assess the value of the observational learning in this market, I estimate the same outcomes for two counterfactual environments: full information and no information. In the full information case, which serves as the baseline model, consumers know exactly the quality of the song, but still do not know their specific tastes. The no information case is one in which consumers only know the prior, or the percentage of songs on Amie Street which are high quality.

I compare the outcomes in these three environments and find that the Amie Street is comparable to perfect information in two of the three: the probability of failure for high quality songs and expected consumer surplus. The reason for this result is the speed at which consumers learn: by the 22nd consumer, consumers are certain which songs are high quality and which are low quality. This means that the failure of high quality songs which can be attributed to uncertainty occurs before the arrival of the 22nd consumer. Results for the third outcome, expected revenue, are different: Amie Street is comparable to the no information case. This is because of the lost revenue of the additional failed high quality songs.\textsuperscript{5}

I also investigate how the choice of pricing scheme can impact market outcomes. Different pricing schemes can act as frictions to learning if they discourage experimentation, which limits the amount of information that is passed to the market. Therefore, the price elasticity in a market with observational learning is not only a function of an individual’s price sensitivity, but it also includes the effect of price on the propensity to experiment. To analyze this, I compare how two different schemes which are used in the industry affect market-level outcomes. While the results of this exercise serve as a general analysis of pricing in these types of markets, they are of particular interest in online music industry because of the tradition of a fixed price of 99 cents per song (e.g., iTunes).\textsuperscript{6} I find that the counterfactual environment with a fixed price of 99 cents has a higher percentage of failed high quality songs, lower expected consumer surplus, and higher expected revenue compared to demand-based pricing. This is a possible justification for the use of the fixed price in the big online music stores where consumer discovery of new artists may not be a high priority.

Finally, to validate the estimation, I examine the failure of high quality songs using an somewhat exogenous measure of song quality: Facebook fan page ‘likes’. This exercise demonstrates two

\textsuperscript{5}A caveat of this is that all the outcomes are measured without considering the fact that successful songs may continue to produce revenue going forward, meaning that the revenue on Amie Street will move further away from no information and closer to full information as time goes on.

\textsuperscript{6}In 2009 iTunes moved to a 3-tiered pricing system with prices of $0.69, $0.99 and $1.29.
things. The first is that it appears that there are high quality songs failing on Amie Street. Second, the model’s predicted number of failed high quality songs is close to the number when measuring quality by ‘likes’.

While this is a study of a specific small market, the results provide some insights on how bigger markets similar to this one may behave. The markets for apps and ebooks immediately come to mind, as well as offline markets such as restaurants and housing. In general, the results suggest that learning and the choice of pricing scheme in these types of environments can be an important determinant of market outcomes.

This paper contributes to several different strands of the literature. Classic theoretical observational learning papers (e.g., Banerjee (1992), Bikhchandani et al. (1992) and 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 et al. (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. In a real-world market, this may be equivalent to a high quality product not being ‘discovered’. I contribute to the above papers by analyzing these long-run effects of observational learning in an empirical setting.

There are many papers 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, while Chevalier and Mayzlin (2006), Sorensen (2006), and Luca (2011) analyze the impact of peer recommendations on the decisions of consumers. Empirical papers studying observational learning in product markets are sparse. In two examples, Cai et al. (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. In a project closely related to the current study, Salganik et al. (2006) show descriptive evidence of the long-run effect of social learning in an experimental music market. Finally, there are many other papers which examine different aspects of the music industry.\(^7\)

This paper’s primary contribution to the empirical literature is the fact that I estimate the effect of observational learning, rather than test whether or not it is occurring. I am able to do this because I estimate a long-run structural model and, therefore, can compare outcomes under observational learning to counterfactual outcomes under other informational environments. Additionally, I run counterfactual experiments which measure the effect of price in a market with observational learning. While Bose et al. (2006, 2008) examine this topic theoretically, I am not

aware of any empirical papers which do the same. Finally, I contribute to the understanding of social learning in online markets by using a new data set from a real-world online music store.

The remainder of the paper 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, while 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

In order 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.\(^8\) While 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 few, when they first post their songs.\(^9\) Another reason many of the artists may be unknown is the fact that any user can post his or her music without it being pre-screened by the website.

Amie Street Music hoped to become “the place to discover new music”\(^10\) by encouraging consumers to learn about new music on the site through features such as demand-based pricing. 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 cents, and then increases by 1 cent for each additional download until it reaches 98 cents (at 98 downloads). Once it reaches 98 cents, 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. One of the goals of this paper is to separately estimate these two effects.

Similar to other online music stores, Amie Street allows consumers to listen to a 1 minute and 30 second long sample of a song at no monetary cost. Unique to Amie Street, however, is the fact

---

\(^8\)The site was active from 2006 to September, 2010 when it was bought by Amazon.com, and subsequently shut down.

\(^9\)“indie” music refers to bands that 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. 36% of the artists in the sample don’t have a Facebook fan page and 58% of the artists who do have less than 1,000 “likes” on Facebook.

\(^10\)Shown on the front page of the Amie Street Web Site.
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’ \footnote{Buzz index is a measure of how much action a song has received recently.}. In Appendix A, I show three screen shots of Amie Street: (1) the home page, (2) song list filtered by genre (Folk/Country) and sorted by popularity, and (3) an album’s home page. On the second and third screen shot, 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 is the number of listens, price, and recommendations, in that order.\footnote{In this study, I ignore the effect of recommendations.} Note that songs listed as ‘FREE’ may be songs that are always given away at no cost, and therefore do not follow the demand-based pricing scheme. 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 the album name and arrives at the album page.

The simplicity and transparency of the features on Amie Street allow me to write down 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.

\section{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 et al. (2012).

\subsection{Consumer Preferences}

Consumer $i$’s utility for a song is given by\footnote{Because songs are assumed to be completely independent, I omit any song subscript.}: 

\[ w_i = X + u_i - p_i \]

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. 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 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 it assumed to be independent of the quality of the song. The normal distribution assumption means it is possible for consumers to have negative preferences for a song, 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\textsuperscript{14}.

