The Surprising Breadth of Harbingers of Failure

Duncan Simester, Catherine Tucker and Clair Yang
MIT

March 2017

PRELIMINARY DRAFT – Please do not circulate

Previous research has shown that there exist “harbinger customers,” who systematically purchase new products that fail (retailers discontinue them). We extend this finding in two ways. First, we show that there are also some zip codes that are “harbinger zip codes”. If households in these zip codes adopt a new product this is a signal that the new product will fail. Second, we show that households in these zip codes make choices that are systematically different from other households across a surprising array of decisions.

We identify harbinger zip codes using purchases of new products from one retailer. We then analyze purchases from these zip codes at a different retailer. We show that the harbinger zip codes identified at the first retailer purchase products from the second retailer that other zip codes are less likely to purchase. They also purchase products that receive less favorable customer reviews and that other customers are more likely to return.

We next use the harbinger zip codes identified at the first retailer to compare donations to congressional election candidates. Households in harbinger zip codes donate to different candidates than households in neighboring zip codes, and they donate to congressional election candidates that are less likely to win. We also compare changes in house prices between 2002 and 2015. House prices in harbinger zip codes increased at slower rates than other zip codes. All of our results hold even when we compare five-digit zip codes with neighboring zip codes that share the same 3-digit zip code.

By studying households that change zip codes we are able to investigate whether harbinger zip codes result from households with harbinger preferences choosing to cluster together, or households learning these preferences by observing their neighbors. The evidence strongly supports the first explanation. When households move from a harbinger zip code they tend to move to another harbinger zip code, while non harbingers do the reverse. However, we find no evidence that households’ preferences change after they move. It appears that harbinger preferences are a sticky trait, and the harbinger zip code effect is more due to where customers choose to live, rather than households influencing the preferences of their neighbors.

Keywords: preference heterogeneity, new product development, real estate prices, campaign donations

1. Introduction

Presumably someone liked Diet Crystal Pepsi. Unfortunately for Pepsi there were not enough of these people and so the product flopped. What is surprising is that the customers who purchased Diet Crystal Pepsi may also have liked other new products that flopped, like Colgate Kitchen Entrees (a range of frozen meals). Anderson, Lin, Simester and Tucker (2015) documented the existence of these customers, who they labelled "harbingers of failure." Harbingers are more likely to purchase products that other customers do not buy. A purchase by these customers indicates that the product appeals to a narrow slice of the marketplace, and is a signal that a new product will fail.

We extend this finding in two ways. First we show that there are not just customers who are harbingers, there are also harbinger zip codes. Purchases of new products by households in these zip codes is a signal that the new product will fail. Identifying harbingers at the zip code level solves an important problem. Firms often lack enough information to distinguish individual harbingers from other customers. Identifying which zip code a customer is from is generally much easier than identifying systematic differences in individual customer decisions.

We highlight the value of identifying harbingers at the zip code level by comparing a wide range of household decisions. To do so we identify harbinger zip codes using new product purchases from a mass merchandise retailer. We show that households in these zip codes purchase new products that tend to fail, and also purchase niche existing products that neighboring zip codes are less likely to purchase.

We next show that the same pattern extends to purchasing decisions at another retailer selling private label apparel. Zip codes identified as harbingers at the mass merchandise store are also harbingers at the apparel retailer. They purchase products that are relatively unpopular. They also purchase products that receive unfavorable reviews and products that other customers are more likely to return.

Our next comparison focuses on contributions to congressional election campaigns. We identify the top two candidates and compare donations in each 5-digit zip code with donations in neighboring zip codes. Zip codes identified as harbingers at the mass merchandise store are systematically less likely to donate to the most popular candidates. They are also more likely to donate to candidates that lose their elections. We also compare changes in house prices between 2002 and 2015. Zip codes identified as harbingers at the mass merchandise store had systematically smaller increases in their house prices during periods that prices were increasing for neighboring zip codes. Collectively, these results reveal that preferences are correlated across a wide range of decisions, including purchasing decisions, political decisions, and housing decisions.

Data from the apparel retailer also allows us to observe households moving between zip codes. This allows us to ask whether the harbinger zip code effect results from households with harbinger preferences clustering together, or whether households learn these preferences by observing their neighbors. We first ask whether households who move from harbinger zip codes tend to move to other harbinger zip codes. The evidence strongly supports this pattern. Households who leave harbinger zip codes tend to go to other harbinger zip codes. The reverse is true for households that start in non-harbinger zip codes; they tend to move to other non-harbinger zip codes. We then ask whether households' preferences change when they move, which might suggest that households' preferences are in part learned from their new neighbors. We find no support for this; there is little evidence that preferences change when households move. We conclude that harbinger preferences are a relatively sticky trait, and the clustering of harbingers within harbinger zip codes appears to be caused more by the preferences households bring than the preferences that they learn.

The Harbinger Effect

Anderson et al. (2015) were the first to document the existence of customers who systematically purchase new products that fail. Their findings were based upon purchases of new products at a chain of convenience stores that sell consumer packaged goods. They divided the new products into two subsets and used the outcomes in the first subset to classify customers as harbingers (or not). They then showed that customers who purchased flops in the first subset of products (the harbingers) were also systematically more likely to purchase flops in the second subset of products. Adoption of a new product by these harbingers was a strong signal that a new product would fail.

Anderson et al. (2015) argue that in the setting they study it is very unlikely that the finding is due to observational learning or information spillovers between customers. Instead they attribute the finding to harbingers having unusual preference that are not representative of other customers. Adoption by the harbingers is a signal that other customers are less likely to adopt the product, which leads to product failure.

The original Anderson et al. (2015) study has since been replicated and extended in two recent working papers. Anderson, Chen and McShane (2017) replicate the findings using an IRI panel dataset including over 100,000 consumers and 400 retailers. Anderson, Chen, Liu and Simester (2017) demonstrate the existence of "harbinger products". While purchases of harbinger customers signal that a new product will fail, harbinger products represent the inverse. Purchases of harbinger products on a customer's first shopping visit signals that the new customer will fail (not return to that retailer).

Perhaps the most surprising aspect of the harbinger customer effect is that the signal extends across consumer packaged goods (CPG) categories. Customers who purchase new oral care products that flop also tend to purchase new haircare products that flop. Anderson et al. (2015) interpret their findings as evidence that customers who have unusual preferences in one product category also tend to have unusual preferences in other categories. In other words, the customers who liked Diet Crystal Pepsi also tended to like Colgate Kitchen Entrees (which also flopped).

If the most surprising aspect of the original study is that the harbinger effect extends across product categories, then the results in this paper magnify that surprise. Anderson et al. (2015) show that the effect extends across CPG categories within a single retailer. We show that the effect extends across different retailers, beyond consumer packaged goods, and beyond purchasing decisions. Customers who purchase products that fail at a mass merchandise retailer, also purchase less popular items at an apparel retailer, support less popular congressional election candidates, and choose to live in zip codes that have smaller house price increases.

Other Related Literature

The harbinger effect is closely related to the preference minority and lead user literatures. The preference minority literature studies customers with unusual preferences. These customers have been used to explain the growth of Internet sales in some product categories. If bricks and mortar stores allocate shelf space according to the preferences that are most typical in their markets, then preference minorities may not find products that fit their preferences (Anderson 1979; Waldfogel 2003). This helps to explain why preference minorities are more likely to purchase from the Internet, and are less price-sensitive when doing so (see Choi and Bell 2011 and Brynjolfsson, Hu and Rahman 2009). A related explanation has also been used to explain why we see a long tail of purchases on niche items on the Internet (see Brynjolfsson, Hu and Smith 2003; Brynjolfsson, Hu and Simester 2011).

Harbingers can be interpreted as the opposite of "lead users". Lead users are customers whose preferences are more likely to identify "breakthrough" ideas (von Hippel 1986). Adoption of new products by these lead users is a signal that the product is more likely to succeed. This is the opposite signal that we should infer from the adoption of new products by harbingers. Notice that the existence of lead users is consistent with a straw man argument that greater adoption by any customer is an indication that a new product will succeed. The signal associated with harbingers is reversed and therefore must overcome this standard relationship. It is this aspect of the harbinger effect that makes it counter-intuitive.

In the product development literature, differences in customer preferences have previously been used to explain why some products have initial success, but then fail to grow over the long term. For example, Geoffrey Moore's (1991) argues that a reason new technology products fail to "cross the chasm" is that the early adopters often have different preferences than the mainstream market and so early success may not signal future success. Van den Bulte and Joshi (2007) investigate this explanation and propose a model with segments of "influentials" and "imitators." They show that this can lead to the diffusion of new products slowing in the middle of the diffusion curve.

