Effects of Online Recommendations on Consumers' Willingness to Pay

Appendix 1: Screenshots

Studies 1 & 2 Only: multiple songs per page

Step 3. Song Pricing

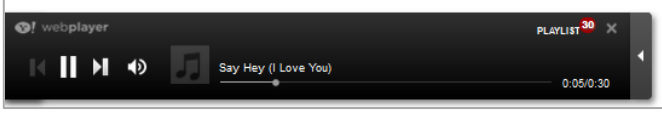
Group 1: Songs with Predicted Ratings

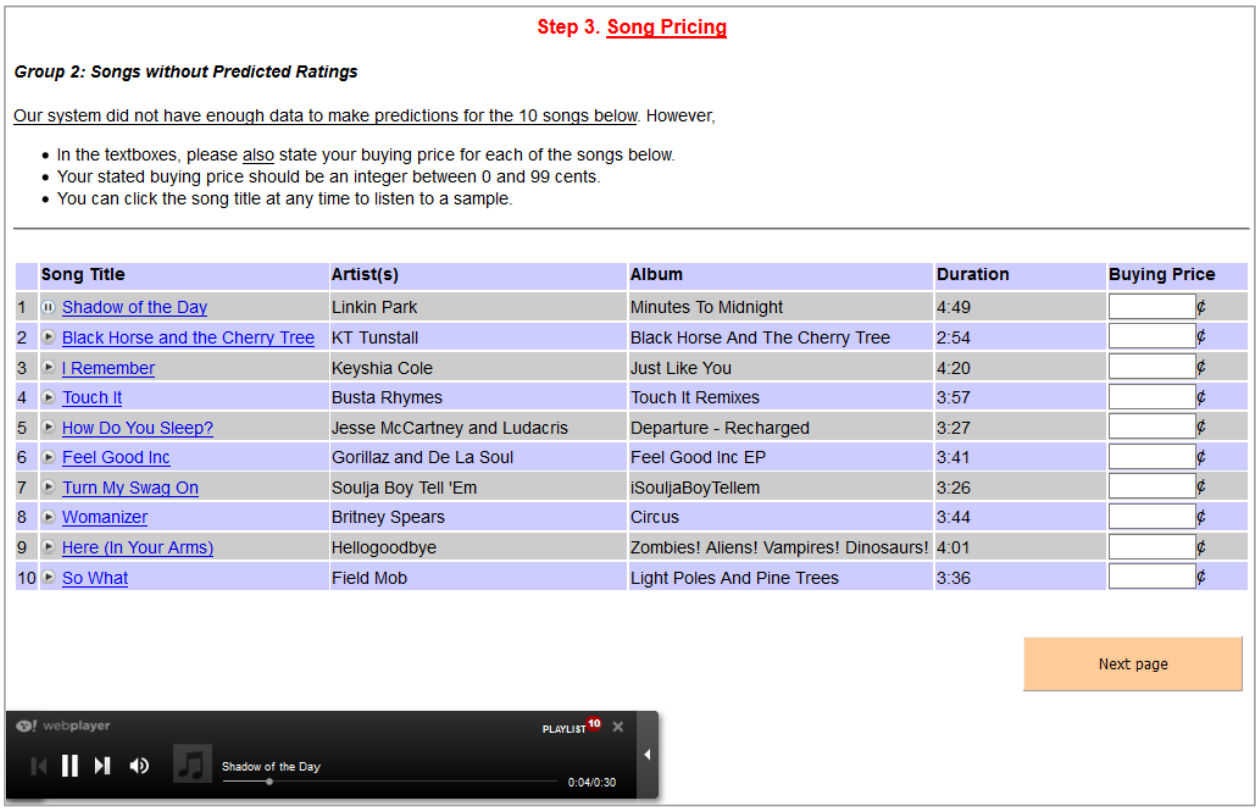
Based on the ratings you provided, our recommender system has predicted how much you would like the songs below. Our system uses several different recommendation techniques for making predictions. Each individual prediction was made by one of these techniques.

- In the textboxes, please state your buying price for each of the songs below.
- Your stated buying price should be an integer between 0 and 99 cents.
- You can click the song title at any time to listen to a sample.

