# ONLINE APPENDIX FOR "PEER EFFECTS IN PRODUCT ADOPTION"

Michael Bailey Drew Johnston Theresa Kuchler Johannes Stroebel Arlene Wong

## A.1 Random Phone Loss Instrument

In this appendix, we provide more details about our approach to identifying public posts about random phone loss events. We also provide evidence that random phone loss shocks are not correlated across individuals and their friends.

### A Random Phone Loss Classification

To construct our primary instrument, we need to identify users who have posted publicly about a random phone loss event. We take two different approaches to this classification: the first applies regular expression searches, while the second uses machine learning techniques. We find that both classifiers perform well at identifying relevant posts, but that the machine learning-based approach is superior to the regular expression-based classifier in terms of reducing both Type I and Type II errors. As a result, in the paper, we construct the random phone loss instrument using posts identified through the machine learning-based classifier.

**Regular Expression Classifier.** To build our regular expression-based classifier, we first compiled a list of common phrases (such as "broke my phone" or "phone got stolen") that were frequently used in Facebook posts concerning random phone loss events. A complete list of phrases is provided in Table A.1. We then automatically scanned all public Facebook posts by individuals in our sample during the period of our study, flagging posts that contained at least one of the phrases on our list.

Using this methodology to construct the instruments generates a strong first stage: about 9% of those individuals whose post is flagged end up purchasing a phone in the week of the post. Nevertheless, the classifier identifies a number of false positives (e.g., "So...I dropped my phone in the toilet yesterday...!! Still works tho!!") while failing to identify a number of more idiosyncratic descriptions of random phone loss (e.g., "R.I.P phone. You will be missed"). These descriptions are often picked up by our second classification model, described below, which uses a natural language processing algorithm.

**Machine Learning Classifier.** Our machine learning classifier is based on word embeddings, which allow us to translate the unstructured text of the public Facebook posts into features appropriate for machine learning models. Word embeddings are one of the most common tools used in Natural Language Processing (NLP), a sub-field of machine learning that aims to extract insights from data expressed in a human language. They are designed to express a word as a real-valued vector, with the direction and magnitude of each word vector learned from a set of training data. The size of the resulting vectors can vary across implementations, but vectors of 100–1,000 dimensions are commonly used; in our implementation, we use a 200-dimensional embedding. Even though these dimensions are often not easily interpretable, the geometry of the vectors represents semantic and syntactic features of each word, such as tense or quantity (see Mikolov, Yih and Zweig, 2013). Similar words tend to be represented by similar vectors (as measured by cosine similarity), and linear combinations of

**Table A.1:** Regular Expressions for Post Classification

| | | | |
|---|---|---|---|
| **Lost Phones** | | | |
| %lost phone% | %lost iphone% | %lost cell% | %lost my iphone% |
| %lost iphone% | %lost my phone% | %lost my cell% | |
| **Dropped Phones** | | | |
| %dropped phone% | %dropped my phone% | %phone dropped% | %cell was dropped% |
| %dropped iphone% | %dropped my iphone% | %cell dropped% | %phone got dropped% |
| %dropped cell% | %dropped my cell% | %phone was dropped% | %cell got dropped% |
| **Broken Phones** | | | |
| %phone%broke% | %broke my iphone% | %brokenphone% | %cellisbroke% |
| %broke phone% | %broke my cell% | %brokeniphone% | %cell?s broke% |
| %broke iphone% | %broken cell% | %brokencell% | %brokecell% |
| %broken phone% | %cell broke% | %cells broke% | %broke my phone% |
| %broken iphone% | %cell is broke% | %cell just broke% | |
| **Destroyed Phones** | | | |
| %destroyed phone% | %destroyed my cell% | %phone is destroyed% | %cell got destroyed% |
| %destroyed iphone% | %phone destroyed% | %cell is destroyed% | %destroyed cell% |
| %cell destroyed% | %phone got destroyed% | %phonedestroyed% | %destroyed my phone% |
| %phones destroyed% | %cell got destroyed% | %destroyed my iphone% | %cells destroyed% |
| %phone was destroyed% | %celldestroyed% | | |
| **Killed Phones** | | | |
| %killed phone% | %phoneisdead% | %phone is dead% | %cell just died% |
| %killed iphone% | %celldead% | %cell is dead% | %phone got killed% |
| %killed cell% | %cellisdead% | %phone has died% | %killed my phone% |
| %phone dead% | %cell has died% | %cell got killed% | %killed my iphone% |
| %cell dead% | %phone died% | %killed my cell% | %phones dead% |
| %cell died% | %phone was killed% | %phonedead% | %cells dead% |
| %phone just died% | %cell was killed% | | |
| **Smashed Phones** | | | |
| %smashed phone% | %smashed my iphone% | %phone smashed% | %phone is smashed% |
| %smashed iphone% | %smashed my cell% | %cell smashed% | %cell is smashed% |
| %smashed cell% | %phonesmashed% | %phones smashed% | %phone was smashed% |
| %smashed my phone% | %cellsmashed% | %cells smashed% | %cell was smashed% |
| **Shattered Phones** | | | |
| %shattered phone% | %shattered my iphone% | %phone shattered% | %phone is shattered% |
| %shattered iphone% | %shattered my cell% | %cell shattered% | %cell is shattered% |
| %shattered cell% | %phone shattered% | %phones shattered% | %phone was shattered% |
| %shattered my phone% | %cell shattered% | %cells shattered% | %cell was shattered% |
| **Busted Phones** | | | |
| %busted phone% | %busted my cell% | %phone busted% | %phone is busted% |
| %busted iphone% | %busted my iphone% | %cell busted% | %cell is busted% |
| %busted cell% | %phone busted% | %phones busted% | %phone was busted% |
| %busted my phone% | %cellbusted% | %cells busted% | %cell was busted% |
| **Damaged Phones** | | | |
| %damaged phone% | %damaged my iphone% | %cell damaged% | %phone was damaged% |
| %damaged iphone% | %damaged my cell% | %phone got damaged% | %cell was damaged% |
| %damaged my phone% | %phone damaged% | %cell got damaged% | |
| **Stolen Phones** | | | |
| %stole my phone% | %phone stolen% | %cell got stolen% | %stole my iphone% |
| %cell stolen% | %phone was stolen% | %stole my cell% | %phone got stolen% |
| %cell was stolen% | | | |
| **Not Working Phones** | | | |
| %phone stopped working% | %phone not working% | %cells not working% | %cell is not working% |
| %cell not working% | %phone isn?t working% | %cell stopped working% | %phones not working% |
| %phone is not working% | %cell isn?t working% | | |
| **Other** | | | |
| %phoneless% | %broke%screen% | %shattered%screen% | %contact me here% |
| %cellless% | %screen%smashed% | %screen%crack% | %no phone% |
| %smashed%screen% | %crack%screen% | %you can reach me% | %screen%broke% |
| %screen%shattered% | %contact me on% | %hit me up on here% | |

**Note:** Table shows the regular expressions used to flag posts about random phone loss. % is a wildcard capturing any number of characters (including 0), ? is a wildcard for any single character.

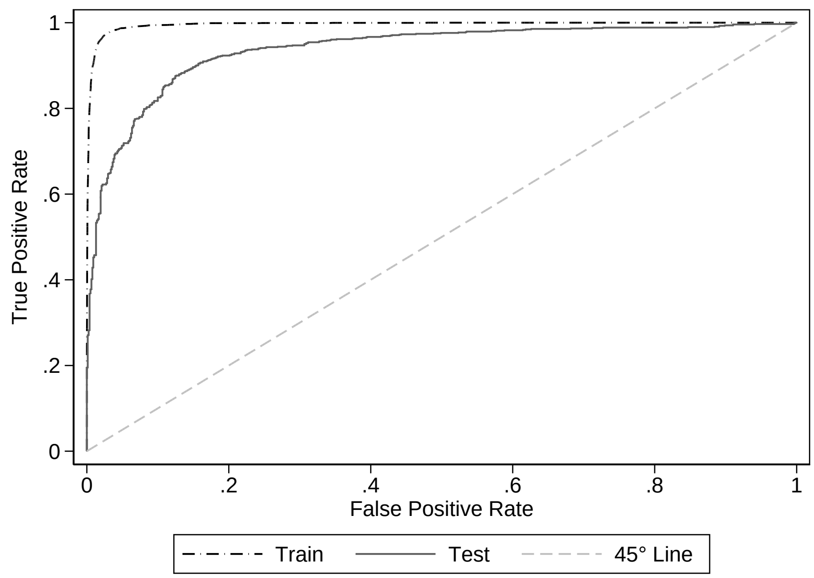
word-embedding vectors contain syntax and semantic meaning. For instance, after converting words to their embeddings, the embedding most similar to $(\overrightarrow{King} - \overrightarrow{Man} + \overrightarrow{Woman})$ is $\overrightarrow{Queen}$.

Several approaches can be taken to learning these vectors from a corpus of text. We chose a skip-gram-based approach, in which a neural network is trained to predict the words surrounding a given term in the corpus. No information is provided to the model about English grammar or syntax—all language features are learned directly from the corpus of text. To train our model, we chose the entire English-language version of Wikipedia as our corpus. Wikipedia is a common corpus in the NLP literature, since it covers a broad range of topics in considerable detail, a feature which helps to train a general-purpose set of embeddings (Bojanowski et al., 2016). We train the embeddings using FastText, a popular open-source library created by researchers at Facebook. After training the word embeddings, we have a model that can transform any word into a 200-dimensional vector. We then concatenate all vectors corresponding to the words in a public post, creating a $200 \times N$ matrix that represents the post, where N is the number of words in the post.