In the original harbingers paper Anderson et al. (2015) review this explanation for their findings but warn that influential customers are less relevant in the CPG setting that they study. In this study, the conspicuous nature of apparel consumption is potentially consistent with some customers influencing the purchasing decisions of others. Our analysis of congressional election donations could also support an explanation in which some households influence the decisions of other households. However, our investigation of apparel customers that changed zip codes suggests that households generally do not change their preferences when they change zip codes. When these apparel customers moved to a different zip code they tended to bring their preferences with them rather than learning their preferences after they arrived.

To our knowledge, this is the first paper to link retail spending behavior to housing choices and political choices. In this respect, our work is related to other work that has explored the link between political events and retail outcomes (see for example Pandya and Venkatesan 2016).

The remainder of the paper is organized as follows. Section 3 describes the data we use to identify harbinger zip codes together with the other datasets that we match to this. Section 4 identifies harbinger zip codes using data from a mass merchandise store. In Section 5 we use this identification from the mass merchandise store to compare purchases of apparel from a private label retailer. These findings demonstrate that the harbinger effect persists across retailers. In Section 6 we investigate the extent to which harbinger zip codes explain variation in donations to political campaigns. Section 7 presents evidence that the effect also extends to changes in house prices. The paper concludes in Section 8.

2. Data and Initial Results

This paper uses multiple datasets. We start by describing the data that we use to identify harbinger zip codes and then briefly describe the other retail, housing and political datasets used in the paper.

To identify harbinger zip codes we use data provided by a mass merchandise store. For confidentiality reasons we cannot identify which store but for ease of exposition we will refer to it as "MassStore". The retailer sells a broad range of products including perishables, sundries, and durables. When customers purchase at this retailer they must identify themselves with their membership card, which allows us to link each transaction to each customer. We use a complete sample of individual customer transactions (for every customer) between January 2013 and July 2016. Each line item in the data identifies a purchase of a unique item by a customer on a unique purchase occasion. The data identifies the customer, store, date and time of purchase, number of units purchased and price paid.

MassStore also provided a second set containing demographic data for each customer together with the 5-digit zip code from each customer's residential address (this address is provided by customers when registering for a membership). We augment this dataset by obtaining 2010 census data that classifies zip codes according to the proportion of households that are urban and rural.[1] The urban households are further divided into those in urbanized areas (very urban), and those in urban clusters (less urban). We use this augmented data to compare the demographic characteristics of harbinger zip codes with other zip codes. Definitions for all of these variables, together with summary statistics, are provided in the Appendix.

The classification of zip codes using the MassStore zip code data is then used to compare purchases from a private label apparel retailer. A confidentiality agreement again prevents us from identifying which retailer provided this data. For ease of exposition we will refer to it as "ApparelCo". The retailer sells through its own dedicated retailer stores, its own catalog stores and its own Internet site. It rarely sells products that do not carry its own brand, and its products are not sold by other retailers (with very few exceptions). We use two datasets provided by this retailer. The first dataset is a complete record of all purchases made through the retailer's catalog and Internet channels. We exclude store transactions as it is not always possible to identify the residential zip code of the customer making the purchase.[2] Like MassStore's transaction data, each line item in the data identifies a purchase of a unique item by a customer on a unique purchase occasion. The data identifies the customer, ordering channel, date and time of purchase, number of units purchased and price paid. We replicate our analysis using transactions from 2010 and 2011. Because the data sample ends on 9 December 2011, the actual data periods are:

2010: 10 December 2009 through 9 December 2010

2011: 10 December 2010 through 9 December 2011.

---

[1] This data was obtained from: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_SF1_H2&prodType=table

[2] The retailer does attempt to match store transactions with Internet and catalog customers. However, this matching process is not complete. When we include the store transactions that are matched we obtain a very similar pattern of findings.

ApparelCo also provided us with customer reviews of their products, including a product rating between 1 and 5 (where 5 is most favorable). Customers who have registered on the ApparelCo's Internet site can review individual products. The data set includes a complete record of all of the reviews submitted between April 2008 and December 2011. We exclude a small number of reviews screened out by a third party due to inappropriate content, such as vulgar language or mentions of a competitor. The data includes 330,975 reviews of 8,283 products. The reviews have an average rating of 4.31 and 86.0% of them recommend the product.

When analyzing the ApparelCo data we identify households using the zip code of the purchaser. In some cases this is different than the zip code of the recipient. Reassuringly, our results replicate when we use the zip code of the recipient instead of the purchaser.

Our investigation of contributions to congressional election candidates was based on a data set combining two sources. The individual campaign contribution dataset was provided by a third party watchdog organization, the Center for Responsive Politics (CRP). CRP digitizes reports published by the Federal Election Commission and organizes them into standardized data sets.[3] The data records any individual contribution made to a congressional candidate, including variables such as the individual's name, address, zip code, donation amount, recipient candidate and date of contribution. We supplement this data with congressional election outcomes from the Clerk of the US House of Representatives. The contribution data span 1990 to 2014, while the election outcomes data extend from 1990 to 2010.

To study the change in house prices we obtained house price data from Zillow.com.[4] This dataset reports median home values by zip code (the Zillow Home Value Index) and includes single-family residences, condominiums and co-op homes.

Throughout the paper we indicate significance (statistically significantly different from zero) using: ** $p < 0.01$, * $p < 0.05$ and $^{\dagger}p < 0.10$.

In the next section we use data from MassStore to investigate whether there are zip codes whose households systematically purchase new products that fail.

3. Harbinger Zip Codes

Recall that the transaction data provided by MassStore extends from January 2013 through July 2016. We identify new items introduced between July 2013 and June 2014.[5] We then use the period from July 2014 to July 2016 to evaluate whether these new items survive 18 months. We label a new product as a failure if its last sale was within 18 months of the product's introduction (identified by the date of its first sale).

[3] Federal campaign finance laws in the US require candidate committees, party committees, and PACs to file periodic reports disclosing the money they raise and spend. Federal candidates and committees must provide the names, occupations, employers and addresses of all individuals who give them more than $200 in an election cycle. The Federal Election Commission maintains this database and publishes the information about campaigns and donors on its web site.

[4] This data was obtained from: http://files.zillowstatic.com/research/public/Zip/Zip_Zhvi_AllHomes.csv.

[5] These items had no sales between January 2013 and July 2013.

We exclude items that survive less than 90 days, which helps avoid short-term or seasonal items, such as St Patrick's Day products. Furthermore, we require that in these 90 days the item is sold in at least a quarter of the retailer's stores. This restriction excludes product that were being tested in a few stores and items with limited geographic appeal (such as a Boston Red Sox hat). We also exclude categories of products that are explicitly introduced with a short-term purpose as labeled by the retailer.

We randomly divide the new items into "classification" items and "holdout" items. There are 2,324 items in the classification category and 2,388 items in the holdout category. As we would expect, the 18-month survival rates are similar for this data split. The classification set had a 64.07% survival rate, and the holdout set had 65.49%.

This success rate is higher than in Anderson et al (2015). In their sample of 8,809 new products, 3,508 (40%) survived for 3 years (12 quarters). The higher success rate reported in this paper may reflect the use of an 18-month window rather than a three-year window. It could also reflect the more mainstream product focus of MassStore. The new products have all survived the retailer's initial market tests and are broadly introduced across the retailer's stores. If we were able to observe the full sample of new products that were either proposed by manufacturers or that were subjected to initial market tests by this retailer, the success rate would be considerably lower.

We classify five-digit zip codes into groups by focusing on purchases of the 2,324 new items in the classification category in the first 90 days after each new item is introduced. We calculate the proportion of new items purchased by customers in each zip code that succeeded (had sales after 18 months). To construct this zip code level average we weight by the number of transactions for each new item. We then rank the zip codes based on this average and organize them into groups based on quartiles of the average success rate.[6]

Group 1	over 85.9% average success rate
Group 2	84.1% to 85.9% average success rate
Group 3	82.1% to 84.1% average success rate
Group 4	under 82.1% average success rate

We then estimate the same OLS models used by Anderson et al. (2015):

$$\text{Success}_j = \alpha + \beta_1 \log \text{Total Orders}_j + \varepsilon_j \quad (1)$$

$$\text{Success}_j = \alpha + \beta_1 \log \text{Group 1 Orders}_j + \beta_2 \log \text{Group 2 Orders}_j + \beta_3 \log \text{Group 3 Orders}_j + \beta_4 \log \text{Group 4 Orders}_j + \varepsilon_j \quad (2)$$

[6] Notice that these average success rates are higher than the average new product survival rates reported above. Recall that the average success rates are weighted by purchase frequency and, as we might expect, the products that survive are purchased more frequently.