	Song Title	Artist(s)	Album	Duration	Predicted Rating	Buying Price
1	Say Hey (I Love You)	Michael Franti, Spearhead and Cherine Anderson	All Rebel Rockers	3:56	4.8	¢
2	You're Gonna Miss This	Trace Adkins	American Man, Greatest Hits Volume II	3:45	4.2	¢
3	Rehab	Amy Winehouse	Back To Black [Explicit]	3:33	4.2	¢
4	Big Green Tractor	Jason Aldean	Wide Open	3:24	1.8	¢
5	Show Stopper	Danity Kane and Yung Joc	Danity Kane (U.S. Version)	3:49	3.0	¢
6	Go Getta	Young Jeezy and R. Kelly	The Inspiration	3:49	2.6	¢
7	One Wish	Ray J	Raydiation	5:37	3.0	¢
8	Into the Night	Santana and Chad Kroeger	Ultimate Santana	3:44	1.8	¢
9	Right Now (Na Na Na)	Akon	Freedom	4:01	2.9	¢
10	Forever	Drake, Kanye West, Lil Wayne and Eminem	More Than A Game	5:57	1.2	¢
11	Sexual Eruption [Explicit]	Snoop Dogg	Sexual Eruption [Explicit]	4:05	2.6	¢
12	Piece of Me	Britney Spears	Blackout	3:32	4.9	¢
13	Lolli Lolli (Pop That Body)	Three 6 Mafia	Lolli Lolli (Pop That Body)	4:18	4.2	¢
14	Promise	Ciara	Promise	4:31	3.3	¢
15	Stay Fly	Three 6 Mafia, Young Buck and 8Ball & MJG	Most Known Unknown (Clean)	3:56	1.3	¢
16	One Time	Justin Bieber	My World	3:36	3.2	¢
17	Break It Off	Rihanna and Sean Paul	A Girl Like Me	3:33	4.7	¢
18	Whatever It Is	Zac Brown Band	The Foundation	3:28	4.1	¢
19	Pump It	The Black Eyed Peas	Monkey Business	3:33	4.6	¢
20	My Love	Justin Timberlake and T.I.	My Love	4:43	1.4	¢
21	Pullin' Me Back	Chingy and Tyrese	Hoodstar (Edited)	3:54	5.0	¢
22	If U Seek Amy	Britney Spears	Circus	3:36	1.7	¢
23	When I'm Gone	Eminem	Curtain Call	4:40	1.5	¢
24	The Way I Live [Explicit]	Baby Boy da Prince	Across The Water [Explicit]	5:25	2.9	¢
25	Call Me When You're Sober	Evanescence	The Open Door	3:34	3.3	¢
26	Walk Away (Remember Me)	Paula DeAnda and The D.E.Y.	Paula DeAnda	3:58	1.3	¢
27	Welcome to the Black Parade	My Chemical Romance	The Black Parade (Amended Version)	5:11	1.7	¢
28	Then	Brad Paisley	Then	4:16	1.5	¢
29	We Be Burnin'	Sean Paul	The Trinity (Domestic Album Version)	3:37	3.2	¢
30	She Wolf	Shakira	She Wolf	3:07	4.7	¢