In the next step, we train a convolutional neural network to classify these matrices. Convolutional neural networks (CNNs) are commonly used in natural language processing, as they allow for the creation of very flexible non-linear models that can capture sentence context. This is important for our task, since the ability to distinguish between sentences like "I broke my friend's phone" and "My friend broke my phone" is crucial. This distinction would have been hard to capture in simpler text classification models that do not respect word order (often called "bag of words" approaches). Convolutional neural networks differ from multi-layer perceptrons (or "vanilla neural networks") in several ways that are useful for working with text data (Kim, 2014). Specifically, CNNs create convolutional filters that transform the underlying data between traditional layers of the neural network. These filters alter the data to amplify the features that are most relevant in the final classification step. The exact features captured by the filters are determined automatically during the training process. In text data, the filters usually capture and transform multi-word patterns. We employ convolutions of widths 2, 3, 4, 5, and 10 in our work. In general, these convolutions are effective at preserving medium-distance relationships between words, allowing the algorithm to distinguish between phrases like "my phone," "his phone," and "a phone." CNNs can also employ max pooling, which is the selective dropping of data perceived by the neural network to be unimportant. Max pooling normally occurs after a layer of convolutions. This step is important when working with text data, since the dimensions of the input matrix vary between observations. Max pooling allows the model to effectively drop data throughout the process until an appropriately-sized array of features is obtained for the final classification step.

We provided substantial training data to create the final model to identify posts concerning random phone loss events. To this end, we hand-classified around 8,000 posts picked up by the regular expression model; hand classification identified about 40% of the posts as false positives. We added another 1,000 hand-classified posts that referenced phones in some way but that were not picked up by the regular expression classifier; hand classification revealed that about 20% of these posts concerned a random phone loss, and therefore they represent false negatives for the regular expression classifier. We supplemented these posts with 25,000 posts that had nothing to do with cell phones. The trained model achieved a 0.94 ROC–AUC on 1,800 labeled examples that were set aside as a test set, and which were therefore not used in training the model (see Figure A.1).[1] To create our primary instrument, we then used the trained model to classify all public posts on Facebook in our sample weeks, identifying those posts that the algorithm detected as being about breaking or losing a phone.

**Figure A.1:** ROC Curves for Post-Detection CNN




**Note:** Figure shows the ROC curves for our post-detection CNN. Both the training curve and the test curve are pictured.

Due to the severe class imbalance in our classification problem, some false positives remained with the machine learning classifier. In most cases, the classifier expressed uncertainty about these posts, with estimated probabilities of between 10% and 30% that the posts concerned a random phone loss. Posts to which the algorithm assigned probabilities above 0.3, on the other hand, are almost always true positives. We thus find that we achieve a stronger first stage if we use the neural network's outputs in tandem with those from the regular expression, in order to give a second check on false positives. In our final model, we therefore set $\mathbb{1}(RandomPhoneLoss_{i,t}) = 1$ if the regular expression condition was true and the CNN's estimated probability was higher than 0.1, or if the regular expression conditions was false and the CNN had an estimated probability above 0.3. This methodology, which is inspired by the notion of ensemble classifiers in machine learning, substantially reduces our false positive rate and generates a stronger first stage than was possible with either of the two classifiers used individually.[2]

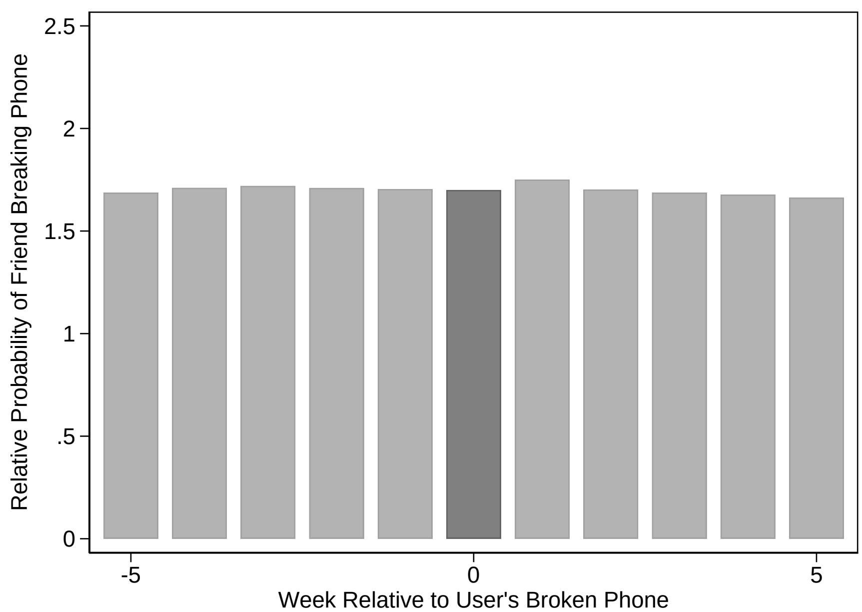
[1] ROC–AUC is a common metric used in machine learning to evaluate the performance of a predictive model. It can be calculated by plotting a graph of a model's true positive rate with respect to the false positive rate across all threshold scores and finding the area under the curve of the line formed. Intuitively, it corresponds to the probability a classifier will rank a random positive example above a random negative example (when tasked with distinguishing positive and negative examples). Regardless of the class balance, a score of 0.5 is awarded to a meaningless model (one as good as random guessing), while 1.0 is a perfect score.

[2] The thresholds may appear to be low, but we chose a threshold that balanced the number of the posts found with the conditional probability of switching of the posters. Increasing the thresholds we use would increase the certainty that any individual post is truly a post about a user who breaks their phone, but it would cause the number of potential posts found to be smaller, which negatively impacts the strength of our first stage. We found that using a lower threshold (which permits somewhat more Type II errors in order to reduce the number of Type I errors) gave us the strongest first stage, although raising the threshold somewhat does not change our main findings.

## B Random Phone Loss Instrument Independence

An important assumption in our empirical analysis is that random phone loss events are not correlated across individuals and their friends. Figure A.2 provides support for this assumption. To construct this graph, we consider two groups: users who post about breaking their phone in week $t = 0$ and users who do not make any such post. For each group, we calculate the average percentage of their friends' posts that are about breaking a phone in weeks $t - 5$ to $t + 5$. We then graph the ratio of the percentages for the first and second group. We see that, even though users who post about breaking their phone tend to have more friends who themselves post about breaking their phone, the level of posting is constant across time. There does not seem to be any indication that users and their friends tend to disproportionately break their phones at the same time. That said, in our baseline regressions we control for the fact that different groups tend to post about breaking their phones at different rates by controlling for the number of the user's friends who break their phone in the 12 months prior to our sample. We also control for the average level of *ProbBuyRandomPhoneLoss* among this group.

**Figure A.2:** Random Phone Loss Among Friends Relative to User Random Phone Loss




**Note:** Figure shows the probability that friends of a user who breaks their phone in week 0 post about breaking their own phones in the weeks before and after. The probability is expressed relative to that of friends of users who do not post about a random phone loss.

## A.2 Purchasing Propensity Predictions

Throughout this paper, we use several predicted probabilities to construct both our instruments and our controls. In a broad sense, all of these classifiers have a similar goal: we aim to predict users' phone-purchasing behavior based on information about their demographics, their social networks, and their current phones. The relationships between these features and phone purchasing are complicated and non-linear, so we fit neural networks to predict these propensities using flexible functional forms.

We train these models using data from weeks 2016-13 to 2016-15 and 2016-25 to 2016-27 (where weeks are indicated in the format yyyy-ww),[3] and we then use the models to predict propensities for all observations in the main sample, which runs from 2016-19 to 2016-22. We use five-fold cross-validation to select the hyperparameters for each model. The selected hyperparameters, as well as the resulting in-sample and out-of-sample ROC–AUC scores, are summarized in Table A.2.

**Table A.2:** Predicting Purchasing Probabilities

| | Test Set ROC–AUC | Regression Sample ROC–AUC | Best Layer Sizes |
|---|---|---|---|
| $ProbBuyRandomPhoneLoss_{i,(t,t+1)}$ | 0.679 | 0.665 | [10] |
| $ProbBuy2y_{i,(t,t+1)}$ | 0.640 | 0.634 | [20, 5] |
| $ProbBuyUncond_{i,t}$ | 0.705 | 0.699 | [10, 5] |
| $ProbBuyRandomPhoneLoss^{c}_{i,(t,t+1)}$ | 0.720 | 0.716 | [10, 5] |
| $ProbBuyUncond^{c}_{i,(t,t+1)}$ | 0.755 | 0.752 | [20, 5] |
| $ProbBuyUncond^{p}_{i,(t,t+1)}$ | 0.786 | 0.782 | [20] |