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 $p_i$.

### 3.2 Arrival

Before outlining the decision making process, it is necessary to talk about the arrival process. I assume that a song belongs to one of two groups: ‘living’ songs or ‘dead’ songs\textsuperscript{15}. 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 of the map’ by reaching the bottom of one of filters or being removed from the site altogether. All songs which haven’t 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 their 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. Assuming a constant arrival for living songs implies that $i$ is also a measure of the time that has passed since the song has first been posted. Therefore, a consumer who knows how long a song has been on Amie Street knows her arrival order $i$.

### 3.3 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 price, and the aggregate number of listens. From the current price ($p$), the aggregate number of listens ($l$) and her arrival order

\textsuperscript{14}See your teenage daughter’s music collection for examples.

\textsuperscript{15}Not to be confused with songs by the incomparable Grateful Dead.
inferred from the release date \((i)\), 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:

\[ S_i = u_i + \epsilon_i \quad \epsilon \sim N(0, \gamma_\epsilon) \]

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 the act of the consumer observing the song name, the artist name, the album name, the album artwork, or any other information that may be visible. This signal provides no information about \(X\).

With \(\mu_i\) and \(S_i\) in hand, consumer \(i\) can calculate her expected benefit of listening:

\[
E[W | S_i, \mu_i] = Pr(w_i \geq 0 | S_i, \mu_i) E[w_i | w_i \geq 0, S_i, \mu_i]
\]

The first element of this expression is the probability that she will purchase the song after she listens, while 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:

\[
E[W | S_i, \mu_i] = \mu_i F\epsilon(H + S_i - p_i)(H + S_i - p_i - E[\epsilon | H + S_i - p_i \geq \epsilon_i]) \\
+ (1 - \mu_i) F\epsilon(L + S_i - p_i)(L + S_i - p_i - E[\epsilon | L + S_i - p_i \geq \epsilon_i])
\]

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 a great deal and there is evidence in the data that may indicate that most consumers listen before purchasing.\(^{16}\) The second

\(^{16}\)The average listen-to-buy ratio, defined as the number of listens over the number of buys for songs not equal to 98 cents, is 2.9 while the median is 1.5.
is that the consumer knows her place in line $i$. This is justified by the fact that the consumer can go to the album home page, learn the album release date, and infer $i$ (constant arrival). 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, $\mu_i$, using the process described in the next section.

### 3.4 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, while the second term is the probability consumer $i$ listens, given she observes $(l, p, i)$. The third term is the probably 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 $\mu(0, 0, 1) = \lambda$.

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

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

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

(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 $i' = \{1, ..., i - 1\}$ that could lead to this listen and purchase combination. Specially, given the initial values, the consumer iterates over the following updating equation:

$$\pi((l, p, i)|X) = \pi((l, p, i - 1)|X)(1 - \alpha(\mu(l, p, i - 1)))$$

$$+ \pi((l - 1, p, i - 1)|X)\alpha(\mu(l - 1, p, i - 1))(1 - \beta(X, \mu(l - 1, p, i - 1))) +$$

$$+ \pi((l - 1, p - 1, i - 1)|X)\alpha(\mu(l - 1, p - 1, i - 1))\beta(X, \mu(l - 1, p - 1, i - 1))$$

(4)

$^{17}$In Appendix B, I show how aggregation can be used to modify the model in order 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 decides to ignore the song. The second term is the probability \( i - 1 \) observed one fewer 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 in order to use this updating equation, the consumer must use equation (3) to calculate \( \mu'_{i'} \) 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.

For a concrete example of this process, assume consumer \( i = 21 \) arrives and observes \((15, 13, 21)\). She knows that 5 people didn’t listen to the song at all, 2 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 \( i' = 1 \) will listen and purchase 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 \( \mu(15, 13, 21) \).

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 1 of the consumers didn’t listen) versus this ‘sophisticated’ learning (what is the probability consumer \( i' = 2 \) didn’t listen) will have different implications.

### 3.5 Long-Run Implications

In this paper, 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.

#### 3.5.1 Long-Run Outcomes

I assume a song has reached its long-run outcome if it has either died somewhere below the 98 cent threshold or it has survived all the way to 98 cents. The former will be classified as a ‘failure’, while the latter will be classified as a ‘success’. While these seem to be loosely defined terms, the
data indicate that the 98 cent threshold is a very important landmark.\footnote{Songs that make it to 98 cents 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.} Therefore, the long-run outcome of a song is either the \((l, p)\) at which it died, or \(p = 98\).

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, \(\beta_d\), 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 cent 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’. I assume that once a song reaches 98 cents, it’s 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 cents are very different than songs that don’t, in terms of their listening rate.\footnote{For any easy example, if \(\beta\) 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 cent songs: the songs which die at higher prices have a significantly lower number of listens compared to songs which make it to 98 cents.}

Including the parameter \(\beta_d\) implies that this is not a standard herding model: it isn’t convergence of beliefs which leads to a song’s death.\footnote{I point out, however, that the traditional herding model is imbedded within the model described: if beliefs converge to a point where no consumers will ever listen again, that song will surely fail.} One of the reasons for this assumption is that it makes convergence to one of these outcomes both easy to define and trivial to prove. Another reason is that a model without this probability 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’.

An alternate assumption, which would be more in line with standard models, would be to define an upper bound on the private signal. However, making this assumption wouldn’t necessarily guarantee convergence.\footnote{Hendricks et al. (2012) show convergence in their model, but this modified version created to mirror the Amie Street market does not necessarily have the same properties.} That being said, I have explored the model with the above assumption and while the qualitative results do not differ a great deal, the model has a much more difficult time fitting the heterogeneity in outcomes in the data.

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.