Next page





Step 3. Song Pricing

Group 2: Songs without Predicted Ratings

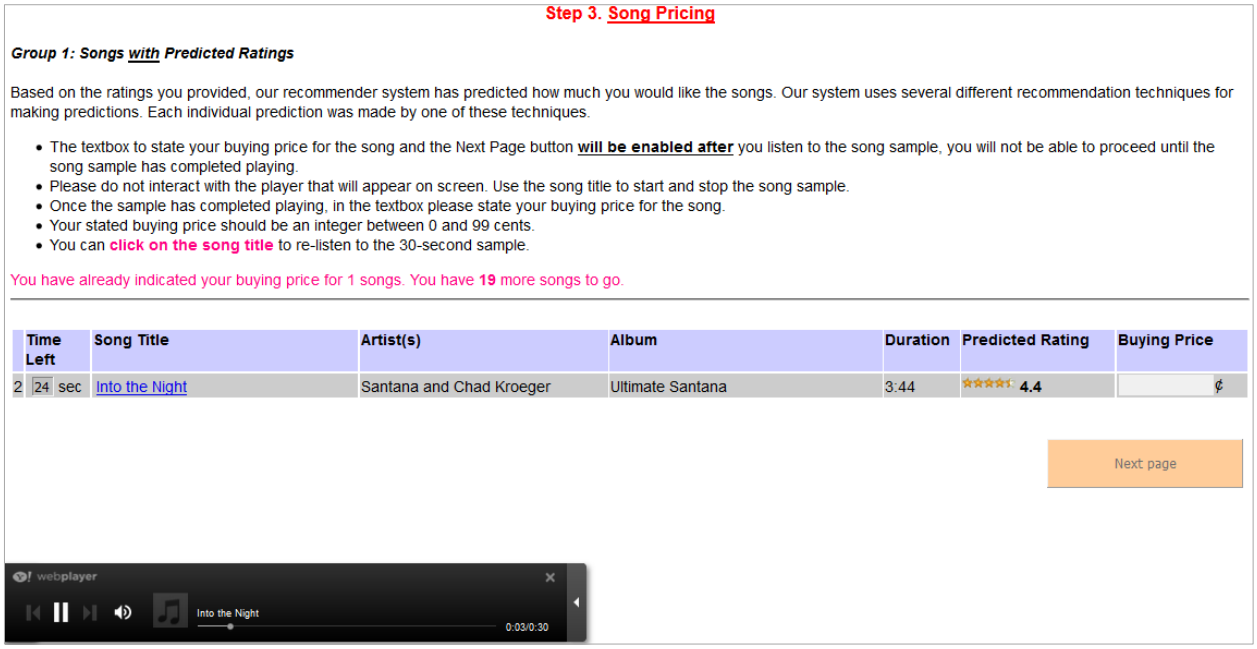
Our system did not have enough data to make predictions for the 10 songs below. However,

- In the textboxes, please also state your buying price for each of the songs below.
- Your stated buying price should be an integer between 0 and 99 cents.
- You can click the song title at any time to listen to a sample.

	Song Title	Artist(s)	Album	Duration	Buying Price
1	Shadow of the Day	Linkin Park	Minutes To Midnight	4:49	¢
2	Black Horse and the Cherry Tree	KT Tunstall	Black Horse And The Cherry Tree	2:54	¢
3	I Remember	Keyshia Cole	Just Like You	4:20	¢
4	Touch It	Busta Rhymes	Touch It Remixes	3:57	¢
5	How Do You Sleep?	Jesse McCartney and Ludacris	Departure - Recharged	3:27	¢
6	Feel Good Inc	Gorillaz and De La Soul	Feel Good Inc EP	3:41	¢
7	Turn My Swag On	Soulja Boy Tell 'Em	iSouljaBoyTellem	3:26	¢
8	Womanizer	Britney Spears	Circus	3:44	¢
9	Here (In Your Arms)	Hellogoodbye	Zombies! Aliens! Vampires! Dinosaurs!	4:01	¢
10	So What	Field Mob	Light Poles And Pine Trees	3:36	¢