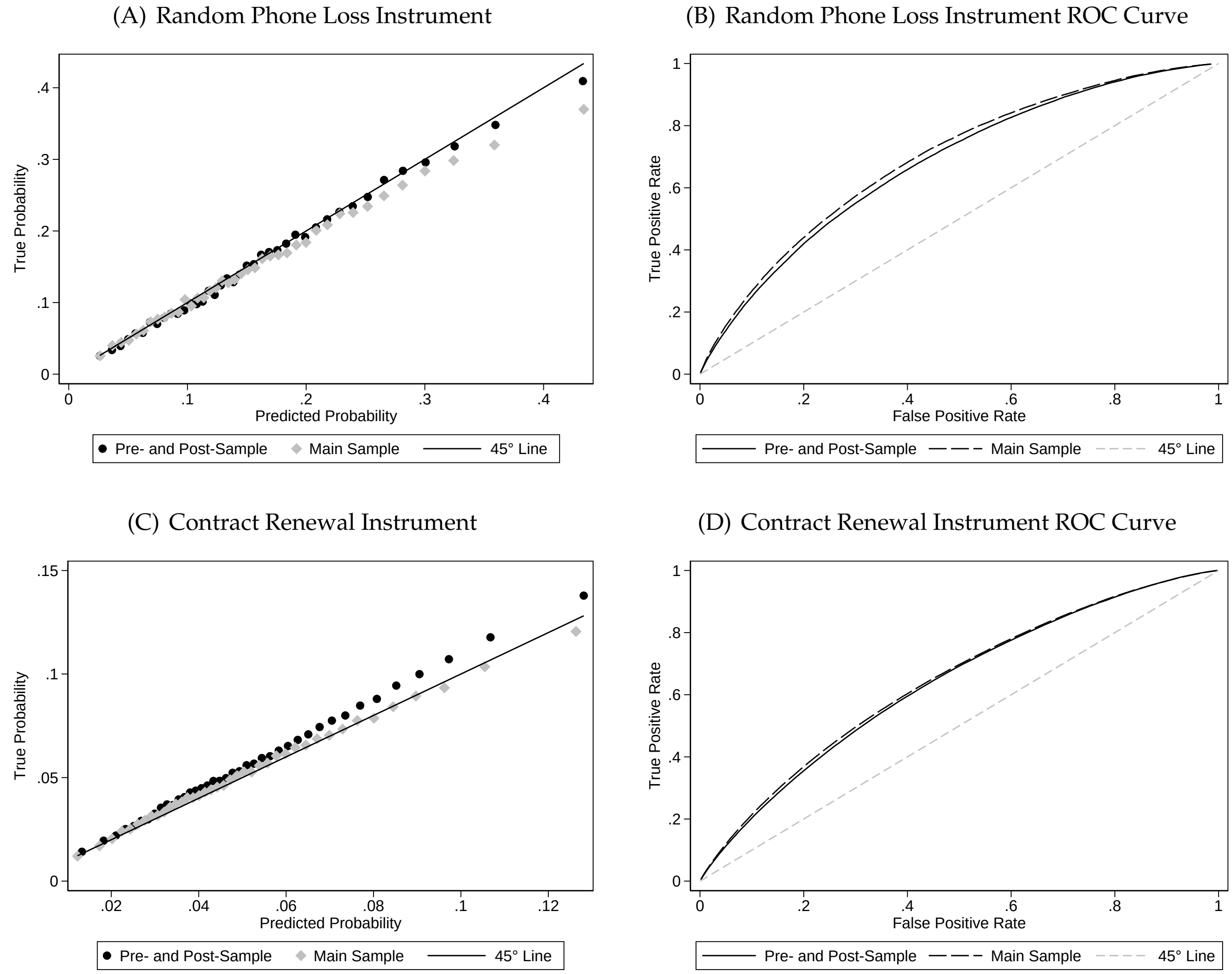
**Note:** Table shows summary statistics on the predictive power of the classifiers used and the best layer size for the classifier with the best performance on a validation sample. All hyperparameters are determined by five-fold cross-validation. The Test Set ROC–AUC is calculated using a held-out test data set drawn from the same sample as the training data (weeks 2016–13 to 2016–15 and 2016–25 to 2016–27). The Regression Sample ROC–AUC is calculated using data from the main period studied in our regressions (weeks 2016–19 to 2016–22). The scores for the final three groups (which each have several classifications for the different brands or phone models) refer to the averaged one-versus-all ROC–AUC scores for each possible classification.

We train both conditional and unconditional models. In the unconditional models, we predict users' purchasing decisions in weeks $t$ and $t+1$ on the basis of their characteristics in week $t$. For all models, the set of observable characteristics used to train the model is as follows: current phone age,[4] current phone model, carrier, age, user browser, Instagram usage flag, state, education level, friend count, activity flags, account age, profile picture flag, number of friendships initiated, gender, and area average income. Since the models are unconditional, we use all user-weeks in our training period as training data. We then predict unconditional probabilities for all users in our main sample, and we use these unconditional probabilities as controls in the regressions in Section III.

[3] Training on held-out data is important in this case, as training using in-sample data would run the risk of overfitting the model, which would bias the IV estimates towards the OLS coefficients.

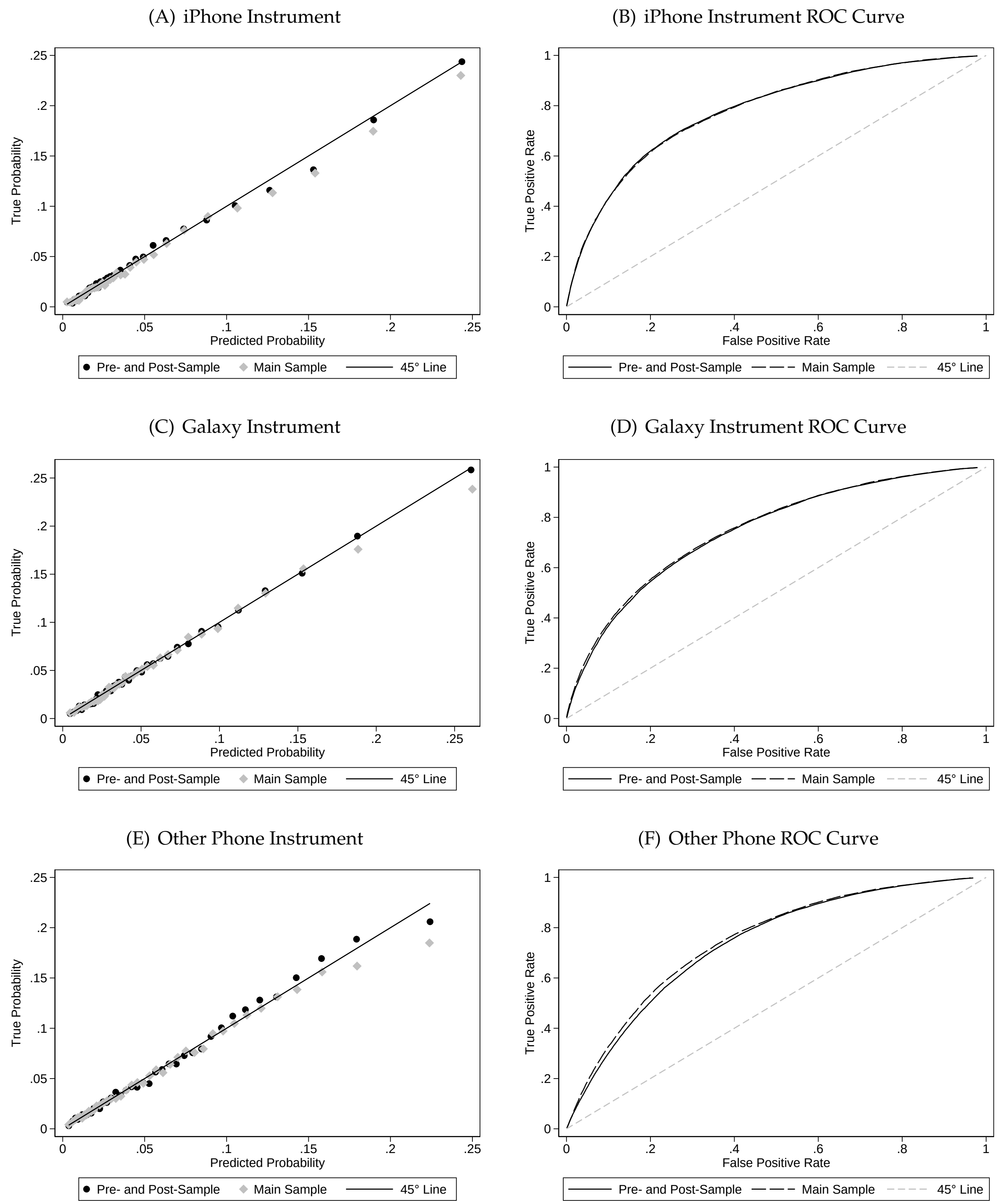
[4] We do not include this characteristic in our models studying switching probability at the contract renewal threshold, as all users in the training set for this classifier have their phone age in a narrow range. This feature makes predicting the conditional probability for users whose phone age is outside this range difficult.

## **Figure A.3:** Baseline Instruments




**Note:** Panels A and C in the left column show binscatter plots of the fit of the probabilities to purchase a new phone given the random events underlying our two instruments: random phone loss in Panel A, and phone age of 2 years in Panel C. Panels B and D in the right column present Receiver Operating Characteristic (ROC) curves for each of these estimated probabilities. All plots only include users for whom $\mathbb{1}(Instrument_{i,t}) = 1$. The regression results using these instruments are shown in Table 2. The "Pre- and Post Sample" is a held-out set of observations from weeks 2016–13 to 2016–15 and 2016–25 to 2016–27, the same weeks that were used to train the data. The "Main Sample" is all observations from the period 2016–18 to 2016–22, the period used to construct our main panel.

**Figure A.4:** Brand Instruments




**Note:** The left column shows binscatter plots of the fit of the probabilities to purchase a new phone of a given brand for individuals with a random phone loss. The right column presents Receiver Operating Characteristic (ROC) curves for each of these instruments. All plots only include users for whom $\mathbb{1}(BrokenPhone_{j,t}) = 1$. The regression results using these three instruments are shown in Table 3. The "Pre- and Post Sample" is a held-out set of observations from weeks 2016–13 to 2016–15 and 2016–25 to 2016–27, the same weeks that were used to train the data. The "Main Sample" is all observations from the period 2016–18 to 2016–22, the period used to construct our main panel.

In the conditional models, we aim to predict a user's probability of purchasing a phone in weeks $t$ and $t+1$, conditional on some behavior or trait in week $t$, such as posting about a random phone loss or having a phone at the contract renewal threshold. We fit one such model for each of the two instruments, and we predict each conditional probability for all user-weeks in our main sample.

In addition to the summary statistics on the model fit provided in Table A.2, Figure A.3 presents information on the performance of the various classifiers for our two baseline instruments. The left column shows binscatter plots of the predicted probability against the realized probability for both a hold-out sample from our training data period ("Pre- and Post-Sample") and the actual regression data set ("Main Sample"). Reassuringly, the training data line up on the 45-degree line, and the regression sample data, which were not used to train the model, also align closely. This finding suggests that our models are relatively stable over time. The horizontal axis shows the range of predicted probabilities. For example, we find the predicted probabilities for purchasing in weeks $t$ and $t+1$ after posting about a random phone loss in week $t$ range to be between 5% and 50%. This result highlights the value of using the computed conditional probability—rather than just the number of friends with a random phone loss—as an instrument for the number of friends buying a phone.