### 3.5.2 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:
\[ \Pi_d(l, p < 98, I|H) \equiv \sum_{i=0}^{I} \beta_d(1 - \beta_d)^{i-1}(\pi(l, p, i|H)) \] (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 \(\beta_d\) 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 cents:

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

The important feature of this equation is the fact that the ‘longer’, or the more consumers, the song stays under 98 cents, 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, assuming it doesn’t die. The probability that this song dies along the way is:

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

as it would only be below 98 cents 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:

\[ \Pi_d(H) = \sum_{(l, p < 98)} \sum_{i=0}^{108} \beta_d(1 - \beta_d)^{i-1}(\pi(l, p, i|H)) + \sum_{i=109}^{\infty} 0 \] (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 cents 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 be unsuccessful. 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 some (if not all) consumers to ignore a high quality song because of the risk of wasting time on a low quality song, which increases the probability of death. The primary goal of this paper is to measure how well the Amie Street model of observational learning performs in dealing with this uncertainty.
4  Data

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 4 month period from March to July, 2010. In March 2010, I used Amie Streets ‘newest added’ filter to begin scraping price and listen information for the last 1,000 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 1,000 artists and a ‘complete’ history for the songs of the 300 artists. This is important because some of the songs from the 1,000 artists had reached their long run outcome before I started scraping the information, meaning I don’t observe their transition to these outcomes. Also, because many of these songs reach their long run outcome within a week of being posted, and the data is 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.

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 the folk/country genre. This is an attempt to keep the heterogeneity in the songs somewhat limited. For restriction 2, I start by eliminating 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 only keep 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. In order to eliminate false release dates, I also drop any artist who has multiple albums released on the same date. The thought is that an artist who has been around a while may have just dumped all their music on Amie Street and put the date they posted it as the release date. Finally, each album usually contains multiple songs, so it may be important to take into account the fact that consumers could learn about an 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 song in order to learn their preferences for that band and/or album, so I include only the first track listed on a given album. Including only the first track eliminates the need to model the learning that occurs between single songs on a given album.

\[22\] The first track listed on an album in the entire data set has on average 198 listens, the second has 133, the third has 126, the fourth has 105, and so on.
album or by a given artist. Therefore, while the data are technically at the song-level, it can also be thought of as artist-level data.

For restriction 3, the data are limited to songs which have reached one of two states by the end of the sample period: (1) it has reached the maximum price of 98 cents or (2) it has not reached 98 cents and it 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), which assumes that inactivity for 2 months implies failure. The sample size totals 276 songs from 276 artists.

The data are separated into two categories: (1) the long-run joint distribution of listens and price for songs which have died below 98 cents and (2) the number of songs which reached 98 cents. Note that I do not use information about the number of listens for 98 cent 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 7% of the songs make it to the 98 cent threshold, even though it only takes 98 downloads to get there. Also, many of the songs fail at very low prices (below 30 cents), while there is a scattering of songs which reach a higher price before failure. This provides some evidence that learning is occurring, as most songs which make it above 30 cents will make it to 98 cents.

Finally, notice that there are some large spikes in the data at 13 and 15 cents. The 13 cent spike is likely due to the fact that songs are free until 13 people have downloaded them, while the 15 cent spike is difficult to explain. It may be possible that songs which are 14 cents are somehow displayed differently on the website, leading to one more purchase, but I have not been able to confirm this. Overall, the price distribution shows the binary nature of the data: songs likely either die at low prices or make it 98 cents.

To further investigate the data, Table 1 shows the average and standard deviation for the number of listens for different subsets of the data. The first row displays the distribution of listens for all the songs which failed below 98 cents, while the second and third row split the data into songs that failed above and below 30 cents. For all failed songs, it takes on average about 58 listens before a song dies. This number is 45 for songs which die under 30 cents and 171 for songs which die above 30 cents. 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. This is evidence that consumers believe songs which reach a higher price must be high quality, and continue to listen to them despite the fact that many other consumers listened and didn’t purchase. A large amount

---

23There are a number of songs that did not reach the point where they were ever purchased for a positive price (i.e., their price is 0 for the entire data set) This could be because the song was not purchased more than 12 times, or it could be that the artist has decided to give the song away for free in order to promote their music. If it is the latter, then the price is not an accurate measure of the number of downloads. I therefore drop any song that never reached a positive price. This will become an issue when matching the model to the data, which I deal with by only using the likelihood of songs that reach at least 13 downloads.

24I have used less restrictive data sets, producing similar results.

25This leads to the question of whether or not artists will try to game the system. In equilibrium only high quality songs would try to do this, which would imply that the learning model is still valid.
of heterogeneity in tastes is one possible explanation for this behavior. Table 1 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. For a more rigorous examination of the data and learning in this market, see Newberry (2013).

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, \( \gamma_u \) is the standard deviation of preferences, \( \gamma_\epsilon \) is the standard deviation of the private signal, \( c \) is the cost of listening, \( \lambda \) is the prior, and \( \beta_d \) is the death rate. Below I describe the procedure to estimate \( \theta \).

5.1 Procedure

Given a vector of parameters, \( \tilde{\theta} \), I calculate the probability that a song has died at \((l, p, i)\), \( \Pi_d(l, p, i|\tilde{\theta}) \), and the probability a song has made it to 98 cents, \( \Pi_{98}(i|\tilde{\theta}) \), 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 \( p < 98 \) 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 cents or made it to 98 cents. This is what I define as convergence of the model. Convergence will occur at the \( i \) such that:

\[ \sum_{(l, p < 98)} \Pi_d(l, p, i|\tilde{\theta}) + \Pi_{98}(i|\tilde{\theta}) = 1 \]

\[ \text{See Appendix B for additional details about the estimation procedure.} \]
An equivalent representation of convergence is the $i$ such that there are no longer any songs below 98 cents that are living, or:

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

With this representation, it is straightforward to show that the model surely converges as $i$ approaches $\infty$. 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 $I$.\textsuperscript{27} Once the model converges, I form the log-likelihood function:

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

where $(l^*, p^* < 98)$ and $K$ are data. $K$ is the number of songs that reached 98 cents. Recall that the data are the joint distribution of $(l, p)$ for songs which died under 98 cents and the number of songs that made it to 98 cents. I do not use any information about the listens for 98 cents songs because the data indicate that these songs are still alive, and in order 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 cents songs would likely be most helpful in identifying $\beta_{98}$, which I am not estimating.