Next page

Study 3 Only: One song per page



Step 3. Song Pricing

Group 1: Songs with Predicted Ratings

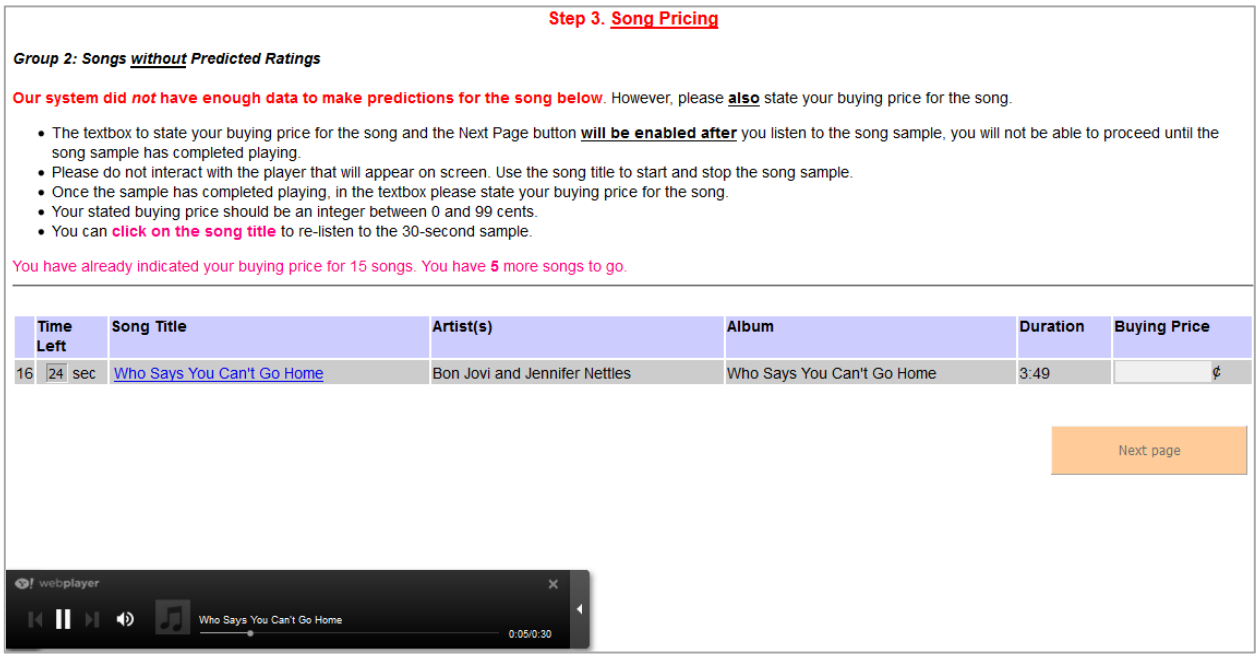
Based on the ratings you provided, our recommender system has predicted how much you would like the songs. Our system uses several different recommendation techniques for making predictions. Each individual prediction was made by one of these techniques.

- The textbox to state your buying price for the song and the Next Page button **will be enabled after** you listen to the song sample, you will not be able to proceed until the song sample has completed playing.
- Please do not interact with the player that will appear on screen. Use the song title to start and stop the song sample.
- Once the sample has completed playing, in the textbox please state your buying price for the song.
- Your stated buying price should be an integer between 0 and 99 cents.
- You can **click on the song title** to re-listen to the 30-second sample.

You have already indicated your buying price for 1 songs. You have **19** more songs to go.

	Time Left	Song Title	Artist(s)	Album	Duration	Predicted Rating	Buying Price
2	24 sec	Into the Night	Santana and Chad Kroeger	Ultimate Santana	3:44	4.4	¢

Next page



Appendix 2: Recommender System

Using the song rating data collected as Task 1 in Study 1, we compared 6 popular recommendation techniques, summarized in Table A1, to find the best-performing technique for our dataset. The techniques included simple user- and item-based rating average methods, user- and item-based collaborative filtering (CF) approaches and their extensions (Breese et al. 1998; Sarwar et al. 2001; Bell and Koren 2007), as well as a model-based matrix factorization algorithm (Funk 2006; Koren et al. 2009) popularized by the recent Netflix prize competition (Bennet and Lanning 2007).

Performance of recommendation algorithms was evaluated using two metrics: accuracy and coverage. Predictive coverage captures the percentage of items for which a given algorithm is able to make estimations of consumer's preference ratings. In many cases with the high sparsity of the dataset (i.e., not enough information is known for some items or users), some algorithms may provide predictions with high quality, but only for a small portion of the popular items that have received a great amount of consumers' ratings. The value of predictive coverage ranges from 0 to 1, with higher values being more desirable. A coverage value of zero means that the algorithm was unable to make any predictions, while a coverage value of 1 means the algorithm was able to predict ratings for all items and users.

The predictive accuracy metric computes the closeness of the system-predicted ratings to users' true ratings. The deviation between the system-predicted rating and the corresponding user's actual rating is computed as a *prediction error*. To compute prediction errors on a given dataset, the root mean squared error (RMSE) is one popular metric widely adopted by real-world applications to aggregate individual prediction errors into a composite measure. RMSE squares each individual error (deviation) before aggregating them, thereby assigning larger penalty for bigger prediction errors. Mathematically, $RMSE = \sqrt{\sum_{(u,i)\in T}(R^*_{u,i} - R_{u,i})^2 / |T|}$, where $R^*_{u,i}$ represents the system-predicted rating for user *u* and item *i*, $R_{u,i}$ is the actual user's rating, and *T* is the set of consumers' true ratings used for performance evaluation (i.e., test dataset). Since RMSE is an error-based accuracy metric, the lower the RMSE value the more accurate is the recommendation algorithm.