We additionally fit models that predict acquisitions of particular phone types. The same features used to generate the general switching predictions are highly predictive of which phone a user buys, and our models therefore have strong predictive power. We create classifiers at two levels of granularity. First, we train a classifier to predict a user's probability of buying any of three mutually exclusive and exhaustive categories of cell phones, iPhones, Galaxies, and other phones, as well as their probability of not buying any phone. Second, as described in Appendix A.4, we predict the user's probability of purchasing each of the 20 most commonly-purchased phones, three residual categories (other iPhone, other Galaxy, other), and no phone at all. For both granularities, we train unconditional models for use as controls in our regressions: $ProbBuyUncond^{c}_{i,t}$ for the brand-level granularity, and $ProbBuyUncond^{p}_{i,t}$ for the model-level granularity. At the brand level, we also train a classifier conditional on a user breaking or losing their phone, $ProbBuyRandomPhoneLoss^{c}_{i,t}$; these predicted probabilities are used to construct our instruments as described in detail in Section IV, and Figure A.4 explores the performance of these predictors.[5] Tables A.2 and A.3 show the performance of the brand-level and model-level propensity predictions.

[5] We do not estimate IV regressions at the model level, since we find that the predictions within a brand tend to be highly collinear, making first stage estimation complicated (see Appendix A.4 for further discussion).

**Table A.3:** Phone-Specific ROC–AUC

| | Unconditional | | Conditional on Losing Phone | |
|---|---|---|---|---|
| | **Test Set ROC–AUC** | **Regression Sample ROC–AUC** | **Test Set ROC–AUC** | **Regression Sample ROC–AUC** |
| **Brand** | **0.755** | **0.752** | **0.720** | **0.716** |
| iPhone | 0.734 | 0.734 | 0.757 | 0.758 |
| Galaxy | 0.772 | 0.769 | 0.729 | 0.726 |
| Other | 0.808 | 0.799 | 0.711 | 0.707 |
| No Purchase | 0.708 | 0.704 | 0.684 | 0.673 |
| **Specific** | **0.786** | **0.782** | | |
| iPhone 6S | 0.782 | 0.785 | | |
| iPhone 6S Plus | 0.711 | 0.718 | | |
| iPhone SE | 0.773 | 0.781 | | |
| iPhone 6 | 0.738 | 0.733 | | |
| iPhone 5S | 0.720 | 0.716 | | |
| iPhone 6 Plus | 0.680 | 0.688 | | |
| Other iPhone | 0.706 | 0.689 | | |
| Galaxy S7 | 0.820 | 0.816 | | |
| Galaxy S7 Edge | 0.790 | 0.786 | | |
| Galaxy Core Prime | 0.797 | 0.793 | | |
| Galaxy Note 5 | 0.808 | 0.809 | | |
| Galaxy J7 | 0.851 | 0.859 | | |
| Galaxy Grand Prime | 0.793 | 0.795 | | |
| Galaxy S5 | 0.758 | 0.741 | | |
| Galaxy S6 | 0.763 | 0.755 | | |
| Other Galaxy | 0.767 | 0.758 | | |
| Tribute 5 | 0.857 | 0.847 | | |
| K10 | 0.844 | 0.846 | | |
| G5 | 0.830 | 0.810 | | |
| G Stylo | 0.851 | 0.844 | | |
| Desire 626s | 0.852 | 0.839 | | |
| One Touch | 0.881 | 0.875 | | |
| Other | 0.800 | 0.793 | | |
| No Purchase | 0.702 | 0.700 | | |

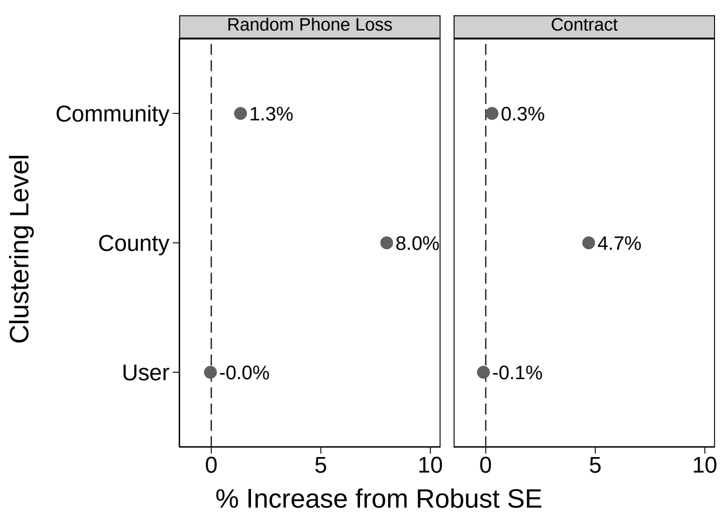
**Note:** Table shows summary statistics on the classifiers used to predict the phone type a user purchases. Since ROC–AUC is a score for binary classification problems, the multi-class ROC–AUC scores (bolded) represent the averaged one-versus-all ROC–AUC scores of each individual classification problem (see also Table A.2). We do not display conditional ROC–AUC scores for specific phones, as we report only OLS versions of these regressions due to the high correlation between within-brand scores as discussed further in Section A.6.

## A.3 Statistical Inference – Robustness Exercises

As discussed in Section C, to construct our baseline standard errors, we follow the suggestions by Eckles, Kizilcec and Bakshy (2016) and Zacchia (2020) and partition the Facebook social graph into a number of communities with limited cross-community dependence, allowing us to cluster standard errors at the community level. Starting from a sparse matrix representing the Facebook social graph, a distributed variant of the Kernighan-Lin algorithm is used to divide the global Facebook social graph into about 21,000 distinct communities. Individuals in our sample are assigned to their communities created by this algorithm. The 0.2% of our sample assigned to communities with fewer than 100 other members of our sample are grouped into a "residual" community (these individuals are likely to be recent immigrants, who are members of communities where most members are outside the United States). Overall, the 81 million users in our primary sample are assigned to 5,140 distinct communities with an average size of 15,910 users. The average user in our sample has 53.4% of her friends within the same community; at the 10th/50th/90th percentiles of our sample, these numers are 21%/54%/84%.

The first row of Figure A.5 compares the standard errors using this clustering approach to baseline heteroskedasticity-robust standard errors for the estimates corresponding to column 4 (left panel) and column 5 (right panel) of Table 2. We find that the community-clustered standard errors are very similar in magnitude to the heteroskedasticity-robust standard errors, suggesting that, in our setting, across-individual spillovers seem to not confound our inference, at least after conditioning on the large set of controls and fixed effects. The other two rows compare standard errors clustered at the county level (middle row) and individual level (bottom row) to heteroskedasticity-robust standard errors. As before, these standard errors are at most 7.2% larger (and, in some cases, even marginally smaller) than heteroskedasticity-robust standard errors.

**Figure A.5:** Comparison of Standard Errors




**Note:** Figure compares standard errors using various clustering approaches to heteroskedasticity-robust standard errors for the instrumental variables estimates corresponding to column 4 (left panel) and column 5 (right panel) of Table 2. The top row uses standard errors clustered at the community level (see Appendix A.3 for discussion on our approach to detecting communities), the middle row uses standard errors clustered at the county level, and the bottom row uses standard errors clustered at the user level.

## A.4 Heterogeneity Models

In Section B, we estimate heterogeneities in influence and susceptibility to influence across individuals and relationships. In this appendix, we provide the regression specifications we use in these analyses.

**Heterogeneity by Relationship and Friend Characteristics.** We first analyze heterogeneities in influence according to relationship characteristics and friend characteristics. In each regression, we consider a mutually-exclusive and exhaustive group of characteristics $G$. For relationship characteristics, one group of characteristics corresponds to strong and weak friendships, and another to geographically proximate or non-proximate friends. For friend characteristics, one group of characteristics corresponds to friend ages, with three conditions $g \in G$ capturing friends aged 18–25, 26–40, and 40+. We use these conditions to create new instruments and endogenous variables for each $g \in G$:

$$Instrument_{i,t}^{Lose,g} = \sum_{j \in Fr(i)} ProbBuyRandomPhoneLoss_{j,t} \cdot \mathbb{1}(LostPhone_{j,t}) \cdot \mathbb{1}(Condition_{j,t}^{g})$$

$$FriendsBuyPhone_{i,(t,t+1)}^{g} = \sum_{j \in Fr(i)} \mathbb{1}(BuysPhone)_{j,(t,t+1)} \cdot \mathbb{1}(Condition_{j,t}^{g})$$

We also create two new sets of controls. The first set counts the number of friends of user $i$ who are members of each group $g$, while the second controls for the average conditional probability among the user's friends in each group:

$$Friends_{i,t}^{g} = \sum_{j \in Fr(i)} \mathbb{1}(Condition_{j,t}^{g})$$

$$AllFriendsAvgProbBuyRPL_{i,t}^{g} = \frac{1}{Friends_{i,t}^{g}} \sum_{j \in Fr(i)} ProbBuyRandomPhoneLoss_{i,t}^{g} \cdot \mathbb{1}(Condition_{j,t}^{g})$$

Using these new variables, we estimate one first stage per condition as well as a single second stage:

$$FriendsBuyPhone_{i,(t-1,t)}^{g} = \sum_{g \in G} \delta_g Instrument_{i,t-1}^{Lose,g} + \sum_{g \in G} \psi_g Friends_{i,t-1}^{g} + \sum_{g \in G} \theta_g AllFriendsAvgProbBuyRPL_{i,t-1}^{g} + \omega X_{i,t} + e_{i,t}$$

$$\mathbb{1}(BuysPhone_{i,t}) = \sum_{g \in G} \beta_g \widehat{FriendsBuyPhone}_{i,(t-1,t)}^{g} + \sum_{g \in G} \Psi_g Friends_{i,t-1}^{g} + \sum_{g \in G} \Theta_g AllFriendsAvgProbBuyRPL_{i,t-1}^{g} + \gamma X_{i,t} + \epsilon_{i,t}$$

In each case, we include the same fixed effects and controls outlined in column 2 of Table 2.