The unit of analysis in both models is a new product (*j*) and the estimation sample is the 2,388 items in the holdout sample. *Total Orders* describes the total number of purchases of new product *j* calculated using all of the households. In Model 2, *Group 1 Orders* measures the total number of purchases of new product *j* by households from zip codes in Group 1 (with analogous definitions for the variables corresponding to the other groups). We use a $\log_{10}$ transformation for all of these total order variables.[7] Recall that the grouping of zip codes is based upon a different sample of new products (the classification items) than the items used to estimate these models (the holdout items).

For completeness we also report findings from an alternative model estimated by Anderson et al. (2015):

$$\text{Success}_j = \alpha + \beta_1 \log \text{Total Orders}_j + \beta_1 \,\%\, \text{Group } 2_j + \beta_1 \,\%\, \text{Group } 3_j + \beta_1 \,\%\, \text{Group } 4_j + \varepsilon_j \qquad (3)$$

The *% Group* 2_j variable measures the proportion of orders contributed by households in zip codes in Group 2. The *% Group* 3_j and *% Group* 4_j variables are defined similarly. The proportions for the four groups sum to 100% and so we omit the *% Group* 1_j variable. Under this specification the coefficient for the *% Group* 2_j variable can be interpreted as the change in the probability of success when the proportion of sales contributed by households in Group 2 increases (and there is a corresponding decrease in the proportion contributed by households in Group 1).

The findings for all three models are reported in Table 3.1. Model 1 confirms that new products that sell more in their first 90 days are more likely to survive for 18 months. However, Models 2 and 3 reveal that the likelihood a new product will survive depends not just upon total purchases; it also depends upon which households are making those purchases. An increase in purchases by households in Groups 1 and 2 is associated with a higher likelihood of success. However, additional purchases from Groups 3 and 4 is a signal that the new product will fail.

In particular, an order of magnitude (10-fold) increase in orders from Group 4 is associated with a 51.72% reduction in the probability that the new item will survive. Similarly, if the proportion of purchases from zip codes in Group 4 increases by 10% (with 10% less from Group 1), there is a 27.8% reduction in the probability that the item will succeed.

Separately controlling for the purchases made by each group sharply increases the explanatory power of the model. The R^2 jumps from 0.0518 in Group 1 to 0.1713 and 0.1769 in Groups 2 and 3. While items with more *Total Orders* in the first 90-days are more likely to succeed (the *Total Orders* coefficient is positive), it is not just the number of orders that is important. It is also important to know which households are placing those orders. If the new product is purchased by households in harbinger zip codes, then more orders signals the reverse outcome - a lower probability of success.

More generally, the findings confirm that there are some zip codes that contain households that systematically purchase new products that fail. They extend the results reported by Anderson et al. (2015) by identifying clusters of harbingers at the zip code level. Recall that Anderson et al. (2015) used household level data to identify harbinger customers; households whose purchases of a new

[7] In this respect the model is different from the model estimated by Anderson et al. (2015). The volume of purchases is much larger in the MassStore data than in the dataset they analyzed.

product are an indication that the product will fail. They acknowledged that an important challenge limiting the application of their finding is that detailed household level data is not always available to identify harbingers. By extending the result to the zip code level we help overcome this challenge and expand the potential applications for their result. Our findings suggest that it may be sufficient to know which zip code a household resides in.

Table 3.1. Harbinger Zip Codes at MassStore

	Model 1	Model 2	Model 3
Log Total Orders	16.45%** (1.44%)		8.86%** (1.40%)
Log Orders Group 1		61.85%** (7.00%)	
Log Orders Group 2		8.22% (17.84%)	
Log Orders Group 3		-8.59% (17.28%)	
Log Orders Group 4		-51.72%** (10.48%)	
% Group 2			41.58% (35.13%)
% Group 3			-92.74%** (28.39%)
% Group 4			-277.64%** (25.85%)
Intercept	10.99%* (4.87%)	28.17%** (4.37%)	127.77%** (9.83%)
R^2	0.0518	0.1713	0.1769

The table reports the coefficients from estimating Equations 1, 2 and 3. Standard errors are in parentheses. In both models the unit of analysis is a new item and the sample size in all models is 2,388. Standard errors are in parentheses.

Our findings in the remaining sections of the paper show that the grouping of harbinger zip codes that we have identified in this analysis explains variation in household decisions across a broad range of decision contexts. The implication is that we do not need to re-identify harbinger zip codes for each decision context. Instead the ranking of zip codes using MassStore new product purchases is sufficient to explain variation in choices in other settings. For this reason, we provide the average success rate of new product purchases by zip code in the Appendix to this paper. This average is constructed using all 4,712 new products from both the classification and holdout sets. It is this measure that we use in other sections in this paper to explore the extent to which the grouping of harbinger zip codes explains variation in decisions in other contexts.

Before extending the analysis to other settings we further investigate the differences between harbinger zip codes and non-harbinger zip codes at MassStore. We begin by comparing purchases of existing items that were not newly introduced.

Niche or Very Niche Items

For the 2013 calendar year we use the number of orders within a three-digit zip code to define items that are “niche” or “very niche”. In particular, we rank order the items within each 3-digit zip code according to how frequently they are purchased. As an indication of the size of a 3-digit zip code, the greater Boston metropolitan area includes eight 3-digit zip codes.[8] Within each 3-digit zip code we identify the least frequently ordered items using different cumulative purchase thresholds:

- Niche: collectively contribute 10% of total orders within a 3-digit zip code.
- Very niche: collectively contribute 1% of total orders within a 3-digit zip code.

These labels identify items that are purchased infrequently within neighboring regional zip codes. We can then ask whether these items are purchased more frequently by harbinger zip codes than other zip codes. For each (5-digit) zip code we calculate a weighted average of the proportion of 2013 orders that are niche, or very niche (weighted by the number of orders for each item in that zip code).

In Table 3.2 and Figure 3.1 we report these averages when grouping the zip codes using the *Average Success Rate* for orders of new products in that zip code. We rank the zip codes using the average success rate of their new product purchases, and group them using quartile splits of this ranking (where Group 1 has the highest average success rate and Group 4 has the lowest).[9] When ranking the zip codes we restrict attention to 5-digit zip codes with at least 200 orders of new products. When calculating the averages we restrict attention to three-digit zip codes with at least 100,000 orders of any items (new or existing).

We see that customers in harbinger zip codes are significantly more likely to purchase niche items compared to neighboring zip codes. There is a monotonic increase in the proportion of niche and very niche items purchased as we move from Group 1 to Group 4. We conclude that households in harbinger zip codes not only purchase new items that are more likely to fail, they also tend to purchase existing items that neighboring households are less likely to purchase. This is consistent with our interpretation that households in these zip codes have preferences that are not representative of other customers.

[8] Three-digit zip codes comprise the first three digits of each five-digit zip code. For example the three-digit zip code for 5-digit zip code 12345 is 123. More information about three-digit zip codes can be found at: http://pe.usps.com/Archive/PDF/DMMArchive20030810/L002.pdf

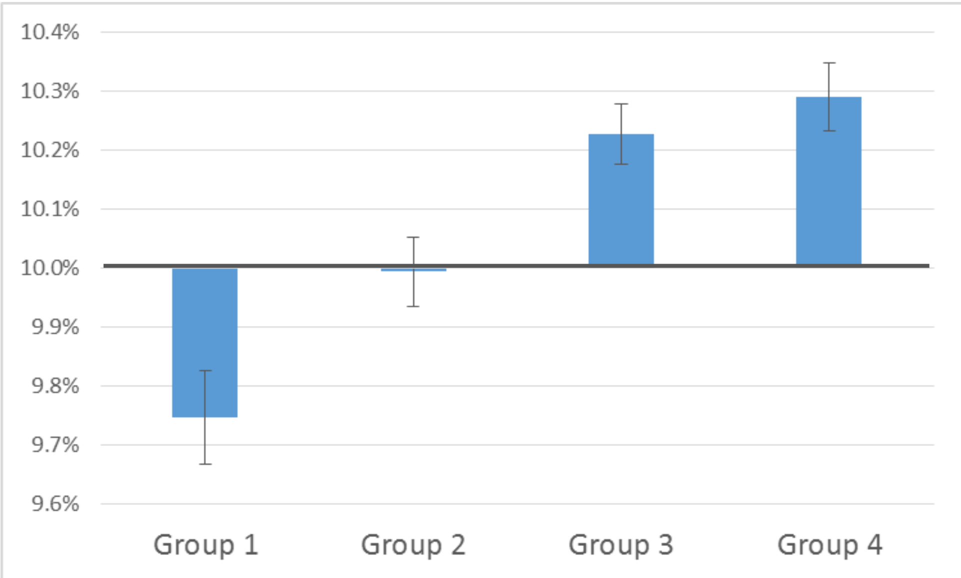
[9] When identifying this grouping, we use all 4,712 new MassStore products (including both the classification and holdout samples). The quartile splits are slightly different for this analysis than for the previous analysis. We provide details of the quartile splits for this and subsequent analyses in the Appendix.