The predictive accuracy of each recommendation technique was evaluated using the standard leave-one-out cross-validation approach commonly applied in the fields of statistics and data mining (Mitchell 1997, p. 432). Using the leave-one-out approach, we exclude one rating from the given dataset and build a predictive model using all other ratings as the training data. The prediction error of the model is then evaluated using the one rating that was previously excluded from model training. This process is repeated on *every* rating in the dataset. Thus, for a dataset with *n* ratings, we construct *n* predictive models, each based on a different training dataset of *n*-1 ratings, and evaluate each model using the corresponding testing rating that is left out from model training. To compute the overall prediction accuracy, the prediction errors of these predictive models are aggregated as RMSE.

Based on the empirical results in Table A1, the predictive coverage values of all six tested recommendation algorithms were 1, meaning that all six algorithms were able to compute predictions for every song in our dataset. In terms of predictive accuracy, collaborative filtering (CF) algorithms provided the highest performance on our dataset; and, consistent with the literature (Deshpande and Karypis 2004; Adomavicius and Tuzhilin 2005), item-based CF performed slightly better than the user-based CF approach. On average, the predicted rating estimated by the item-based CF algorithm was about 0.779 stars off the

actual user ratings on a 5-star rating scale. Based on our results, we chose to use the item-based CF (IBCF) approach as the technique for our recommender system (for all three studies).

Table A1. Comparison of Recommendation Techniques on Song Rating Dataset.

Methodology	**Description**	**Predictive Accuracy (RMSE)**	**Predictive Coverage**
Item Average	Predicts each user-item rating as an average rating of that item (or user, for user-based approach).	0.8936	1
User Average		0.8505	1
User-Item Average	Computes unknown ratings with baseline (i.e., "global effects") estimates of corresponding users and items.	0.8190	1
Item-based CF	For each user-item rating to be predicted, finds the most similar items that have been rated by the user (or finds the most similar users who have rated the same item, for user-based approach) and computes the weighted sum of neighbors' ratings as the predicted rating. Similarity is computed based on Pearson correlation coefficient (Konstan et al. 1997; Sarwar et al. 2001).	0.7790	1
User-based CF		0.7893	1
Matrix Factorization	Decomposes the rating matrix to two matrices so that every user and every item is associated with a user-factor vector and an item-factor vector. Prediction is done by taking the inner product of the user-factor and item-factor vectors.	0.8120	1

Specifically, the item-based collaborative filtering approach makes predictions of unknown ratings for an item on the basis of the ratings previously received by this item's "nearest neighbors", i.e., other items that have similar rating patterns. That is, the value of the unknown rating for user *u* and item *i* is computed as an aggregate of the same user *u*'s ratings on similar items. A common aggregation approach is to use the weighted sum of the neighbors' ratings, where the similarity measure between two items is used as a weight. The more similar an item *j* and the target item *i* are, the more weight will be carried by the rating provided on item *j* by user *u* in the weighted sum when computing the prediction for item *i*. Specifically, the predicted rating for user *u* on item *i* is computed as:

$$R_{ui}^{*} = b_{ui} + \frac{\sum_{j \in N(u,i)} sim_{ij} \times \left(R_{uj} - b_{uj}\right)}{\sum_{j \in N(u,i)} \left|sim_{ij}\right|}$$

Here $N(u,i)$ is a set of "neighbors" with similar rating patterns to item i that have been previously consumed by user u, sim_{ij} is the similarity between items i and j, and b_{ui} is the baseline estimate for user u on item i, i.e., $b_{ui} = \mu + b_u + b_i$, where μ is the overall average rating of the dataset (i.e., global effect), b_u is the average observed deviation from μ on ratings provided by user u (i.e., user effect), and b_i is the average observed deviation from μ on ratings given to item i (i.e., item effect). Normalizing the dataset by removing the baseline estimate from each rating is a suggested data pre-processing step in the literature to further improve the predictive performance of collaborative filtering recommendation algorithms (Bell and Koren 2007). In our implementation, the similarity between two items is calculated as the Pearson correlation coefficient between their rating vectors (based on the users who rated these items in common), as suggested by the literature (Sarwar et al. 2001). We first computed the song similarities based on initial rating data collected in the song-rating task (in which each participant rated at least 50 songs). We then applied the item-based CF algorithm to estimate consumers' preference ratings on the 40 songs that were indicated as not owned by the participant.