**Heterogeneity by User Characteristics.** In Section B, we also analyze how susceptibility to influence varies according to user characteristics. We employ a similar approach to that outlined above to understand these heterogeneities:

$$Instrument_{i,t}^{Lose,g} = \mathbb{1}(Condition_{i,t}^{g}) \cdot \sum_{j \in Fr(i)} ProbBuyRandomPhoneLoss_{j,t} \cdot \mathbb{1}(LostPhone_{j,t})$$

$$FriendsBuyPhone^{g}_{i,(t,t+1)} = \mathbb{1}(Condition^{g}_{i,t}) \cdot \sum_{j \in Fr(i)} \mathbb{1}(BuysPhone)_{j,(t,t+1)}$$

Our regressions are similar to those outlined above, but when studying user characteristics, we do not need to include separate controls indicating whether user $i$ meets condition $g$, because each of these conditions is already included as a fixed effect at a finer level of granularity. As before, we have one first stage for each group $g$, as well as one second stage regression:

$$FriendsBuyPhone^{g}_{i(,t-1,t)} = \sum_{g \in G} (\delta_g Instrument^{Lose,g}_{i,t-1}) + \omega X_{i,t} + e_{i,t}$$
$$\mathbb{1}(BuysPhone_{i,t}) = \sum_{g \in G} (\beta_g \widehat{FriendsBuyPhone}^{g}_{i,(t-1,t)}) + \gamma X_{i,t} + \epsilon_{i,t}$$

In each case, we include the same fixed effects and controls outlined in column 2 of Table 2.

**Heterogeneity by Pairwise Characteristics** In Section B, we also study variation in peer effects according to user and friend characteristics simultaneously. We consider the same set $G$ of characteristics for both the users and friends. We construct new instruments and endogenous variables as follows:

$$\begin{aligned} Instrument^{Lose,g_1,g_2}_{i,t} =& \mathbb{1}(Condition^{g_1}_{i,t+1}) \cdot \\ & \sum_{j \in Fr(i)} [ProbBuyRandomPhoneLoss_{j,t} \cdot \mathbb{1}(LostPhone_{j,t}) \cdot \mathbb{1}(Condition^{g_2}_{j,t})] \end{aligned}$$
$$FriendsBuyPhone^{g_1,g_2}_{i,(t,t+1)} = \mathbb{1}(Condition^{g_1}_{i,t+1}) \cdot \sum_{j \in Fr(i)} [\mathbb{1}(BuysPhone)_{j,(t,t+1)} \cdot \mathbb{1}(Condition^{g_2}_{j,t})]$$

We use these variables to construct first stages that allow us to separately gauge the influence of friends of each type on users of each type. The regressions for the first and second stages, respectively, are

$$\begin{aligned} FriendsBuyPhone^{g_1,g_2}_{i,(t-1,t)} =& \sum_{g_1 \in G} \sum_{g_2 \in G} (\delta_{g_1,g_2} Instrument^{Lose,g_1,g_2}_{i,t-1}) + \sum_{g \in G} \psi_g Friends^{g}_{i,t-1} + \\ & \sum_{g \in G} \theta_g AllFriendsAvgProbBuyRPL^{g}_{i,t-1} + \omega X_{i,t} + e_{i,t} \end{aligned}$$

$$\begin{aligned} \mathbb{1}(BuysPhone_{i,t}) =& \sum_{g_2 \in G} \sum_{g_1 \in G} (\beta_{g_1,g_2} \widehat{FriendsBuyPhone}^{g_1,g_2}_{i,(t-1,t)}) + \sum_{g \in G} \Psi_g Friends^{g}_{i,t-1} + \\ & \sum_{g \in G} \Theta_g AllFriendsAvgProbBuyRPL^{g}_{i,t-1} + \gamma X_{i,t} + \epsilon_{i,t} \end{aligned}$$

In each case, we the same fixed effects and controls outlined in column 2 of Table 2. For each individual, we additionally control for the number of friends in each group.

## A.5 Theoretical Model: Peer Effects, Demand Elasticities, and Prices

In this appendix, we describe a simple model of price setting under monopolistic competition that allows us to illustrate an important channel through which peer effects in consumption can affect demand elasticities and markups.[6]

### A Consumer Preferences

There are $N$ consumers, and the consumption of each consumer is infinitesimally small relative to total demand. An individual consumer $i$'s utility function is given by:

$$U_i = \left( \int_0^n R_i \left( Q_j \right) q_{ij}^{\rho_i} dj \right)^{\frac{1}{\rho_i}}, \tag{A.1}$$

where $q_{ij}$ is individual $i$'s consumption of variety $j$, $Q_j = \{q_{1j}, q_{2j}, ..., q_{Nj}\}$ is a vector of quantities consumed by all other individuals, and $n$ is the mass of varieties available to consumers. The parameter $\rho_i$ is a measure of substitutability across product varieties, and is allowed to vary across individuals.

The function $R_i(.)$ provides a reduced-form way of capturing the dependence of a consumer's consumption utility on the consumption of others through peer effects. $R_i(.)$ can be individual-specific, so that the quantity consumed by others enter differentially into different consumers' utility; in other words, different individuals are allowed to be differentially affected by the consumption of their peers.

One channel through which peers' consumption can influence an individual's own utility from a particular good is a desire of people to consume similar types of products as others. This could come about because of the proverbial "keeping up with the Joneses," or because of network externalities that make a particular good more useful if others are also using it (e.g., being on the iOS operating system is more useful to me if my friends are also on iOS, since we can then communicate using Facetime). In this case, the function $R$ would be positive and increasing in $q_{-ij}$, ensuring that individual $i$'s utility for consuming a product is higher when the product is also widely consumed by others.

An alternative channel through which an individual's own consumption of goods may be affected by the consumption of others is through social learning. For instance, suppose that a product only enters the consumer's choice set if they learn about the existence of the item from friends that already bought the item. In this case, we could specify $R$ to reflect the probability that the individual learns about the item from their friends, where $R$ is an increasing function of the purchases made by friends. At the extreme, if the individual has one friend and she only learns about the item if the friend purchased the item, then $R(q_{-i,j}) = 1$ when $q_{-i,j} > 0$, and $R(q_{-i,j}) = 0$ otherwise.

**Parametric example.** For the purposes of illustrating the implications of peer effects for consumption and markups, consider the following parametric form:

$$R_i(Q_j) = \left( \int_{F(i)} \psi_f \times q_{fj} df \right)^{\eta_i}, \tag{A.2}$$

[6] As discussed in the paper, it is possible that peer effects can also influence prices and markups through other channels. In that case, the overall effect will depend on the relative strength of the forces discussed here, and any additional mechanisms.

so that the utility for individual $i$ is given by

$$U_i = \left( \int_0^n q_{ij}^{\rho_i} \left( \int_{F(i)} \psi_f \times q_{fj} df \right)^{\eta_i} dj \right)^{\frac{1}{\rho_i}}, \tag{A.3}$$

where $F(i)$ denotes the set of peers that influence individuals $i$, and $q_{fj}$ denotes the quantity of variety $j$ consumed by individual $i$. The two scalar variables $\eta_i$ and $\psi_f$ affect how sensitive individual $i$'s consumption is to the consumption of her peers. $\eta_i$ reflects how susceptible individual $i$ is to the peer effects exerted by others. When $\eta_i = 0$, then there are no peer effects from others and we have the usual CES utility function. When $\eta_i > 0$, then an individual's utility is dependent on the consumption of their peers. The scalar $\psi_f$ reflects how influential individual $f$'s consumption for the utility and consumption of $f$'s friends. Larger values for $\psi_f$ mean the consumption of individual $f$ has a greater impact on the utility of her friends.

Together, $\eta_i$ and $\psi_f$ affect the overall magnitude of the peer effects in driving individual $i$'s consumption. We can see this in the elasticity of expression A.2 with respect to quantity consumed by friend $f$:

$$\frac{\partial \ln R_i(Q_j)}{\partial \ln q_{fj}} = \eta_i \tilde{\psi}_f, \tag{A.4}$$

where

$$\tilde{\psi}_{fj} = \frac{\psi_f q_{fj}}{\int_{F(i)} \psi_g q_{gj} dg}. \tag{A.5}$$

The higher the value of $\eta_i$ and $\psi_f$, the more sensitive is individual $i$'s consumption of variety $j$ to changes in peer $f$'s consumption of variety $j$.

We next derive the analytical solutions for prices and markups using the parametric specification in equation A.2. However, the implications for elasticities of consumption and markups hold for various forms of $R_i(Q_j)$ that have the property where $R_i(Q_j)$ is weakly increasing in $Q_j$.