Table 3.2. Purchases of Niche Items at MassStore

	Group 1	Group 2	Group 3	Group 4
Niche	9.746% (0.041%)	9.994% (0.030%)	10.228% (0.026%)	10.291% (0.029%)
Very Niche	0.958% (0.006%)	0.992% (0.005%)	1.034% (0.004%)	1.075% (0.005%)
Sample Size	1,457	1,458	1,456	1,458

The table reports the proportion of purchases from MassStore of existing items that were niche or very niche. The averages are calculated separately for each group of zip codes, where the zip codes are organized according to the success rate of new product purchases. The unit of analysis is a zip code. Standard errors are in parentheses.

Figure 3.1. Purchases of Niche Items at MassStore



The figure reports the proportion of purchases from MassStore of existing items that were niche. The averages are calculated separately for each group of zip codes, where the zip codes are organized according to the success rate of new product purchases. The unit of analysis is an item. Sample sizes are reported in Table 3.2. Error bars indicate 95% confidence intervals.

We conducted several robustness checks. First, we repeated the analysis using transactions from the 2014 calendar year (instead of the 2013 calendar year). Second, we used transactions from 2013 to identify niche items and transactions from 2014 to calculate the average "nicheness" of the orders. Finally we also repeated the analysis when using 4-digit zip codes to identify niche items. The pattern of findings survives all of these robustness checks.

Our next analysis uses the demographic data provided by MassStore and the US Census to compare the demographic characteristics of harbinger zip codes with other zip codes.

Demographic Characteristics

In Table 3.3 we summarize the average demographic measure by zip code group, where the unit of analysis is a zip code. We also include the pair-wise correlation between the demographic measures and the average success rate for new products purchased in each zip code.

The findings reveal that harbinger zip codes are distinctively less urban than other zip codes. None of MassStores stores are in rural locations, and so the non-urban locations can generally be interpreted as suburban locations. Perhaps consistent with rural locations, the zip codes contain fewer households and they tend to have lower household incomes and home values, older heads of households, and a higher proportion of single family homes. They are also located further away from both MassStore's stores and its competitors' stores.

Table 3.3. Demographic Characteristics

Variable	Group 1	Group 2	Group 3	Group 4	Correlation with Avg. Success Rate
Age	51.96	53.08	53.68	54.75	-0.285**
Home Value	$346,907	$279,091	$251,305	$215,200	0.260**
Income	$84,002	$88,404	$83,079	$72,807	0.069**
Single Family	63.94%	80.04%	82.87%	85.76%	-0.359**
Multi-family	35.72%	19.48%	16.39%	13.29%	0.366**
Distance	8.38	9.99	11.62	16.21	-0.178**
Competitors Distance	9.37	12.55	13.83	23.18	-0.291**
Number of Households	8,914	7,300	6,750	4,238	0.298**
% All Urban	84.14%	75.40%	69.44%	47.94%	0.351**
% Very Urban	80.59%	70.74%	63.83%	37.90%	0.368**
% Less Urban	3.55%	4.66%	5.61%	10.04%	-0.132**
% Rural	15.86%	24.60%	30.56%	52.06%	-0.351**

The table reports the average of each measure within each zip code group. Column 5 reports the pair-wise correlation between each measure and the average success rate of new product purchases in that zip code. The demographic variables are defined in the Appendix, where we also provide summary statistics. The unit of analysis is a zip code. The sample size for the pair-wise correlations is 5,202 zip codes. The sample sizes for the group averages are 1,300, 1,301, 1,300 and 1,301 for Groups 1 through 4 (respectively), except the urban and rural measures, where the sample sizes are 1,410, 1,376, 1,366 and 1,473 (respectively).

This evidence that harbinger zip codes are less urban is again consistent with our interpretation that households in these zip codes have preferences that are not representative of preferences in other zip codes. Households in suburban zip codes could reasonably be expected to have preferences that are different from households in urban zip codes. However, while harbinger zip codes are more likely to be suburban, this is only a partial explanation for the differences between harbinger zip codes and other zip codes. Simply attributing the harbinger effect to suburban preferences is not a complete explanation for the effect. We can illustrate this by retaining the residuals from the following OLS model, where the unit of analysis is a zip code (*z*):

$$\text{Average Success Rate}_z = \alpha + \beta_1 \text{ \% Very Urban}_z + \beta_2 \text{ \% Less Urban}_z + \varepsilon_z \tag{3}$$

The estimated coefficients (and standard errors) from this model are:

- Very Urban: 0.0272 (0.0009)
- Less Urban: 0.0082 (0.0023)

These coefficient confirm that harbinger zip codes are much less likely to be urban. The residuals from this model contain the variation in the average success rate that is not explained by the degree to which a zip code is urban. In the Appendix we report the pair-wise correlation between these residuals and the demographic variables. They reveal that even after controlling for the degree to which zip codes are urban we see a significant correlation between the average success rate and almost all of the demographic variables. The only correlation that is no longer statistically significant is the relationship with household income.

We conclude that while some of the variation in the success rate associated with the demographic variables can be attributed to urban vs. suburban locations, this urban vs. suburban difference does not fully explain these relationships. This should not be surprising. For example, we would expect that households with different age profiles would have different preferences irrespective of whether they are urban or suburban.

Summary

Using data from MassStore we have shown that there exist harbinger zip codes. Households in these zip codes are consistently more likely to purchase new products that fail than households in other zip codes. They are also more likely to purchase niche products that are rarely purchased by neighboring zip codes. These harbinger zip codes tend to be located in suburban rather than urban areas. They also tend to be located further away from MassStore's (and its competitors') stores, they contain fewer households, and they have lower average home values, more single family homes, and older heads of households.

The evidence that we can identify harbingers at the zip code level greatly simplifies the challenge of identifying harbingers. This will be particularly helpful if the same zip codes exhibit harbinger preferences across a broad range of decision contexts. We investigate this in the reminder of this paper, where we use the classification of harbinger zip codes in the MassStore data to explain variation in purchasing decisions at another retailer (Section 5), donations to congressional election campaigns (Section 6), and changes in house prices (Section 7).

4. Harbinger Zip Codes at MassStore and Purchases from ApparelCo.

Unlike MassStore where successful new products are continued and unsuccessful new products are discontinued, ApparelCo often updates its products between seasons irrespective of whether the products are successful. This means that observing that a new product was discontinued does not imply that the product flopped. ApparelCo does not have an obvious basis for evaluating new product success or failure, and so we are unable to repeat the same new product analysis that we conducted with the MassStore data.

Instead, we focus our analysis of the ApparelCo data on four questions. First we investigate whether harbinger zip codes were more likely to purchase “niche” items that households in neighboring zip codes were less likely to purchase. Second, ApparelCo provided data describing customer product reviews. We investigate whether harbinger zip codes are more likely to purchase products with unfavorable reviews. Third, we investigate whether harbinger zip codes are more likely to purchase products that other customers return. Finally, we identify a sample of households that changed zip codes. We use this sample to ask whether the clustering of harbingers in harbinger zip codes is because of where households choose to live, or because households change their preferences when they move into a harbinger zip code.

An important difference in this analysis is that we classify the zip codes into harbinger and non-harbinger zip codes using new product purchases from one company (MassStore) and use these classifications to compare purchases from a different company (ApparelCo). We start with the analysis of niche purchases.

Do Harbingers Purchase Niche Items from ApparelCo?

We use the same approach as we used to study niche purchases at MassStore. In particular, we use the number of purchases within a three-digit zip code to define items that are “niche” (see the earlier definition). We then calculate a weighted average of apparel orders in a zip code that are niche.

We construct the data sample using the same approach as our earlier analysis. We restrict attention to three-digit zip codes with at least 100,000 purchases of any item from ApparelCo. We also restrict attention to zip codes with at least 200 orders of new products from MassStore.

In Table 5.1 we report the average proportion of ApparelCo purchases that are niche. To evaluate robustness we report separate results for ApparelCo purchases in 2010 and 2011. We group the zip codes into 4 groups using the average success rate for orders of new products from MassStore. When constructing this grouping we use all 4,712 new MassStore products, including both the classification and holdout samples. The grouping uses the same approach as our earlier analyses; the zip codes are rank ordered by the average success rate and then grouped using quartile splits of this ranking.