Appendix 3. Robustness Checks Analyses for Studies 1-3

To check the robustness of the results, we conduct two additional sets of analysis for Studies 1-3:

- Fixed effect analysis: see Table A2a, A3a and A4a;
- Random-effect analysis with bootstrapping error estimation: see Table A2b, A3b and A4b.

Results from the two sets of analysis are highly consistent with the main results in the paper.

Table A2a. Study 1 fixed effects regression results

	Model 1 GLS, FE	**Model 2 LogNorm, FE**	**Model 3 Tobit**
ShownRating	3.552(0.92)***	0.163(0.03)***	3.533(0.78)***
PredictedRating	6.343(1.66)***	0.402(0.13)***	6.410(1.84)***
Constant	-5.237 (4.43)	0.698(0.37)	
R^2	0.0587	0.0543	
$Ç^2$	27.28***	65.52***	29.54***

Table A2b. Study 1 random effects regression results (with bootstrapped error term)

	Model 1 GLS, RE	Model 2 LogNorm, RE	Model 3 Tobit, RE
ShownRating	3.533(0.87)***	0.162(0.03)***	3.530(0.85)***
PredictedRating	6.235(1.26)***	0.396(0.10)***	6.215(1.76)***
Controls			
male	-9.466(7.57)	-0.910(0.43)*	-9.463(7.00)
undergrad	-3.819(22.37)	-0.335(1.25)	-3.826(21.91)
age	-1.366(3.49)	-0.069(0.22)	-1.362(3.80)
usedRecSys	-12.437(8.47)	-1.152(0.53)*	-12.439(8.20)
country	2.608(2.80)	0.204(0.17)	2.611(4.68)
rock	1.634(4.03)	0.275(0.23)	1.635(5.13)
hiphop	-0.468(2.91)	0.076(0.26)	-0.459(6.33)
pop	3.176(3.86)	0.091(0.24)	3.176(7.53)
recomAccurate	-3.963(4.77)	-0.238(0.38)	-3.961(6.45)
recomUseful	4.193(3.64)	0.294(0.33)	4.200(6.64)
buyingFreq	1.921(4.82)	0.227(0.21)	1.927(4.20)
songsOwned	-4.120(5.04)	-0.276(0.36)	-4.109(6.26)
constant	18.543(96.05)	1.220(5.77)	18.425(107.64)
R^2	0.214	0.268	
$Ç^2$	89.35***	144.28***	60.36***

Table A3a. Study 2 fixed effects regression results

	Model 1 GLS, FE	Model 2 LogNorm, FE	Model 3 Tobit
ShownRating	2.245(0.880)*	0.121(0.049)*	2.245(0.392)***
PredictedRating	8.787(1.414)***	0.567(0.082)***	8.797(0.832)***
Constant	7.613(23.674)	1.065(1.564)	7.587(25.841)
R^2	0.0587	0.0543	
$Ç^2$	27.28***	65.52***	29.54***

Table A3b. Study 2 random effects regression results (with bootstrapped error term)

	Model 1 GLS, RE	Model 2 LogNorm, RE	Model 3 Tobit, RE
Perturbation	2.245(0.85)*	0.121(0.05)*	2.245(0.84)*
PredictedRating	8.787(1.55)***	0.567(0.07)***	8.797(1.14)***
Controls			
Male	-2.203(5.17)	0.108(0.39)	-2.203(5.31)
Undergrad	-4.814(5.35)	-0.173(0.31)	-4.814(7.28)
Age	-0.351(1.08)	-0.007(0.08)	-0.351(1.36)
usedRecSys	3.944(4.9)	0.116(0.33)	3.944(4.94)
country	-2.646(2.17)	-0.166(0.17)	-2.647(2.4)
rock	-1.92(1.83)	-0.245(0.14)	-1.92(2.18)
hiphop	0.146(2.76)	-0.046(0.13)	0.146(2.21)
pop	-0.921(2.61)	-0.065(0.19)	-0.921(2.56)
recomAccurate	1.089(2.51)	0.081(0.21)	1.089(2.94)

recomUseful	3.886(2.68)	0.218(0.17)	3.886(3)
buyingFreq	2.712(2.21)	0.334(0.15)*	2.711(2.52)
songsOwned	-3.576(3)	-0.139(0.19)	-3.576(3.43)
constant	7.613(27.86)	1.065(1.84)	7.587(29.96)
R^2	0.1580	0.1850	
χ^2	72.07***	183.64***	111.81***