## B Consumer Demand

Consumer $i$'s constrained maximization problem is given by

$$\max_{q_{ij}} U_i^{\rho_i} - \lambda_i \left( \int_0^n p_j q_{ij} dj - I_i \right),$$

where $\lambda_i$ is the multiplier on the budget constraint, and $I_i$ denotes total income. From the consumer's maximization problem, the Frisch demand function is given by

$$q_{ij} = \left( \frac{\lambda_i p_j}{\rho_i \left( \int_{F(i)} \psi_f \times q_{fj} df \right)^{\eta_i \rho_i}} \right)^{\frac{1}{\rho_i - 1}}. \tag{A.6}$$

The relative demand for varieties $j$ and $k$ is given by

$$\frac{q_{ij}}{q_{ik}} = \left( \frac{p_j}{p_k} \frac{\left(\int_{F(i)} \psi_f \times q_{fk} df\right)^{\eta_i \rho_i}}{\left(\int_{F(i)} \psi_f \times q_{fj} df\right)^{\eta_i \rho_i}} \right)^{\frac{1}{\rho_i - 1}}. \tag{A.7}$$

Let $\sigma_i \equiv \frac{1}{1-\rho_i}$, which is the elasticity of substitution when $\eta_i = 0$. Hence,

$$q_{ij} = q_{ik} \left( \frac{p_j}{p_k} \right)^{-\sigma_i} \left( \frac{\int_{F(i)} \psi_f \times q_{fk} df}{\int_{F(i)} \psi_f \times q_{fj} df} \right)^{\eta_i (1-\sigma_i)}. \tag{A.8}$$

Multiplying both sides by $p_j$ and integrating with respect to all varieties yields

$$\int_0^n p_j q_{ij} dj = \int_0^n q_{ik} p_j^{1-\sigma_i} p_k^{\sigma_i} \left( \frac{\int_{F(i)} \psi_f \times q_{fk} df}{\int_{F(i)} \psi_f \times q_{fj} df} \right)^{\eta_i (1-\sigma_i)} dj.$$

The left-hand side is total expenditure, which we assume equals total income $I_i$. Hence, we can write the individual consumer's demand for variety $k$ as

$$q_{ik} = \frac{p_k^{-\sigma_i} \left(\int_{F(i)} \psi_f \times q_{fk} df\right)^{\eta_i(\sigma_i - 1)}}{\int_0^n p_j^{1-\sigma_i} \left(\int_{F(i)} \psi_f \times q_{fj} df\right)^{\eta_i(\sigma_i - 1)} dj} I_i. \tag{A.9}$$

### Elasticity of Consumer Demand

The individual consumer's demand for variety $k$ (and the individual's price elasticity of demand) now depends on the consumption of others. From equation A.8, we can compute the elasticity of an individual consumer $i$'s demand for variety $k$ relative to the demand of others (e.g., friend $f$). This is given by

$$\frac{\partial \ln q_{ik}}{\partial \ln q_{fk}} = (\sigma_i - 1)(1 - s_{ik}) \eta_i \tilde{\psi}_{fk} \tag{A.10}$$

where $q_{ik}$ is the quantity consumed by individual $i$, $q_{ik}$ is the quantity consumed by individual $f$, $Q_j = \{q_{1j}, q_{2j}, ..., q_{Nj}\}$ is a vector of quantities consumed across individuals, and $s_{ik}$ is the expenditure share of variety $k$ for consumer $i$.

As discussed above, the scalar $\eta_i$ reflects the susceptibility of individual $i$ to peer effects, and $\tilde{\psi}_{fj}$ is the influence of friend $f$ defined in equation A.5. The larger is $\eta_i$, the more sensitive is the individual's demand for variety $j$ to the consumption of others; similarly, increases in consumption by more influential friends (high-$\tilde{\psi}_{fj}$ friends) have a larger effect on own demand for variety $j$. When $\eta_i = 0$, the individual's demand does not depend on consumption of others.

### Price elasticity of demand

The price elasticity of consumer $i$'s demand for variety $k$ is given by:[7]

$$\frac{-\partial \ln q_{ik}}{\partial \ln p_k} = \sigma_i + (\sigma_i - 1)\eta_i \int_{F(i)} \tilde{\psi}_{fk} \times \frac{-\partial \ln q_{fk}}{\partial \ln p_k} df. \tag{A.11}$$

Equation A.11 highlights that, for the standard case with $\sigma > 1$, peer effects lead individuals to have larger price elasticities than in the benchmark model (which only includes the first term). In addition, the equation shows that individuals can have different price elasticities of demand for at least three reasons.

1. They may have different elasticities of substitution between varieties, i.e. heterogeneity in $\sigma_i$. The larger $\sigma_i$, the more responsive is consumer $i$'s demand to price changes. This effect, which corresponds to the first term in equation A.11, exists even in the absence of peer effects.

2. Individuals can have different price elasticities because their utility is differently affected by the consumption of others. In other words, individuals may be differentially susceptible to peer effects, generating heterogeneity in $\eta_i$. The larger is $\eta_i$, the more price elastic is the demand of individual $i$, all else equal, because the increase in consumption of their peers in response to a given price drop has a larger effect on their own utility (and thus demand).

3. Individuals may also have heterogeneous price elasticities of demand because their set of friends is different in terms of both their influence and their price elasticities. Specifically, peer effects can lead to higher price elasticity of demand when (i) a person's friends are more influential (the first term within the integral, $\tilde{\psi}_{fk}$), and (ii) for a given $\tilde{\psi}_{fk}$, the consumption of peers is more price sensitive (the second term within the integral, $\frac{-\partial \ln q_{fk}}{\partial \ln p_k}$). Putting the two forces together, we can see that there is a larger price elasticity of demand when each of these two terms is higher, as well as when there is a higher covariance between the peer influence of a friend and the price sensitivity of the friend. Our empirical micro estimates are informative about the magnitudes of the covariance between the peer influence of a friend and the price elasticity of that friend.

## C Firm's problem and Price Setting

Production of a good involves a fixed cost $F$ in addition to a constant marginal cost $c$, so the average cost is decreasing in quantity. To keep things simple, there are no economies of scope to producing multiple varieties. Therefore, there is a continuum of firms, where each firm produces one variety and there is one firm per variety. A firm that produces variety $k$ has profits

$$\pi_k = p_k x_k - c x_k - F,$$

[7] This derivation is based on the assumption that there is a continuum of firms and varieties produced, and the function $R$ is bounded. Therefore, $\frac{\partial \int_0^n p_j^{1-\sigma_i} R_i(Q_j)^{(\sigma_i - 1)} dj}{\partial p_k} \approx 0$.

where $x_k$ denotes the quantity of variety $k$ produced. The firm sets prices $p_k$ to maximize profits. Solving the firm's problem yields:

$$p_k = c + \frac{-x_k}{\partial x_k / \partial p_k}. \tag{A.12}$$

In equilibrium, the free entry condition implies zero profits. Hence, $x_k = F/(p_k - c)$. The market clearing condition is given by $x_k = \int_i q_{ik} di$, where $q_{ik}$ is the quantity of variety $k$ consumed by individual $i$. From the market clearing condition, we can derive the following partial derivative expression:

$$\frac{-\partial x_k}{\partial p_k}\frac{1}{x_k} = \int_i \frac{-\partial q_{ik}}{\partial p_k} di \times \frac{1}{\int_f q_{fk} df} = \frac{1}{p_k}\int_i \frac{-\partial \ln q_{ik}}{\partial \ln p_k} w_{ik} di, \tag{A.13}$$

where $w_{ik} = \frac{q_{ik}}{\int_f q_{fk} df}$, and the price elasticity of demand $\frac{\partial \ln q_{ik}}{\partial \ln p_k}$ was previously derived in equation A.10.

Equation A.13 can be interpreted as the weighted-average aggregate price elasticity of demand, where the individual price elasticities of demand are weighted by their quantity of consumption relative to total aggregate consumption. Substituting A.13 into the firm's first-order conditions gives the following markup expression:

$$\frac{p_k}{c} = \frac{1}{1 - \theta_k}, \tag{A.14}$$

where

$$1/\theta_k \equiv \int_i \left( \sigma_i + (\sigma_i - 1)\eta_i \int_{F(i)} \tilde{\psi}_{fk} \times \frac{-\partial \ln q_{fk}}{\partial \ln p_k} df \right) w_{ik} di \tag{A.15}$$

Notice that when there are no consumption peer effects, i.e. $\eta_i = 0$, then the markup equals $\frac{\sigma_k}{\sigma_k - 1}$, where $\sigma_k \equiv \int_i \sigma_i w_{ik} di$. This corresponds to the markup expression in the usual monopolistic competition Dixit-Stiglitz model, allowing for heterogeneity in $\sigma$ across individuals.

In the presence of consumption peer effects, where $\eta_i > 0$, markups are strictly less than $\frac{\sigma_k}{\sigma_k - 1}$. Peer effects are stronger (and markups are smaller) when individuals' utilities of a variety is more sensitive to the consumption of others (large $\eta_i$). Markups are also smaller when price changes lead to larger changes in the amount of peer influence exerted, either because each friend is more influential, or because each friend is more price sensitive (and therefore more likely to buy and exert influence), or because the most price sensitive friends are the most influential ones. Equation A.15 also highlights another covariance that determines the overall influence of peer effects on markups: the correlation between an individuals' own susceptibility to peer influence ($\eta_i$) and the overall peer effect exerted on this individual, which depends on which people are friends with that individual ($F_i$): markups are particularly low if the most susceptible individuals are friends with the most influential individuals.[8]

Equation A.14 has also has implications for the dispersion of markups across firms and over time. In particular, it shows that markups can vary across firms because of differences in the strength of peer effects for different products produced by firms. Indeed, our estimates imply that the strength of peer effects do vary significantly across different mobile phone brands, and there are likely even stronger heterogeneities across different product categories that could contribute to markup dispersion. In related work, we explore this further (see Kuchler, Stroebel and Wong, 2021)

[8] In our setting, we did not find a large amount of heterogeneity in susceptibility to peer influence across demographic groups, but these parameters will differ in other settings.

## A.6 Phone Model-Level Analysis

In this appendix, we explore peer effects at the model level. Specifically, we analyze whether having a friend buy an iPhone 6s primarily increases a person's own probability of also purchasing an iPhone 6s, or whether it increases the individual's probability of purchasing an iPhone in general. To do so, we explore purchases of each of the 20 most common models that individuals switch to in our sample, in addition to three residual categories. The full set of phones $P$ that we study includes iPhone 6s, iPhone 6s Plus, iPhone SE, iPhone 6, iPhone 5s, iPhone 6 Plus, Other iPhone, Galaxy S7, Galaxy S7 Edge, Galaxy Core Prime, Galaxy Note 5, Galaxy J7, Galaxy Grand Prime, Galaxy S5, Galaxy S6, Other Galaxy, LG Tribute 5, LG K10, LG G5, LG G Stylo, HTC Desire 626s, Alcatel One Touch, and Other.