Because the likelihood function is not smooth and has many local minima, I use the ARS (accelerated random search) algorithm in Appel et al. (2003) to find starting points to plug into the maximization routine in Python.

### 5.2 Identification

The discussion of identification is informal. Intuitively, identification of the parameters comes from the fact that each parameter changes the consumers’ decision process in different ways. 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 at a given belief. Changing the decision process of consumers changes the probability of certain paths, which will ultimately change the probability of observing a song die at a given $(l, p)$. While this tells the general story of identification, I move on to a parameter-by-parameter discussion. I start by stating the following hypotheses about how the change of a parameter will effect the long-run joint

\textsuperscript{27}For the results presented, I set $\varepsilon$ to 0.01. I have also experimented with other values between 0.001 and 0.01, with little change in the results.
distribution of listens and price for $H$ songs and $L$ songs separately:

1. $\uparrow H$: a high quality song is now less likely to die. Also, any high quality song that died at $p$, is now more likely to make it to a higher $p$. Result: a higher mean price at which high quality songs die.

2. $\uparrow L$: a low quality song which received the ‘average’ draw of consumers and died at $p$, is now more likely to make it to a higher $p$. Result: a higher mean price at which low quality songs die.

3. $\uparrow \gamma_u$: a low quality song which received the ‘best’ draw of consumers and died at $p$ is now more likely to make it to a higher $p$ (i.e., the ‘best’ draw is now better). Result: a larger spread of prices at which low quality songs die.

4. $\uparrow \lambda$: there are now more high quality songs in the market than before. Result: mass will switch from the distribution of low quality songs to high quality songs.

5. $\uparrow c$: consumers are now less likely to listen to any song at every belief. Any song that died at $(l, p)$ is now more likely to die at $(l - 1, p)$. Result: a decrease in the location of the conditional listen distribution for all songs.

6. $\uparrow \gamma_c$: consumers are now more likely to listen to songs which they aren’t actually going to buy and ignore songs they would buy. Any song that died at $(l, p)$ is now more likely to die at $(l + 1, p)$ and $(l - 1, p)$. Result: the spread of the conditional listen distribution will increase.

7. $\uparrow \beta_d$: more songs of both quality are going to die. Result: increase the total amount of songs which die.

What the above imply is that if I knew exactly which songs are high quality and which are low quality, then the two separate distributions should be able to identify the parameters. While I do not have an exogenous measure of quality, a result of the model is that high quality songs are more likely to reach higher prices before they die. That is, any song that dies at $(l, p)$ has a higher probability of being $H$ than a song that dies at $(l, p - 1)$. This is simply because people are both more likely to listen to high quality songs as learning takes place and more likely to purchase them given they listen. For an example, I present Figure 3, which is the estimated joint distribution of price and listens for all songs (light) and the joint distribution for high quality songs (dark). It is clear a song which died at $p = 80$ is more likely to be a high quality song than one which died at $p = 20$. Given this fact, I argue that the parameters that mostly affect the high quality songs ($H$) should be identified by the data at the high end of the price distribution, while the parameters that affect the low quality songs ($L, \gamma_u$) should be identified by the data at the low end. Parameters that affect the relationship between the two distributions ($\lambda$) should be identified by the relationship
between the data at the high end of the price distribution and the data at the low end. Finally, parameters that affect both high quality and low quality songs more or less equally ($\beta_d$, $c$, $\gamma\epsilon$) should be identified by the overall distribution of the data.

There are two additional points about identification that deserve discussion. First, by assuming an exogenous death rate instead of the standard assumption of an upper bound on $S_i$, I believe I am able to identify all the parameters, including $\beta_d$. Identification of an upper bound on the private signal, $\tilde{S}$, on the other hand, is not so clear. The primary reason is that $c$ and $\tilde{S}$ have similar effects: an increase in $\tilde{S}$ and a decrease in $c$ both lead to a lower belief needed in order for a song to die. Second, while most empirical learning papers use individual consumer level data to identify the parameters, I argue above that I am able to identify the model using only aggregate data. While the transition data may be helpful for more precise identification, in many cases this type of data may not be available. In this paper, I show that it is possible to estimate a learning model with only long-run, aggregate data.

6 Results

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

6.1 Estimates

The estimates are found in Table 3 with standard errors in parentheses. The average value of a high quality song is $2.26, while the average value of a low quality song is $\sim$0.48. The negative value for a low quality song is interesting, but again could be because of space on the computer, the time it takes to download, or that it is just not very nice to listen to. Recall that any user can post a song on Amie Street, meaning there could be some very low quality music for sale on the site. I imagine most consumers would not download a song by the author, even if they were paid $0.50. The heterogeneity of consumer preferences is $0.85, meaning 29% of consumers would download a low quality song if it were free and about 7% of consumers would not purchase a high quality song at the maximum price of 98 cents. The fact that there is not much overlap in values for the high quality and low quality songs implies that learning may happen very quickly: if a consumer does not download a free song, it is almost surely low quality.

The estimates of $\lambda$ and $\beta_d$ are 0.39 and 0.009, respectively, indicating that about 40% of the music on Amie Street is high quality and each song has less than a 1% chance of falling of the map for each consumer arrival. The cost of listening is quite low, $0.01, but this could be coming from the fact that I am assuming consumers must listen before they purchase. Relaxing this assumption would increase $c$ because there will be some chance that an observed purchase is completed without an associated listen. Finally, $\gamma\epsilon$’s estimate is large in magnitude, implying that the private signal is
not very informative. Standard errors indicate that all of the parameters are estimated precisely.

I compare how well the model fits the data with Figure 4, which 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 98 cents songs, but this is presumably due to the inability to pick up the large jumps at the lower end of the distribution. Considering I am assuming a model of only two qualities, I find the fit of the price distribution impressive.