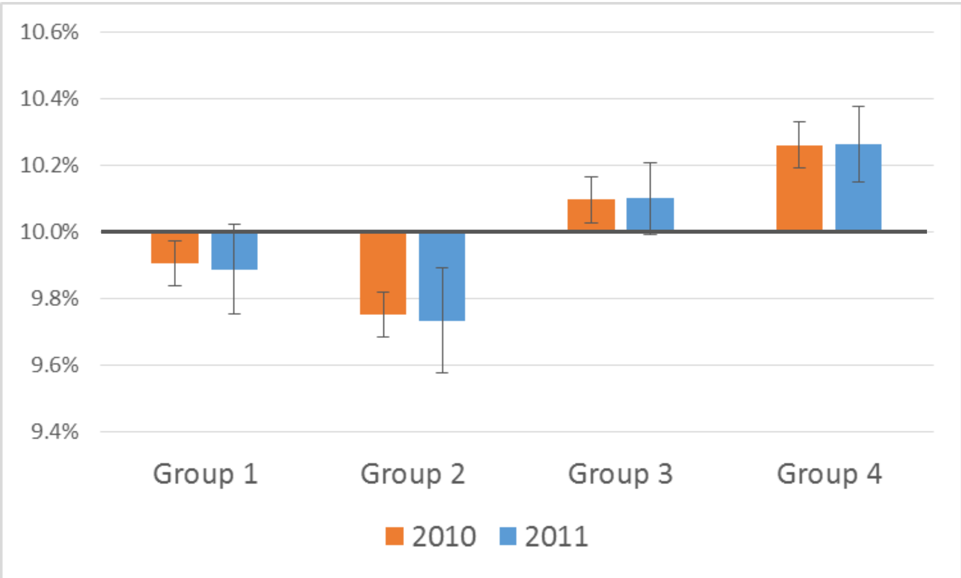
The findings confirm that harbinger zip codes do purchase more niche items from ApparelCo compared to neighboring zip codes (that share the same 3-digit zip code). We illustrate this relationship in Figure 5.1. In both years the proportion of niche items purchased by zip codes in Groups 1 and 2 are significantly lower than in Groups 3 and 4 ($p < 0.01$). We also repeat this analysis for “very niche” items (the bottom 1% of purchases). The findings reveal a similar pattern.

Table 5.1. Percentage of ApparelCo Purchases that are Niche

		Group 1	Group 2	Group 3	Group 4
Niche	2010	9.889% (0.074%)	9.736% (0.081%)	10.103% (0.055%)	10.265% (0.057%)
	2011	9.892% (0.069%)	9.760% (0.082%)	10.118% (0.055%)	10.167% (0.054%)
Very Niche	2010	0.987% (0.012%)	0.972% (0.012%)	1.011% (0.009%)	1.019% (0.010%)
	2011	0.987% (0.012%)	0.975% (0.012%)	1.015% (0.010%)	1.015% (0.011%)
Sample Size	2010	498	496	498	497
	2011	493	492	492	493

The table reports weighted averages describing the proportion of ApparelCo purchases that were niche or very niche. The unit of analysis is a zip code. The zip codes are grouped according to the success or failure of MassStore's new product transactions. When calculating the averages we weight by the number of units purchased of each item in that zip code. Standard errors are in parentheses.

Figure 5.1. Percentage of ApparelCo Purchases that are Niche



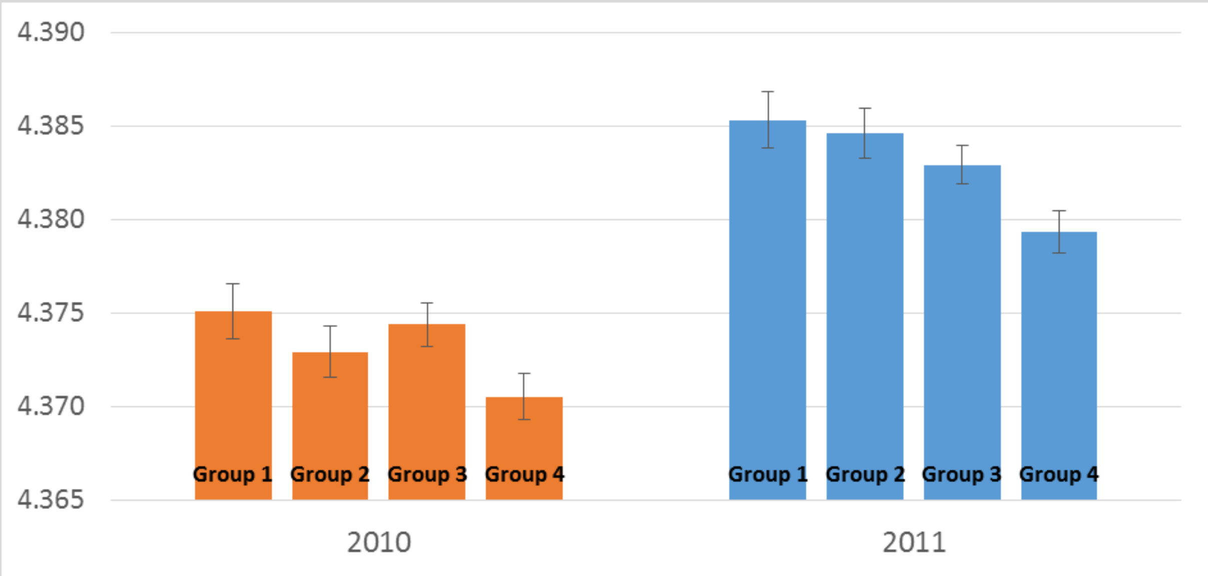
The figure reports weighted averages describing the proportion of purchases that were niche. The averages are calculated separately for each group of zip codes, where the zip codes are organized according to the success rate of new product purchases at MassStore. The error bars indicate 95% confidence intervals.

We next investigate whether customers in harbinger zip codes are also more likely to purchase ApparelCo items that have less favorable product reviews.

Do Harbingers Purchase Items that Receive Less Favorable Reviews from Other Customers?

As we discussed in Section 3, the product review data includes a product rating ranging from 1 to 5, where 5 is the highest rating. We again group the zip codes according to failure rates for new products purchased at MassStore. We then calculate an average product rating for items purchased in each zip code, weighting by the number of purchases of each item in each zip code. When calculating this average we exclude reviews written by households in the focal zip code.[10] This allows us to measure reviews contributed by “other customers” where the identity of the “other customers” varies for each zip code. We restrict attention to the 724 items with at least 100 reviews. The findings are summarized in Figure 5.2 with complete details reported in the Appendix.

Figure 5.2. Customer Reviews of Items Purchased from ApparelCo



The figures report the average product rating in customer reviews for products purchased in each group of zip codes. The unit of analysis is a zip code. The zip codes are grouped according to the success of new product purchases at MassStore. The averages are weighted by the number of ApparelCo purchases of that item in that zip code. Error bars indicate 95% confidence intervals. Detailed findings (including sample sizes) are provided in the Appendix.

Harbinger zip codes at MassStore purchase ApparelCo items that have lower average product ratings (compared to other zip codes). In both years we see that zip codes in Group 4 buy products with significantly lower product ratings. It is possible that this finding may in part reflects regional difference in which products customers purchase.[11] To investigate this possibility we can explicitly control for regional differences using the following OLS model:

$$\text{Average Product Rating}_z = \alpha + \beta_1 \text{ Average Success Rate}_z + \boldsymbol{\beta} \textbf{ 3-Digit} + \varepsilon_z \qquad (5.1)$$

[10] This has little impact on the results.

[11] Notice that the previous analysis of niche purchases controls for regional differences by identifying niche products using purchases within a 3-digit zip code.

The unit of analysis is a zip code (*z*). The **β 3-Digit** denotes fixed effects identifying each 3-digit zip code, which controls for regional variation in the products that customers purchase. The coefficient of interest is β_1. The inclusion of the fixed effects ensures that β_1 is only identified by variation within each 3-digit zip code (and not by variation across 3-digit zip codes). It measures how the *Average Product Rating* varies within a 3-digit zip code according to the *Average Success Rate* of new product purchases at MassStore. If harbinger zip codes purchase products that receive lower product ratings than neighboring zip codes then we would expect to see a positive coefficient for this variable, indicating a lower *Average Product Rating* when the *Average Success Rate* is lower. We repeat the analysis separately using 2010 and 2011 transactions and report the coefficient of interest in Table 5.2.

Table 5.2. Do Harbingers Purchase Items that Receive Less Favorable Product Reviews?

	2010	**2011**
Average Success Rate	7.56%** (2.43%)	3.92%† (2.22%)
R^2	0.1064	0.0940
Sample Size	5,946	5,948

The table reports the *Average Success Rate* coefficients (β_1) when estimating Equation 5.1. The unit of analysis is a zip code. Fixed effects were estimated but are omitted from the table. Standard errors are in parentheses.