For this analysis, we cannot use an instrumental variables research design, as we do in the main body of the paper. In particular, while observable characteristics allow us to predict relatively well whether a given individual would purchase an iPhone or a Galaxy, it is much harder to predict whether an individual would buy an iPhone 6 or an iPhone 6s, and the resulting estimated probabilities are highly collinear.[9] Instead of an instrumental variables specification, we therefore run the following OLS regressions for each model $p' \in P$:

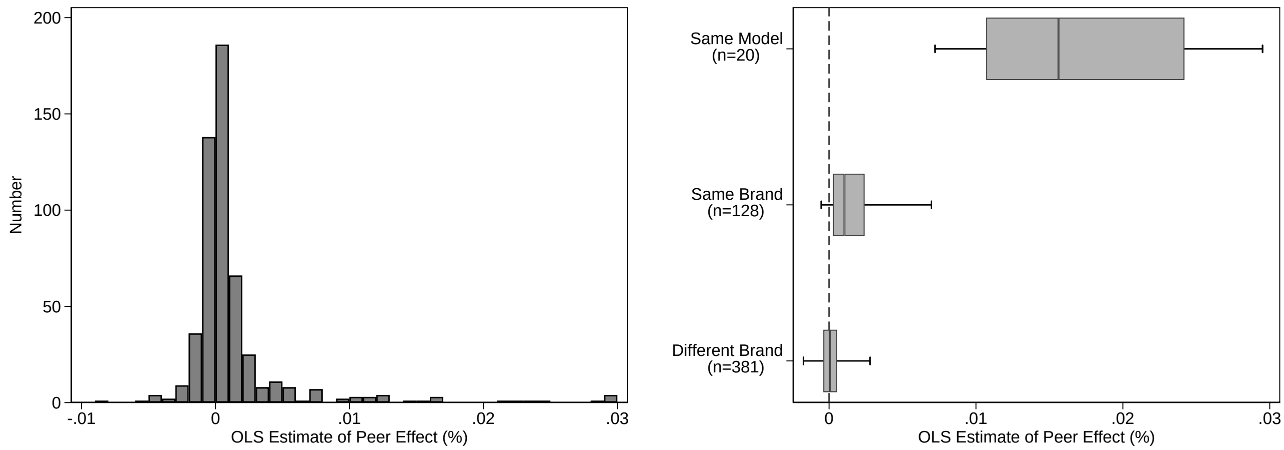
$$\begin{aligned} \mathbb{1}(BuysPhone)_{i,t}^{p'} = & \textstyle\sum_{p\in P} \beta_p^{p'} FriendsBuyPhone_{i,t-1}^{p} + \\ & \textstyle\sum_{p\in P} \pi_p^{p'} ProbBuyUncond_{i,t}^{p} + \\ & \textstyle\sum_{p\in P} \xi_p^{p'} AllFriendsAvgProbBuyUncond_{i,t}^{p} + \gamma X_{i,t} + \epsilon_{i,t} \end{aligned} \tag{A.16}$$

The central controls in these specifications are the unconditional probabilities of purchasing each phone $p \in P$ of person $i$, as well as the averages of these probabilities across all of person $i$'s friends. The vector $X_{i,t}$ includes additional controls analogous to those in Equation 11. The coefficients of interest are the parameters $\beta_p^{p'}$, of which there are at total of $|P| \times |P| = 529$. The similarity of the OLS and IV estimates in Section III suggests that biased coefficients due to correlated shocks may not be a first-order concern for the analysis, especially after conditioning on time fixed effects interacted with our large set of control variables. In addition, even if these OLS estimates were biased on average, patterns across these coefficients can still be informative about the nature of the underlying peer effects—especially since many common shocks should affect these estimates in similar ways.

The left panel of Figure A.6 shows a histogram of the estimated $\beta_p^{p'}$ coefficients. About 64% of the estimated coefficients are positive. The right panel shows the distributions of $\beta_p^{p'}$ coefficients for three groups. The first group includes the 20 same-model coefficients that capture the effects of an individual buying a specific model on the probability of her friends buying the same model. The second group corresponds to same-brand, different-model peer effects. They capture, for example, the effect of an individual buying a Galaxy S7 on his friends' probability of purchasing a different Galaxy model, such as a Galaxy Note 5. The final group includes all peer effects to models from a different brand.

[9] In Appendix A.2, we describe how we estimate the unconditional probabilities that individuals purchase a phone of each type $p$ in a given week, $ProbBuyUncond_{i,t}^{p}$. Many of these probabilities are highly correlated within individuals. For example, the predicted propensities to purchase an iPhone 6s and an iPhone 6 have a correlation of 0.89, and the predicted propensities to purchase a Galaxy J7 and a Galaxy S5 have a correlation of 0.90.

**Figure A.6:** Specific Model Peer Effects (OLS)

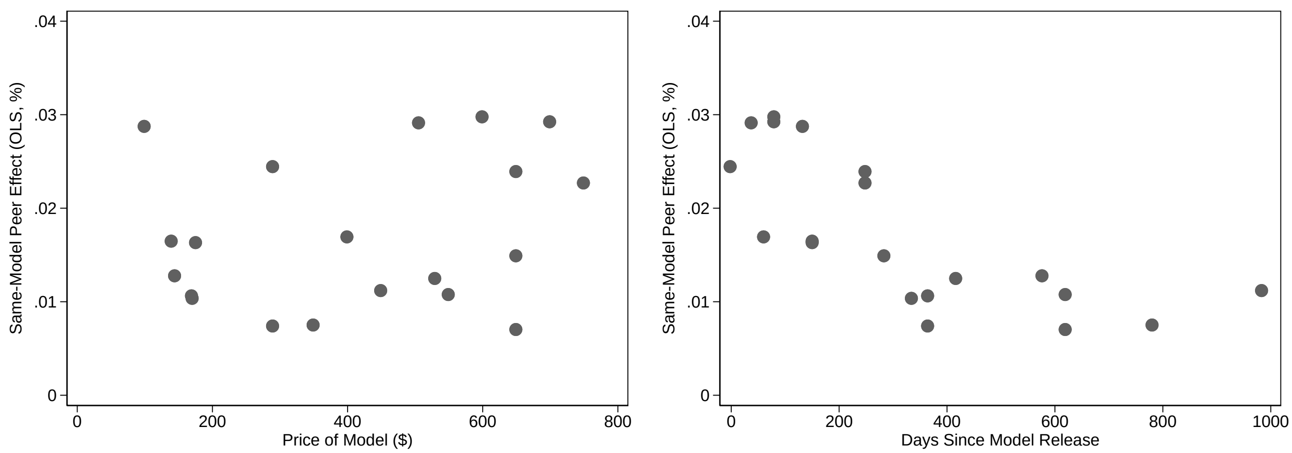



**Note:** Figure shows the distributions of $\beta_p^{p'}$ coefficients from regression A.16. The left panel shows a histogram of the 529 coefficients. The right panel splits these up into three groups. The top bar shows the coefficients when $p$ and $p'$ correspond to the same model; the middle bar corresponds to the coefficients when $p$ and $p'$ correspond to different models of the same brand, and the bottom bar corresponds to the coefficients when $p$ and $p'$ correspond to different brands. We include the coefficients for "Other iPhone" to "Other iPhone" and "Other Galaxy" to "Other Galaxy" in the "Same Brand" category. We include the "Other" to "Other" coefficient in the "Other Brand" category, even though some of these peer effects could still correspond to "Same Model" or "Same Brand" purchases for phones that were so uncommon that they were not split out independently. Among the 20 precise models we split out, there are 3 unique brands with more than one model: iPhone, Galaxy, and LG. The box plots show the $5^{\text{th}}$, $25^{\text{th}}$, $50^{\text{th}}$, $75^{\text{th}}$ and $95^{\text{th}}$ percentiles of the distribution.

By far the largest peer effects are concentrated within the same model. Indeed, the smallest same-model peer effect we estimate is larger than the largest different-model peer effect. There is also substantial heterogeneity across the estimated same-model peer effects. In Figure A.7, we plot the 20 same-model $\beta_p^{p'}$ coefficients against the market price of the model during our sample period (left panel), and against the time since the market introduction (right panel). There is no correlation between model price and the estimated same-model peer effect. In other words, low-end models such as the LG Tribute 5 have similarly sized same-model peer effects as more upscale models such as the iPhone 6s Plus. On the other hand, we find that same-model peer effects are substantially larger for more recent models, regardless of the price of these models. As before, these patterns point towards social learning as an important driver for the estimated peer effects (the importance of which should decline as the model becomes more well-known over time). The evidence for a "keeping up" effect, which would likely be more important for more expensive phones, is more limited, though, as we have discussed, we cannot rule out that it also contributes to the overall peer effects.