Another area where the model does not do very well is predicting the heterogeneity in the listens, and hence, the heterogeneity in the joint distribution. Table 4 displays the estimated mean and standard deviation of the listens with the same measures in the data in parentheses. The model overestimates the mean of the listen distribution for songs which failed under 30 cents and underestimates the mean for songs which fail above. The reason for this is that the model is trying to fit both of these moments, but can only end up somewhere in the middle with a relatively low estimate of $c$. One way to rectify this would be to make the consumer arrival process endogenous, but with the available data, I lack anything that would identify arrival. The model does a reasonably good job of matching the spread of the listen distribution, although it is underestimated when conditioning on price. The reason is that it is hard for the model to predict some songs dying with very few listens and some dying with many listens.

6.2 The Effects of Observational Learning

In order to assess the impact of observational learning, I compute three long-run outcomes and compare them across three different informational environments. The three outcomes I measure are: (1) the probability a high quality song fails (2) the expected consumer surplus and (3) the expected revenue. The first row of Table 6 displays the results for the estimated model. The expected consumer surplus and revenue are calculated assuming $\beta_d = \beta_{98}$, meaning the numbers presented are essentially lower bounds. Results indicate that 63.7% of high quality songs fail on Amie Street (AS), or about 67 of the folk/country songs in the data. The expected consumer surplus per song is $76.91$, or about about 13 cents per consumer, and the expected revenue is $24.45$ per song, or $6,602$ for the entire market.

I calculate the same outcomes for two counterfactual environments: one in which consumers have perfect information (PI) and one in which consumers have no information (NI). In the perfect information environment, consumers know $X$, but they still do not know $u_i$, meaning some uncertainty remains. In the no information environment, consumers only know the prior, or that there is a 0.39 probability that the song is high quality. While the 63.7% number gives me an overall measure of unsuccessful high quality songs, it does not actually tell me how many songs were unsuccessful due to the uncertainty that exists in the market. The comparison between Amie Street and the perfect information market allows me to do that. At the same time, the comparison
between Amie Street and no information market allows me to estimate how well the observational learning does in reducing the cost of uncertainty, or how much value the learning provides the market. The ability to perform this counterfactual analysis is an additional reason why the assumption of a death rate is preferred to an upper bound on the private signal. In the environments where there is no learning occurring, the traditional herding model would never converge. In the death rate model, however, even environments where beliefs aren’t changing eventually converge, allowing me to compare outcomes across learning and non-learning models.

The second and third rows of Table 6 show the percentage differences in outcomes between the given counterfactual environment and Amie Street. Row (2) indicates that the AS learning environment performs well in two dimensions, as there is less than a 1% difference between AS and PI in the first two outcomes. However, the expected revenue for a song is about 5% lower with the AS model. The simple explanation for this is that a slight increase in the number of failed high quality songs leads to a comparatively large reduction in revenues. This is due to the fact that a high quality song which does not die produces at least $48 in revenue.

Row (3) of Table 6 shows that removing any ability for consumers to learn results in an increase in failed high quality songs of around 4%, a reduction of consumer surplus and a small reduction in revenue. Note that the assumption of a constant death rate after 98 cents implies that these differences are lower bounds, as the successful songs will generate more revenue and surplus if $\beta_d > \beta_{98}$. Overall, these results suggest that learning on Amie Street helps high quality products get discovered and is beneficial to consumers, but does little to enhance revenue for the website. Therefore, in this case, observational learning benefits producers of high quality products and consumers, but not necessarily the firm in charge of the market. Finally, the fact that the failure rate of high quality songs is similar to PI means that the learning on Amie Street is very efficient.

The efficiency of the learning in this market is directly tied to the speed at which consumers learn the true value of a song: the more consumers it takes to learn the true quality, or the longer uncertainty exists, the more likely it is that a high quality song will die. Figure 5 shows that the market learns the true quality of a song by the 25th consumer. The x-axis is the arrival order of consumers while the y-axis is the expected surplus for the corresponding consumer. The three different curves represent the different informational environments discussed above (AS,PI,NI). Complete learning is achieved when the AS curve is the same shape as the PI curve, which happens at approximately consumer 22. Theses results imply that any failure of high quality songs that can be attributed to uncertainty happens before the arrival of the 22nd consumer.

The expected surplus curves for PI and AS are not equal under complete learning because of the demand-based pricing scheme: consumers buy at a higher rate under PI, meaning later consumers will observe a higher expected price. This implies that it actually pays to be one of the later consumers in AS: the market has learned the true $X$ with a lower expected price. In fact, there

---

28This is the approximate revenue for a song which is purchased 98 times.
are points when the NI environment dominates the PI environment for this same reason.

In order to make the intuition of the graph more clear, I remove any effects of the demand-based pricing scheme by fixing the price at 99 cents. Under this pricing scheme complete learning is achieved when the expected surplus under AS exactly equals the expect surplus under PI. Figure 6 shows that under AS learning and a fixed price of 99 cents, this occurs at about consumer 40. I discuss this counterfactual in more detail below.

6.3 Fixed Price

In markets with observational learning, a change in the price will not only affect an individual’s utility, it will also impact the propensity to experiment with new music. When consumers are experimenting less, there exists more uncertainty in the market, and individuals will be more sensitive to prices than they would be in the full information environment. Therefore, lower prices will lead to more efficient learning and an increase in demand for high quality products. However, a firm must balance this with the fact that this would also reduce the revenue earned from each purchase. The demand-based pricing (DBP) scheme is a clever way to try to accomplish both efficient learning and also earn revenue into the future. In this section, I run counterfactual experiments to assess the impact of this pricing scheme. In general, this exercise is meant to be a first look at how pricing impacts demand in markets with social learning.