Compared to neighboring zip codes, harbinger zip codes purchase products that receive lower product ratings. This finding is consistent across both 2010 and 2011, although it is only marginally significant in 2011. It confirms that the evidence that harbingers purchase items that receive less favorable reviews cannot simply be explained by regional differences.

We also investigated whether households in harbinger zip codes write more or less favorable reviews than households in other zip codes. When controlling for regional differences (within a 3-digit zip code) we found no evidence of any difference in the average product ratings.

In our next analysis we switch the focus to product return decisions. We compare whether harbinger zip codes are more or less likely to purchase items that other customers return. If harbingers have unusual preferences then we might expect that products purchased by harbingers would tend to have relatively high return rates from other customers.

Do Harbingers Purchase Items that Other Customers Return?

We begin by calculating the average return rate for each item. Like the review analysis, when calculating the average return rate for an item we exclude return decisions in the focal zip code.[12] This allows us to measure return decisions by "other customers". We then use purchases by customers in each zip code to calculate the average return rate (by other customers) for items

[12] This again has little impact on the results.

purchased in that zip code, weighting by the number of units purchased. We then re-estimate Equation 5.1 using the *Average Return Rate* (by other customers) as the dependent variable.

The coefficient of interest (β_1) is reported in Table 5.3. In this analysis the coefficient of interest measures how the *Average Return Rate* (by other customers) varies within a 3-digit zip code according to the *Average Success Rate* of new product purchases at MassStore.[13] If harbinger zip codes are more likely to purchase items that other customers return, then we would expect to see a negative coefficient for this variable, indicating a higher *Average Return Rate* when the *Average Success Rate* is lower.

Table 5.3. Do Harbingers Purchase Items that Other Customers Return?

Dependent Variable	2010	2011
All Return Reasons	-1.59%* (0.62%)	-1.57%** (0.61%)
Did Not Like the Item	-0.67%** (0.21%)	-0.76%** (0.20%)
Wrong Size	-0.72%† (0.37%)	-0.75%* (0.38%)
Defective	0.03% (0.03%)	0.05%† (0.03%)
Miscellaneous	-0.23%** (0.08%)	-0.11% (0.07%)
Sample Size	5,948	5,950

The table reports the *Average Success Rate* coefficients (β_1) when estimating Equation 5.1. The unit of analysis is a zip code x item and the observations are weighted by the number of purchases in each zip code. Fixed effects were estimated but are omitted from the table. Standard errors are in parentheses.

The findings confirm that harbinger zip codes are more likely than neighboring zip codes to purchase items that other customers return. A 10% decrease in the *Average Success Rate* in a zip code is associated with a 0.16% increase in the average return rate by customers in other zip codes. This effect is almost identical when using transactions from 2010 or 2011.

One interpretation of a high return rate on an item is that it indicates many of the other customers did not like the item. Thus, a tendency to purchase items that other customers return suggests that harbingers purchase items that many others do not like. This argument would be stronger if we knew that other customers returned the item because they did not like it. Fortunately ApparelCo asks customers to indicate the reason that an item is returned. We group the return reasons into four groups and summarize the frequency of their occurrence in Table 5.4. We provide additional details about these groupings in the Appendix.

[13] The inclusion of the fixed effects ensures that β_1 is only identified by variation within each 3-digit zip code (and not by regional variation across 3-digit zip codes).

Table 5.4. Average Return Rate by Return Reason

Return Reason	2010	2011
Did Not Like The Item	4.13%	4.25%
Item is the Wrong Size	6.72%	7.27%
Item is Defective	0.32%	0.37%
Miscellaneous	2.15%	1.92%
All Reasons	13.32%	13.80%

The table describes the average return rate by return reason. They are calculated separately using return transactions in 2010 and 2011. We only consider transactions in zip codes for which we have average success rates from MassStore. Sample sizes are 6,190,833 (2010) and 6,226,019 (2011). Additional details are provided in the Appendix.

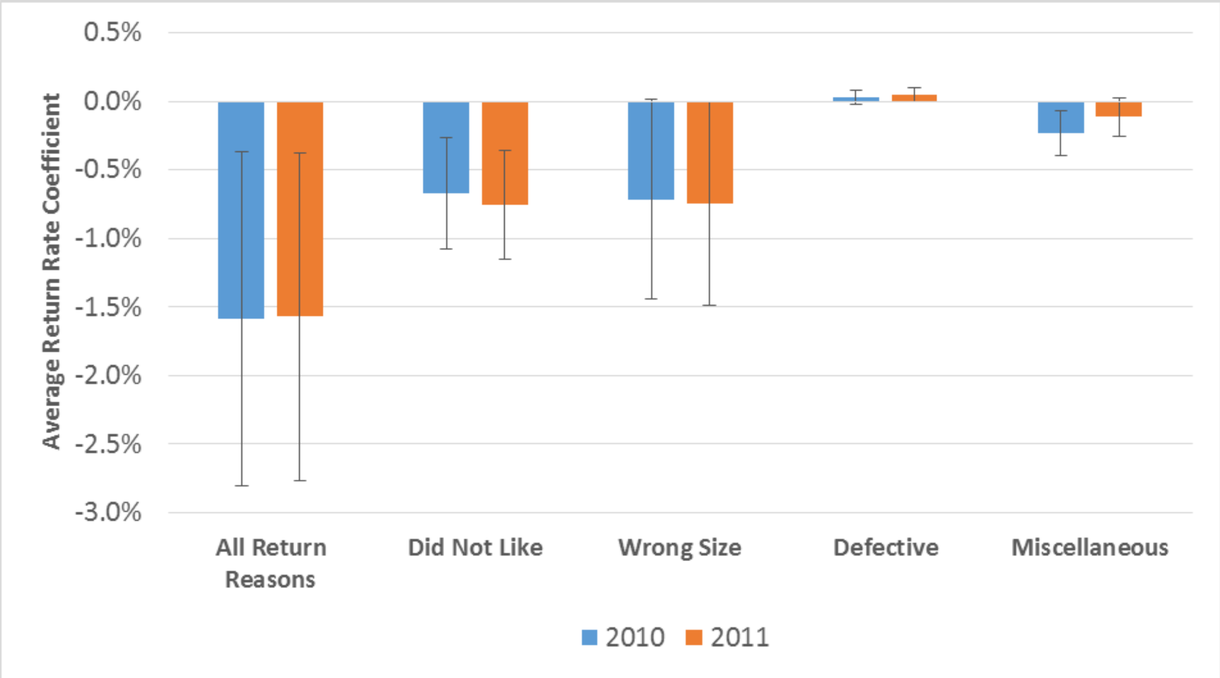
The most common reason an item is returned is because it is the wrong size. The next most common reason is that the customer did not like the item when it arrived. This category includes returns due to color, material or styling. There are also returns due to quality defects, although these are relatively uncommon. These return reasons allow us to further investigate our interpretation that harbingers have preferences that are different from other customers. To do so we separately re-estimate Equation 5.1 using different dependent variables. The dependent variables measure the proportion of items returned (by other customers) for each of the four return reasons. The coefficients of interest are reported in Table 5.3 and summarized in Figure 5.3.

The analysis of return reasons reveals that the tendency of harbingers to buy items that other customers return is concentrated in returns due to (a) size and (b) because customers did not like the item. Preferences for the color, material or styling, together with the size and fit of an item are idiosyncratic, reflecting customers' body shapes and personal tastes. The evidence that harbingers are more likely to purchase items that other customers return for these idiosyncratic reasons is consistent with our interpretation that harbingers have preferences that are different from other customers.

In contrast, quality defects are not idiosyncratic and we would expect returns due to defects to affect all customers in approximately the same way. It is therefore notable that we do not see the same effect for returns due to defects. The *Average Success Rate* coefficients when estimating Equation 5.1 using returns due to defects are not negative in either year. This reinforces the conclusion that the tendency for harbinger zip codes to purchase items that other customers return reflects idiosyncratic differences in tastes.

Returns due to defects are relatively uncommon, and it is possible that this may have contributed to why we do not see a significant negative coefficient for this return reason. However, this finding does not appear to just reflect a lack of power. The signs of the *Average Success Rate* coefficients are reversed in the defective returns models. Moreover, the confidence intervals around these parameter estimates are relatively tight. To further investigate this possibility we re-estimated all of the models when restricting attention to product families that had higher than median returns due to defects. The pattern of findings was unchanged.

Figure 5.3 Do Harbingers Purchase Items that Other Customers Return?