Figure A.6 also shows that same-brand, different-model peer effects are almost three times as large as different-brand peer effects, with an average $\beta_p^{p'}$ coefficient of 0.0036 vs. 0.0013. To further explore the same-brand, different-model peer effects, we split them up according to the three major brands in our data. The left panel of Figure A.8 shows that these same-brand, different-model peer effects are largest among the iPhones in our sample and smallest among the LG phones. This finding is consistent with the relatively independent conduct of the marketing campaigns for the LG models (and the model names not indicating any relationship between phones), while both Apple and Samsung tended to

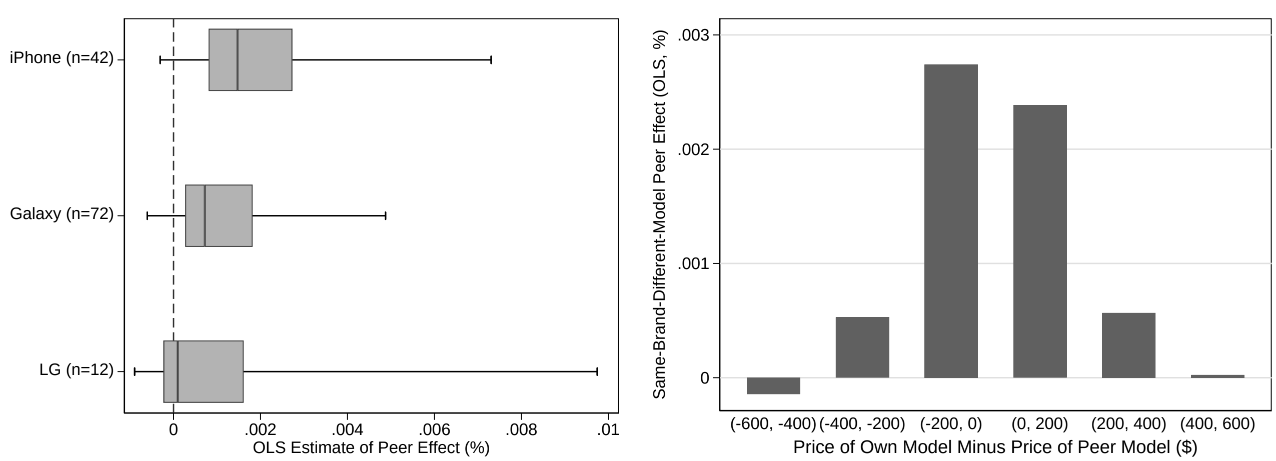
**Figure A.7:** Same-Model Peer Effects




**Note:** Figure shows scatterplots of the 20 same-model $\beta_p^{p'}$ coefficients from regression A.16 against the price of the model (left panel) and the time since the model's release (right panel).

jointly market their entire range of phones under a common brand identity. These results suggest that umbrella branding campaigns of different phone models can generate valuable same-brand spillovers through peer effects (see Erdem, 1998, for academic analyses of the spillovers of marketing activities due to umbrella branding). The results also highlight additional benefits resulting from line extension strategies beyond the direct effects of advertising spillovers documented by Balachander and Ghose (2003).

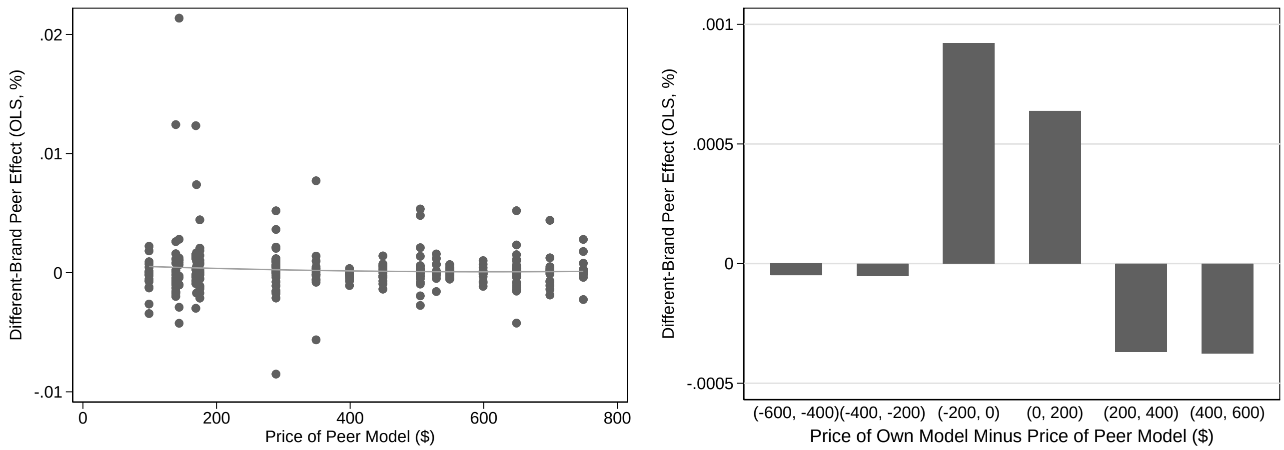
**Figure A.8:** Same-Brand, Different-Model Peer Effects




**Note:** Figure shows the same-brand, different-model $\beta_p^{p'}$ coefficients from regression A.16. The left panel shows the distribution of the effects separately by the three main brands in our sample. The box plots show the 5$^{th}$, 25$^{th}$, 50$^{th}$, 75$^{th}$, and 95$^{th}$ percentiles of the distribution. In the right panel, we form groups on the basis of the price difference between the models (i.e., $Price(p') - Price(p)$), and plot the average OLS coefficient on the basis of that price difference. Positive numbers capture the peer effects of a friend buying a cheaper model on a person's probability of buying a more expensive model.

The right panel of Figure A.8 shows the same-brand, different-model peer effects split out by the price difference between the two models. Positive numbers on the horizontal axis capture the peer effects

**Figure A.9:** Different-Brand Peer Effects




**Note:** Figure shows the different-brand $\beta_p^{p'}$ coefficients from regression A.16. The left panel shows the distribution of the effects separately by the price of the peer model including a quadratic fit line. In the right panel, we form groups on the basis of the price difference between the models, $Price(p') - Price(p)$, and plot the average OLS coefficient on the basis of their price difference. Positive numbers capture the peer effects of a friend buying a cheaper model on a person's probability of buying a more expensive model.

of a friend buying a cheaper model on a person's probability of buying a more expensive model. We find that peer effects are larger for similarly-priced models than they are for models that are either substantially more expensive or substantially cheaper. Importantly, this finding is not just the result of an individual and her friends having similar incomes or being similarly old. Indeed, in all regressions, we directly control for individuals' estimated unconditional probability of purchasing a certain phone model, in addition to the average of these probabilities across their friends. These findings are consistent with evidence from the marketing literature that across-product spillovers decrease in magnitude as products become more dissimilar (e.g., Janakiraman, Sismeiro and Dutta, 2009).

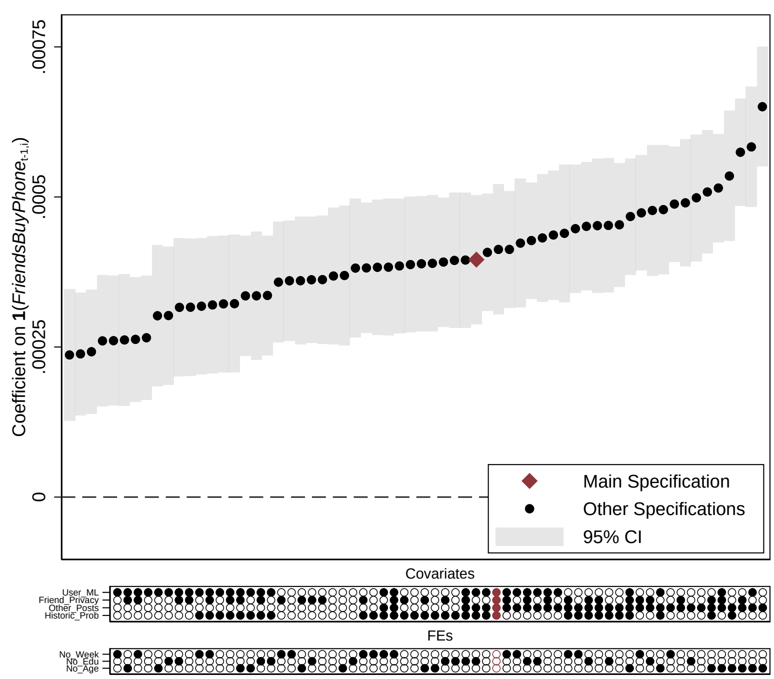
In the last part of the analysis, we split out different-brand peer effects. The left panel of Figure A.9 shows that these peer effects do not vary, on average, with the price of the models. The right panel shows that most of these different-brand peer effects are concentrated on different brand models in a similar price range as the phone purchased by the peer. This is consistent with the patterns and associated interpretations for same-brand peer effects documented above.

**Summary of Model-Level Findings.** A number of take-aways result from our model-level analysis. First, same-model peer effects are more than an order of magnitude larger than different-model peer effects. Second, these same-model peer effects do not vary with the cost of the model, but rather they are decreasing in the time since the model release, providing additional evidence for an important social learning channel behind the peer effects. Third, same-brand, different-model peer effects are more than twice as large as different-brand peer effects. These effects are largest for brands that co-brand their devices and are also largest within each brand for models of similar value. Finally, across-brand peer effects do not vary with model value, and are largest for models of similar value.

## A.7 Robustness Checks – Specifications

In our baseline regressions, we include a number of controls and fixed effects, which we describe in detail in Section C. In this appendix, we show that our results are robust to changes in the set of controls and fixed effects we incorporate. In Figure A.10, we present the estimated $\beta$-coefficients for 64 different variants of our baseline regression 6. Each variant is depicted as a dot, with the 95% confidence intervals shaded in gray. The red diamond corresponds to our baseline specification. In the panel below the graph, we indicate which fixed effects and controls are included in each regression. We consider three modifications to the set of fixed effects. In the regressions marked *No_Week*, we omit the week interaction term from each of the three sets of fixed effects. Correspondingly, in the regressions marked *No_Edu*, we remove the education fixed effect, while in the regressions marked *No_Age*, the age fixed effect is omitted. For each of the four variants of fixed effects (our baseline plus three adjustments), we run a number of permutations, excluding some of our baseline set of controls. In the regressions marked *User_ML*, we control for if a user posts about breaking their own phone. In the regressions marked *Friend_Privacy*, we control for the number of a user's friends whose posts are public by default. In the regressions marked *Other_Posts*, we control for the number of a user's friends who post about anything other than a broken phone. In the regressions marked *Historic_Prob*, we control for the number of a user's friends who broke their phone in the 12 months prior to our sample, or for the conditional probability of buying a new phone among these users. In our baseline specification, we control for all of these variables, but our results are largely similar if we only control for a subset of them or if we adjust the fixed effects.

**Figure A.10:** Robustness of Alternative Specifications of Controls and Fixed Effects




**Note:** Figure compares estimates of the $\beta$-coefficients for 64 different variants of our baseline regression 6. Our baseline specification is marked with a diamond; the panel below the graph indicates the specification corresponding to each