The demand-based pricing scheme has two effects on the consumer’s decision making process. The first, which I call the ‘experimental effect’, encourages consumers to listen to new music because these songs are cheaper, while the second, which I call the ‘learning effect’, allows consumers to learn about the quality of a song through the price. These two forces are intertwined, as the more experimenting consumers do, the more information will be transmitted, and the faster the market will learn. I try to separate the learning effect from the experimental effect by fixing the price of a song to 99 cents, but keeping the learning environment the same. The choice of 99 cents is due to the fact that this was the traditional pricing scheme of the industry during the early 2000s. This counterfactual examines the question of how the market may have looked had Amie Street adopted the pricing scheme popular to the industry. It also serves as an analysis of how pricing and learning are related, and how a firm may be able to use different pricing schemes to encourage experimentation.

Row four of Table 6 shows the results if Amie Street employed a fixed price but kept the same learning environment. The effect on the success rate of high quality songs is similar to the effect of removing the ability to learn (5% higher rate of failure compared to AS), implying that fixing the price actually hampers learning a great deal. The consumer surplus falls by 29% because the consumers who are purchasing the song now have to pay 99 cents for it. This also increases expected revenue by about 66%. Again, the speed of learning plays a big role in these outcomes. Figure

---

\[29\] This is an upper bound due to the assumption of the death rate. However, because only 5% fewer songs make
6, shows that the market doesn’t learn the true quality under the fixed pricing scheme until the 40th consumer, which is about 15 more consumers than it takes under the demand-based pricing scheme.

The same pattern emerges when using a fixed price in the full and no information environments. While not a perfect comparison, I believe that the bigger music shops like iTunes and Amazon might fall somewhere in-between AS (row 4) and NI (row 6). This is because there are some social learning signals on these sites (e.g., ratings on iTunes and recommendations on Amazon) and the price is fixed at 99 cents. If this assumption is true, one would conclude that these environments are beneficial to the music store (higher revenue), but hurt the high quality artists (higher probably of failure) and the consumer (lower consumer welfare). This, perhaps, is a justification why the music stores with a higher market share may decide to price this way. Why would Amie Street choose to price the way it did? While I cannot answer this question directly, the techniques they employ lead to speculation that they are a small player hoping to gain a loyal consumer base and attract unknown, but emerging high quality artists.

### 6.4 Increased Search Cost

A fixed pricing scheme is only one of the things that may slow down learning. In this section, I discuss the results of a counterfactual in which the cost of listening (or the search cost) is increased.

Suppose that Amie Street charged consumers to listen to the sample of the song, or removed the ability to listen to the sample all together, increasing the cost a consumer must pay to learn $X + u_i$. While in my model this is $c$, or the cost of listening, in general we can think of it as a search cost. In the following counterfactual, I analyze how increasing the search cost may affect the market by slowing down the learning process. A high cost of listening discourages some consumers from experimenting with a song, limiting the amount of information that is passed to the market. This slows down the learning process and increases the chance of failure for high quality songs. Figure 7 shows that the learning process is greatly affected by increasing the cost of listening to just 10 cents. This figure displays the expected surplus for consumers under the fixed pricing scheme and the increased value of $c$. Because the fixed pricing scheme makes the effect of learning much more obvious in the graph, the comparison is to Figure 6.

Results indicate that learning doesn’t even begin to occur until about the 40th consumer. Recall, that purchases aren’t observable until 13 people have bought a particular song. With the high cost of listening, many people will ignore the song, and the 13th purchase won’t occur until at least 40 consumers have arrived. The figure shows that the market won’t learn the true quality of the song until about the 150th consumer. This exercise demonstrates that policies which discourage experimentation will hamper learning. Depending on their goals, this result implies that Amie Street may want to try to make learning about a particular artist or song as easy as possible by it to 98 cents, I believe this number would still be significantly large under a different value for $\beta_{98}$. 

---

22
lowering the frictions to the learning process.

6.5 High Quality Songs Failing

In order to take the results seriously, we must believe that the model does a good job of matching the true environment on Amie Street. In this section, 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 where 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. In order for this to be valid, I assume that the number of FB likes collected in November of 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 7 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 cents 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. While 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 67.

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 cents (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 (62). In the second approach, I find the FB likes of the 67th ‘best’ failed song (588), and assume this is the lower bound of FB likes for high quality artists. I then compare this to the FB likes of the 98 cents songs. Again, the lowest number of FB likes for a 98 cents song is 632. I believe these results, although not very scientific, show some evidence of the predicting power of the model. At the very least, this exercise shows that there are likely some high quality songs which failed on Amie Street. Table 8 lists some possible examples of these artists.

7 Conclusion

In this paper, 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 music is a market where songs quickly either succeed or fail and because of the transparency of the learning environment. I found that the observational learning on Amie Street Music is valuable to
producers of high quality songs and consumers, but not very valuable for the website itself. I also explored how different pricing schemes may hamper the learning process, and found that the scheme traditional used by the industry may lead to worse outcomes for the consumer, but better outcomes for the firm. While this paper studied one market in particular, I believe the results can be applied to online markets with similar learning features. In general, the results provide evidence that (1) observational learning can be a valuable tool for producers who sell their goods in an online market and consumers who purchase goods in an online market and (2) the pricing scheme can have a significant impact on the learning process. Possible future work on this issue includes studying the long-run impact of other forms of learning which may be prevalent in bigger markets, such as word of mouth.

References


A Screen Shots

Screenshot 1: Home Page

Get great deals on new music at Amie Street:

- Quality music over popular music
- We help you discover up-and-coming artists and re-discover classic albums
- Find great deals and free songs every day
- Most songs start FREE and rise in price based on popularity
- An active community of music fans

Join Now for Free and get 75 FREE TRACKS!

Today's Album Overview

Amie's Top 4
Best-selling albums of the week

- Let It Sway
  by Someone Still Loves You Boris Yeltsin
  Released: Aug 17, 2010
  12 Tracks

- Modern Rituals
  by Chief
  $7.49

- The Budos Band III
  by The Budos Band
  $6.98

- The Week Never Starts Round Here
  by Aesop Rock
  $9.10

Community Buzz
Albums people are talking about right now

- El Sur de mis zapatos
  by Javier Carretero
  Released: Nov 30, 1999
  15 Tracks

- From The Barrio With Hate
  by Darkroom Farmla
  $4.21

- The Best Of Grammercy Gospel
  by Various Gospel Artists
  $2.99

- Out Of Darkness
  by In The Gods Of Lom
  $3.10

See more REC's...