The table reports the *Average Success Rate* coefficients when estimating Equation 5.1 using different dependent variables. The unit of analysis is a zip code and the sample sizes are 5,948 (2010) and 5,950 (2011). Error bars indicate 95% confidence intervals.

In our final ApparelCo analysis we shift focus and ask whether harbinger zip codes result from households with harbinger preferences choosing to cluster together, or whether households learn these preferences by observing their neighbors. We do so by studying ApparelCo customers that changed zip codes.

Do Households Bring Their Preferences or Learn Them?

In this analysis we identify households' zip codes using the addresses to which they shipped orders.[14] All of the households placed at least one order in both 2007 and 2011. Their 2007 orders were all shipped to one zip code and their 2011 orders were all shipped to another zip code (we exclude households that shipped to multiple zip codes within either calendar year).[15] We first ask whether households that started in harbinger zip codes in 2007 moved to other harbinger zip codes in 2011.

[14] In the other analysis in this section we identify households' zip codes using their registered addresses. However, we only have registered addresses for a subset of the households in 2007. The registered address is generally the billing address for their credit cards (we only received data describing the 5-digit zip code, not the complete street address).

[15] Note that because the data period ends on 9 December 2011 the years are defined as ending on 9 December of that year and so include 22 days from the previous calendar year. There is a small risk of error because households sometimes ship items as gifts. Although, this is unlikely to provide an alternative explanation for

There were 28,476 households that changed zip codes from 2007 to 2011. We grouped zip codes into quartiles using the MassStore new product survival rates, and mapped the transfers of these 28,476 households between each group of zip codes. This mapping is summarized in Table 5.5.

Table 5.5 Do Households in Harbinger Zip Codes Move to Other Harbinger Zip Codes?

	Ended in Group 1	**Ended in Group 2**	**Ended in Group 3**	**Ended in Group 4**	**Number of Households**
Started in Group 1	4,097	2,296	1,477	1,080	8,950
Started in Group 2	1,609	2,042	1,910	1,310	6,871
Started in Group 3	1,154	1,613	1,790	1,723	6,280
Started in Group 4	740	1,082	1,664	2,889	6,375
Nbr of Households	7,600	7,033	6,841	7,002	28,476

The table reports the movements of 28,476 households that changed zip codes from 2007 to 2011. The zip codes are grouped according to the average survival rate of new products purchased from MassStore. The rows indicate the grouping of zip codes that households moved from in 2007, and the columns indicate the grouping of zip codes that the households lived in by 2011.

We see clear evidence that when households in harbinger zip codes move, they move to another harbinger zip code. Of the 6,375 households that started in a Group 4 zip code, 2,889 (45.3%) moved to another Group 4 zip code, but just 740 (11.6%) moved to a Group 1 zip code. In contrast, of the 8,950 that started in a Group 1 zip code, just 1,080 (12.1%) moved to a Group 4 zip code, while 4,097 (45.8%) moved to another Group 1 zip code. Across all 28,476 households that changed zip codes, the pair-wise correlation between the average new product survival rate in 2007 and 2011 is 0.63.

These findings reveal systematic differences in where households choose to live. They suggest that one reason we see harbinger zip codes is that harbingers choose to move into zip codes occupied by other harbingers, while non-harbingers choose to do the opposite. For example, if non-harbingers like ocean views but harbingers do not then we would expect to see clusters of harbingers in zip codes that do not view the ocean.

We conducted two robustness checks. First, we grouped the households according to whether they moved within a 3-digit zip code or to a separate 3-digit zip code. The pattern of findings survives in either case. Second, instead of identifying households from the zip codes to which they mailed purchases, we identified them from the address they have registered with the retailer. In particular, we compared the zip codes in each household's registered address as of 19 October 2007 and 9 December 2011. We only know the registered addresses in 2007 for a randomly selected subsample of the households, which reduced the sample size. Nevertheless, the pattern of findings was again unchanged.

---

our findings, we will also replicate the results for a subset of the customers for which we have data describing their registered addresses in both 2007 and 2011.

We can also use households that changed zip codes to address a second question; do households change their purchasing decisions when they move into a harbinger zip code? In particular, we investigate whether households' purchases of niche items changed after they moved zip codes. In this analysis we focus on 16,904 households that changed zip codes between 2007 and 2009 and did not change zip codes between 2009 and 2011.[16] These 16,404 households all satisfy the following criteria:

a. Made at least one order in 2007, 2009 and 2011.
b. Shipped all of their orders to 1 zip code in 2007.
c. Shipped all of their orders to 1 zip code in 2009.
d. Shipped all of their orders to 1 zip code in 2011.
e. The zip code was different in 2007 and 2011.
f. The zip code was the same in 2009 and 2011.

We measure the change in niche item purchases after they moved by comparing the proportion of niche items purchased in 2011 with the proportion purchased in 2009. We summarize this identification approach in Figure 5.4.

Figure 5.4. Composition of the Sample

2007	2008	2009	2010	2011
Moved zip code		Did not move zip code		
		% niche in 2009		% niche in 2011

In the table below we report the change in the % of niche items purchased. Recall that the niche items (for a given year) are identified using rank ordered purchase frequencies within a 3-digit zip code. The niche items cumulatively contribute 10% of total annual revenue in that year within that 3-digit zip code.

The change in the % of niche items purchased is calculated by pooling across the households in each combination of groups and calculating the total % of niche items purchased in 2009 and 2011. We then subtract the proportion in 2011 from the proportion in 2009 to calculate the change in the % of niche items purchased. For example, there were 577 households that were in a Group 1 zip code in 2007 and moved to a Group 4 zip code by 2011. In 2011 these 577 households purchased 4,185 items, of which 10.61% were niche items. In 2009 the same 577 households purchased 3,971 items of which 10.95% were niche items. The difference between 10.60% and 10.95% represents a -0.35% change in the proportion of niche items purchased by these 577 households. The change in % of niche items purchased for each combination of starting and ending zip code groups is reported in

[16] More precisely, the households had moved by the time of their first purchase in 2009. It is possible that the move occurred in 2009 before their first 2009 purchase. Notice that, despite this possibility, our measure of the % of niche items purchased in 2009 only measures purchases after households had moved.

Table 5.5. The number of households and items used to calculate these outcomes are reported in the Appendix.

Table 5.5 Change in the Percentage of Niche Items Purchased between 2009 and 2011

	Ended in Group 1	**Ended in Group 2**	**Ended in Group 3**	**Ended in Group 4**
Started in Group 1	-0.18%	0.61%	0.05%	-0.35%
Started in Group 2	0.26%	1.61%	0.22%	0.82%
Started in Group 3	0.96%	0.71%	1.33%	0.02%
Started in Group 4	1.26%	1.23%	1.21%	0.09%

The table reports the change in the percentage of niche items purchased between 2009 and 2011 by different groups of households. All of the households changed zip codes between 2007 and 2009 and did not change zip codes between 2009 and 2011, and so the change in the percentage of niche items purchased is the change after they moved. The zip codes are grouped according to the average survival rate of new products purchased from MassStore and we group households according to their zip codes in 2007 and 2011. The rows indicate the grouping of zip codes that households moved from in 2007, and the columns indicate the grouping of zip codes that the households lived in by 2011. Sample sizes are reported in the Appendix.

There is no evidence that when households move into harbinger zip codes they purchase more niche items. Among the households that started in Group 1 and moved to a Group 4 zip code, there was actually a small decrease in the % of niche items purchased (-0.35%). If they moved from a Group 1 zip code to another Group 1 zip code the decrease was slightly smaller (−0.18%).

There is also no evidence that when households move out of a harbinger zip code into a non-harbinger zip code they purchase fewer niche items. Among the households that started in a Group 4 zip code and moved to a Group 1 zip code (a non-harbinger zip code) the percentage of niche items actually increased (by 1.26%).

We also repeated this analysis when restricting attention to households that moved zip codes within the same 3-digit zip code. There is again no evidence that when households move into harbinger zip codes they purchase more niche items. We conclude that the data does not indicate that households acquire their harbinger preferences from their new neighbors.

Summary

The findings reported in this section represent the first evidence that the identification of harbinger zip codes is stable across decision contexts. Identifying harbinger zip codes from purchases at one company helps to explain variation in purchasing at another company. As we discussed, this has an important implication. If the same zip codes exhibit harbinger behaviors across a broad range of decision contexts then we do not need to identify harbingers separately for each decision context. Instead we can use a common list of harbinger zip codes to identify households whose choices will not be representative of other households.