Table A4a. Study 3 fixed effects regression results

	Model 1 GLS, FE	**Model 2 LogNorm, FE**	**Model 3 Tobit**
ShownRating	1.993(0.65)*	0.099(0.03)***	2.003(0.72)*
PredictedRating	8.463(1.35)***	0.607(0.08)***	8.674(1.41)***
DisplayOrder	0.119(0.16)	-0.004(0.01)	0.117(0.14)
DisplayOrder × ShownRating	0.085(0.11)	-0.002(0.01)	0.092(0.1)
Constant	1.993(0.65)*	0.216(0.3)	-9.882(4.5)*
R^2	0.0635	0.0729	
χ^2	72.15***	65.86***	44.87***

Table A4b. Study 3 random effects regression esults (with bootstrapped error term)

	Model 1 GLS, RE	**Model 2 LogNorm, RE**	**Model 3 Tobit, RE**
ShownRating	1.998(0.7)*	0.1(0.04)*	1.999(0.58)***
PredictedRating	8.458(1.46)***	0.604(0.09)***	8.457(1.57)***
DisplayOrder	0.119(0.12)	-0.004(0.01)	0.119(0.17)
DisplayOrder × ShownRating	0.089(0.09)	-0.002(0.01)	0.09(0.08)
Controls			
Male	-1.412(3.72)	-0.129(0.27)	-1.412(5.58)
Undergrad	-0.837(7.28)	0.278(0.37)	-0.836(6.04)
Age	-0.086(1.08)	0.013(0.07)	-0.086(0.99)
usedRecSys	5.141(5.47)	0.359(0.32)	5.141(4.64)
country	3.935(1.83)*	0.241(0.12)*	3.935(2.05)
rock	1.457(2.23)	-0.036(0.15)	1.457(2.91)
hiphop	0.18(1.84)	-0.02(0.12)	0.18(2.07)
pop	0.548(1.87)	0.092(0.14)	0.548(2.18)
recomAccurate	-3.263(2.34)	-0.208(0.11)	-3.262(2.69)
recomUseful	4.288(2.03)*	0.331(0.12)*	4.288(2.41)
buyingFreq	2.806(2.23)	0.108(0.14)	2.806(2.94)
songsOwned	-1.819(3.01)	-0.122(0.21)	-1.819(3.21)
constant	-32.288(27.71)	-1.592(1.95)	-32.286(30.63)
R^2	0.1881	0.2455	
χ^2	175.71***	140.54***	124.70***

Appendix References:

Adomavicius, G., and Tuzhilin, A. 2005. "Toward the Next Generation of Recommendation System: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering* (17:6), pp. 734-749.

Bell, R.M., and Koren, Y. 2007. "Improved Neighborhood-Based Collaborative Filtering," *KDD Cup'07*, San Jose, California, USA: ACM, New York, NY, pp. 7-14.

Bennet, J., and Lanning, S. 2007. "The Netflix Prize," *KDD Cup and Workshop*, p. www.netflixprize.com.

Breese, J.S., Heckerman, D., and Kadie, C. 1998. "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI.

Deshpande, M., and Karypis, G. 2004. "Item-Based Top-N Recommendation Algorithms," *ACM Trans. Information Systems* (22:1), pp. 143-177.

Funk, S. 2006. "Netflix Update: Try This at Home." *Netflix Update: Try This at Home*, 2010, from http://sifter.org/~simon/journal/20061211.html

Koren, Y., Bell, R., and Volinsky, C. 2009. "Matrix Factorization Techniques for Recommender Systems," *IEEE Computer Society* (42), pp. 30-37.

Mitchell, T. 1997. *Machine Learning*. McGraw-Hill Science/Engineering/Math.

Sarwar, B., Karypis, G., Konstan, J.A., and Riedl, J. 2001. "Item-Based Collaborative Filtering Recommendation Algorithms," *the 10th International WWW Conference*, Hong Kong, pp. 285 - 295.