New Releases
Most recently added albums

- till I ever
  by Prose In Rosette
  $4.29

- LOVES WILL FIND THE WAY
  by Abad Guitera (Songwriter)
  FREE

- electric x-t-t
  by unlevels
  FREE

See all new releases...
Screenshot 2: Folk/Country Filtered by Popularity

Folk / Country Genre
Check out all the Folk / Country music Amie Street has to offer.

Price: Any  |  Sort By: Best Selling  |  Show: 15  |  Page: 1 of 11380

- **Atman** by Rodrigo y Gabriela
  on Atman (Singla)  
  1,299 plays  
  **FREE**  
  $0.98

- **Can’t Go Back Now** by The Weepies
  on Hillsaway  
  4,499 plays  
  **FREE**

- **L.A. County Blues** by The Band of Heathens
  on One Foot In The Ether  
  1,181 plays  
  **FREE**

- **Dreams Come True Girl** by Cass McCombs
  on Cest La Vie  
  1,481 plays  
  **FREE**

- **Early November** by Miranda Lee Richards
  on Early November  
  2,175 plays  
  **$0.98**

- **Bird** by Susan Enan
  on Plainsong  
  1,269 plays  
  **$0.98**

- **San Francisco** by Brett Dennen
  on Hope For The Hopeless  
  4,671 plays  
  **$0.98**

- **Who Will** by Will Stratton
  on No Wonder  
  534 plays  
  **FREE**

- **Lovers Prayers** by Ida
  on Lovers Prayers  
  4,179 plays  
  **FREE**

- **A Long Dream** by Tyler Ramsey
  on A Long Dream About Swimming Across The Sea  
  2,216 plays  
  **FREE**

- **Charlie Darwin** by The Low Anthem
  on Oh My God, Charlie Darwin  
  1,474 plays  
  **FREE**

- **How Long** by Eric Davich
  on How Long / Moonlight In The Morning (Demos)  
  516 plays  
  **FREE**

- **Monoplain** by Susan Enan
  on The Acoustic Sessions  
  3,847 plays  
  **FREE**

- **First Person** by Jenny Owen Youngs
  on Transmitter Failure  
  1,899 plays  
  **FREE**

- **I Heard Your Voice In Dresden** by Elvis Perkins in Dearland
  on Elvis Perkins In Dearland  
  922 plays  
  **FREE**

---

Folk / Country Styles

- All Folk / Country
- Acoustic
- Americana
- Bluegrass
- Country
- Folk
- Folk Rock
- Rockabilly
- Roots
- Zydeco

Gens

- All
- Pop
- Rock
- Electronic
- Hip Hop / Rap
- Soul / R&B
- International
- Jazz / Blues
- Classical
- Miscellaneous
B Estimation

In this section I discuss a few more details of the estimation procedure. For any reader who is interested, I used Python for the web scraping and a combination of Python along with Fortran for estimation.

B.1 Calculating Cutoff $S$

The first issue is the calculation of the cutoff level of the private signal $\bar{S}(\mu_i)$. With the parametric assumptions, there is no closed for solution for this. Therefore, I use the bisection method to calculate the $S_i$ that solves:

$$0 = \mu_i F_\epsilon(H + S_i - p_i)(H + S_i - p_i - E[\epsilon|H + S_i - p_i \geq \epsilon_i])$$
$$+ (1 - \mu_i) F_\epsilon(L + S_i - p_i)(L + S_i - p_i - E[\epsilon|L + S_i - p_i \geq \epsilon_i]) - c$$  (9)
I calculate this for every possible \( (l, p, i) \) combination.

### B.2 Calculating the Expected Value of Listening

The second issue is the fact that 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 1,000 values for \( u_i \) and \( \epsilon_i \) and calculate the percentage of consumers who would purchase, given they listen.

### B.3 0 Probabilities

The third issue is the fact that the computer will predict 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 \( \Pi_d(l, p, I|\theta) \) or \( \Pi_{98}(I|\theta) \) 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 \( 1 \times 10^{-150} \) to each likelihood the model predicts, and then recalculate probabilities so they sum to one.

### B.4 Convergence and \( \beta_d \)

Next, the length of time before convergence of the model is heavily dependent on the value of \( \beta_d \). As the value of this parameter falls, the time it takes to compute the objective function increases substantially. I therefore set a lower bound on the value of \( \beta_d \) when I run the ARS algorithm to search for starting points. In most iterations I use 0.008 as the lower bound because this implies that a song will surely have less than a 1% chance of being alive by the time the 600th consumer arrives. 600 seemed like a reasonable estimate of the maximum number of consumers who arrived at a dead song considering 400 is the maximum amount of listens for any dead song in the data.

### B.5 Learning under DBP

Finally, I presented the updating procedure in the body of the paper 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 number 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.

C Figures and Tables

C.1 Figures

Figure 1: Purchase (Price) Distribution

Notes: The figure displays the distribution of long-run prices for the 276 folk songs in the data.
Figure 2: Listens for 98 and Non-98 Cent Songs

Notes: The figure displays the average number of listens for songs depending on how long they have been on Amie Street and whether or not they are a 98 cent song.

Figure 3: Estimated Purchase (Price) Distribution Separated by Quality

Notes: The figure displays the estimated long-run price distribution for all songs and $H$ songs.
Figure 4: Purchase (Price) Distribution Fit

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

Figure 5: Evolution of Consumer Utility (DBP)

Notes: AS refers to the Amie Street learning environment, while PI is perfect information and NI is no information. The figure displays how the expected utility of consumers changes over time (consumers) under the three different environments and demand-based pricing (DPB). AS and NI dominate PI at times because under DBP prices will increase faster under PI compared to the other environments. This means later consumers will face a higher expected price. Complete learning occurs when the AS curve is the same shape as the PI curve.
Figure 6: Evolution of Consumer Utility (p=99 cents)

Notes: AS refers to the Amie Street learning environment, while PI is perfect information and NI is no information. The figure displays how the expected utility of consumers changes over time (consumers) under the three different environments and a fixed price of 99 cents. Complete learning occurs when the AS curve and the PI curve meet.

Figure 7: Evolution of Consumer Utility (c=10 cents)

Notes: AS refers to the Amie Street learning environment, while PI is perfect information and NI is no information. The figure displays how the expected utility of consumers changes over time (consumers) under the three different environments, a fixed price of 99 cents, and a listening cost of 10 cents. Complete learning occurs when the AS curve and the PI curve meet.
### Table 1: Listen Distribution for Unsuccessful Songs

<table>
<thead>
<tr>
<th>Buys (Price)</th>
<th>Mean Listens</th>
<th>St. Dev. Listens</th>
</tr>
</thead>
<tbody>
<tr>
<td>[13, 97]</td>
<td>58</td>
<td>73</td>
</tr>
<tr>
<td>[13, 30]</td>
<td>45</td>
<td>48</td>
</tr>
<tr>
<td>[30, 97]</td>
<td>171</td>
<td>133</td>
</tr>
</tbody>
</table>

### Table 2: Estimates

<table>
<thead>
<tr>
<th>$H$</th>
<th>$L$</th>
<th>$\gamma_u$</th>
<th>$\gamma_c$</th>
<th>$c$</th>
<th>$\lambda$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.26</td>
<td>0.48</td>
<td>0.85</td>
<td>2.75</td>
<td>0.01</td>
<td>0.39</td>
<td>0.009</td>
</tr>
<tr>
<td>(0.04)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.07)</td>
<td>(0.00)</td>
<td>(0.03)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

Note: Standard errors calculated numerically in parentheses.

### Table 3: Model Fit

#### Listen Distribution for Unsuccessful Songs

<table>
<thead>
<tr>
<th>Buys (Price)</th>
<th>Mean Listens</th>
<th>St. Dev. Listens</th>
</tr>
</thead>
<tbody>
<tr>
<td>[13, 97]</td>
<td>100 (58)</td>
<td>69 (73)</td>
</tr>
<tr>
<td>[13, 30]</td>
<td>77 (45)</td>
<td>38 (48)</td>
</tr>
<tr>
<td>[30, 97]</td>
<td>125 (171)</td>
<td>84 (133)</td>
</tr>
</tbody>
</table>

Note: Estimates displayed, data in parentheses.
Table 4: Outcomes Compared to Amie Street

<table>
<thead>
<tr>
<th>Row</th>
<th>Mechanism</th>
<th>Environment</th>
<th>Probability of Failure for H Songs</th>
<th>Expected Surplus per Consumer</th>
<th>Expected Revenue per Song</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>DBP</td>
<td>AS</td>
<td>63.7%</td>
<td>$0.13</td>
<td>$24.45</td>
</tr>
<tr>
<td>(2)</td>
<td>DBP</td>
<td>PI</td>
<td>-0.5%</td>
<td>0.0%</td>
<td>4.5%</td>
</tr>
<tr>
<td>(3)</td>
<td>DBP</td>
<td>NI</td>
<td>-1.6%</td>
<td>-0.4%</td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>Fixed 99</td>
<td>AS</td>
<td>4.7%</td>
<td>-29.0%</td>
<td>66.4%</td>
</tr>
<tr>
<td>(5)</td>
<td>Fixed 99</td>
<td>PI</td>
<td>3.6%</td>
<td>-28.5%</td>
<td>67.5%</td>
</tr>
<tr>
<td>(6)</td>
<td>Fixed 99</td>
<td>NI</td>
<td>10.2%</td>
<td>-31.5%</td>
<td>60.7%</td>
</tr>
</tbody>
</table>

Notes: AS refers to the Amie Street learning environment, while PI is perfect information, NI is no information and DBP is demand-based pricing. The table displays the estimated outcomes for the Amie Street learning and pricing environment in the first row and then the percentage difference between these results and counterfactual outcomes in the remaining rows.

Table 5: Facebook Fan Page Likes by Purchases

<table>
<thead>
<tr>
<th>Buys (Price)</th>
<th>FB Likes Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>[13, ∞)</td>
<td>91</td>
</tr>
<tr>
<td>[98, ∞)</td>
<td>9,221</td>
</tr>
<tr>
<td>[13, 97]</td>
<td>53</td>
</tr>
<tr>
<td>[13, 29]</td>
<td>27</td>
</tr>
<tr>
<td>[30, 97]</td>
<td>2,014</td>
</tr>
</tbody>
</table>

Notes: I set FB likes to 0 if the artist did not have a Facebook fan page which was approximately 36% of them. The table displays the median FB likes for different price ranges of folk songs on Amie Street.

Table 6: Failed High Quality Artists?

<table>
<thead>
<tr>
<th>Artist</th>
<th>Buys</th>
<th>Listens</th>
<th>FB Likes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shooter Jennings</td>
<td>15</td>
<td>17</td>
<td>57,860</td>
</tr>
<tr>
<td>Jesse Cook</td>
<td>32</td>
<td>40</td>
<td>17,550</td>
</tr>
<tr>
<td>Hope Sandoval</td>
<td>48</td>
<td>62</td>
<td>94,381</td>
</tr>
<tr>
<td>Rodrigo y Gabriela</td>
<td>55</td>
<td>75</td>
<td>288,542</td>
</tr>
<tr>
<td>Kate Voegel</td>
<td>69</td>
<td>389</td>
<td>277,820</td>
</tr>
<tr>
<td>Sonia Leigh</td>
<td>24</td>
<td>24</td>
<td>22,863</td>
</tr>
</tbody>
</table>

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