The findings also have important implications for recommendation engines. While we might have thought we should rely on the advice of one friend for restaurants, another friend for movies and a third friend for books, these findings indicate that preferences are correlated across categories. This suggests that as long as you choose a friend whose preferences match yours, the friend's recommendations may be valuable in multiple categories. This is particularly important for firms such as Amazon, when selecting which items to recommend to customers. If preferences are correlated across categories they can identify customers who purchased similar items in one category to make product recommendations in another category.[17]

We also used a sample of ApparelCo customers that changed zip codes to investigate whether the harbinger zip code effect is due to where harbingers choose to live, or due to changes in preferences when households move into a harbinger zip code. There is strong support for the first interpretation; households leaving a harbinger zip code tend to move to another harbinger zip code, while the reverse is true of those leaving non-harbinger zip codes. However, there is no evidence that households learn harbinger preferences when they move to a harbinger zip code. This suggests that harbinger preferences are a sticky trait, and the harbinger zip code effect is more due to where customers choose to live, rather than households learning the preferences of their neighbors.

In the next section we further explore the extent to which preferences are correlated across different decisions. We focus on donations to congressional election candidates and investigate whether MassStore's harbinger zip codes contribute to the same candidates as households in neighboring zip codes. We also investigate whether the candidates supported by harbinger zip codes are more or less likely to win their elections.

5. Contributions to Federal Election Campaigns

In the previous sections we have shown that zip codes identified from new product purchases at MassStore make decisions that are different from households in other zip codes. We interpret these findings as evidence that households in these zip codes have preferences that are relatively unusual. What is most surprising about these results is that they extend across product categories and across retailers. In this section we investigate whether the effect also extends beyond purchasing preferences to political preferences.

We start by comparing whether harbinger zip codes make contributions to election candidates in proportions that are systematically different than their neighboring zip codes (other zip codes in the same 3-digit zip). We use data describing both the number of donors and the total $ amount contributed to congressional campaigns between 1990 and 2010. We first describe the variables used to analyze the number of donors, and then discuss the $ amount analysis.

Number of Donors

For each 3-digit zip code we identified the two candidates that received the largest number of contributions (ranked by number of donors). We then calculated the proportion of contributions that each of these candidates received:

[17] We thank Nader Tavassoli for the observations in this paragraph.

$$zip3\ nbr = \frac{nbr\ donors\ for\ candidate\ with\ most\ donors}{total\ nbr\ of\ donors\ for\ top\ 2\ candidates}$$

We also calculated a similar measure at the 5-digit zip level, where the top 2 candidates were defined by the number of donors at the zip3 level:

$$zip5\ nbr = \frac{nbr\ donors\ for\ candidate\ with\ most\ donors}{total\ nbr\ of\ donors\ for\ top\ 2\ candidates}$$

Our outcome measure is a binary indicator *More Donors for Top Candidate*, which is equal to 1 if *zip5 nbr* is greater than *zip3 nbr*, and zero otherwise. This binary indicator helps to reduce noise introduced if zip codes have relatively few donors. Given these definitions, a value of 1 for *More Donors for Top Candidate* indicates that there were proportionately more donors to the top candidate in this 5-digit zip code than in the zip 3 average.

$ Contribution Analysis

The $ contribution analysis is analogous to the number of donors analysis. For each 3-digit zip code we identified the two candidates that received the largest $ contributions (instead of largest number of donors). We then calculated the proportion of dollars that each of these candidates received:

$$zip3\ amount = \frac{\$\ for\ candidate\ with\ most\ \$}{total\ \$\ for\ top\ 2\ candidates}$$

We also calculated similar measures at the 5-digit zip level, where the top 2 candidates were defined by the total $ contributions at the zip3 level:

$$zip5\ amount = \frac{\$\ for\ candidate\ with\ most\ \$}{total\ \$\ for\ top\ 2\ candidates}$$

We again use a binary outcome measure; *More $ for Top Candidate*, which is equal to 1 if *zip5 amount* is greater than *zip3 amount*, and zero otherwise. A value of 1 for *More $ for Top Candidate* indicates that this 5-digit zip code contributed a higher proportion of dollars to the most popular candidate than the zip 3 average.

In Table 6.1 and Figure 6.1 we report the findings for our two dependent measures, with the observations grouped according to the success rate of new products purchased from MassStore. The unit of analysis is a (5-digit) zip code x election cycle x congressional district.[18]

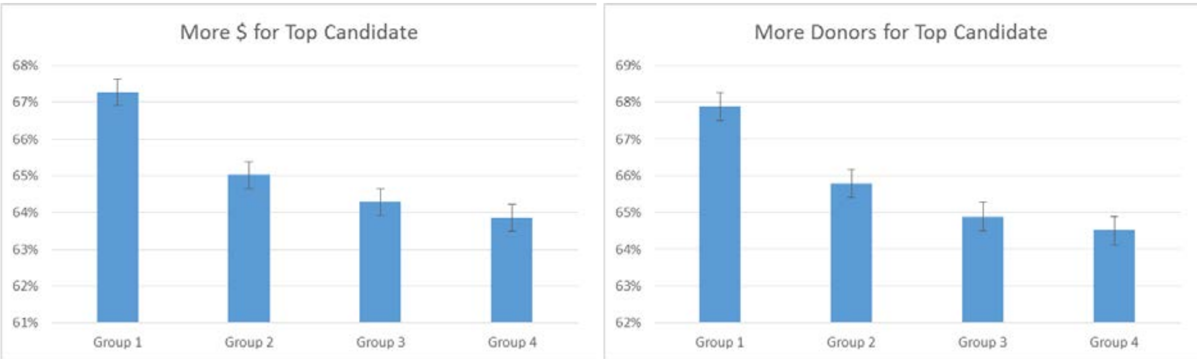
[18] For some congressional elections the total contribution was identical for the top two candidates in a 3-digit zip code. We omitted these observations. This was more likely to affect the number of donors analysis than the $ contribution analysis (and so the sample sizes are smaller for the number of donors analysis). We also restrict attention to zip codes with at least one donation to the top 2 candidates. This also contributes to a difference in sample sizes between the $ contribution and number of donors sample sizes (as the identity of the top 2 candidates in a 3-digit zip code depends upon which metric we are using). Finally, we also exclude a small number of observations in which a 3-digit zip code only had contributions from a single 5-digit zip code

Table 6.1. Donations to Political Candidates

	Group 1	Group 2	Group 3	Group 4
More $ for Top Candidate	67.271% (0.182%)	65.028% (0.186%)	64.288% (0.186%)	63.857% (0.187%)
Sample Size	66,202	65,749	66,250	65,675
More Donors for Top Candidate	67.883% (0.195%)	65.791% (0.196%)	64.894% (0.198%)	64.512% (0.199%)
Sample Size	57,558	58,386	58,027	57,946

The table reports the average of *More $ for Top Candidate* and *More Donors for Top Candidate*. The zip codes are grouped according to the success or failure of MassStore's new product transactions. The unit of analysis is a zip code x election cycle x congressional district. Standard errors are in parentheses.

Figure 6.1. Donations to Political Candidates



The figures report the average of *More $ for Top Candidate* and *More Donors for Top Candidate* as defined above. The zip codes are grouped according to the success or failure of MassStore's new product transactions. The unit of analysis is a zip code x election cycle x congressional district. Error bars indicate 95% confidence intervals.

The findings indicate that harbinger zip codes are less likely than other zip codes to donate to the candidates that receive the most donations in their 3-digit zip code. While 67.3% of zip codes in Group 1 contributed more total dollars to the most popular candidate in their regions, this proportion dropped to just 63.9% in Group 4. This relationship extends from both the $ amount contributed to also include the number of donors. For both measures the relationship is monotonic, with the proportion dropping consistently from Group 1 to Group 4.

In the previous analysis we grouped the zip codes into four buckets according to the average success rate of new products purchased at MassStore. We can also conduct the inverse analysis: comparing the average success rate when grouping the observations according to the binary election

for that election cycle (by construction the proportion contributed by the 5-digit zip code is the same as the 3-digit zip code in these observations).

contribution variables (*More $ for Top Candidate* and *More Donors for Top Candidate*). These results are summarized in Table 6.2.

These findings confirm that there is a significant relationship between the outcome of new products purchased by households in a zip code and contributions to congressional elections campaigns. Zip codes that contribute more to candidates who receive the most donations in that 3-digit zip code also tend to purchase new MassStore products that succeed. In contrast, harbinger zip codes both give less to these candidates and are more likely to purchase new MassStore products that fail. Notice that the t-values for these two difference calculations both exceed 10.

Table 6.2. Donations to Political Candidates

		Average Success Rate of New Products Purchased	Sample Size
Amount Contributed	More $ for Top Candidate	85.812% (0.007%)	171,819
	Fewer $ for Top Candidate	85.676% (0.009%)	92,057
	Difference	0.136%** (0.011%)	
Number of Donors	More Donors for Top Candidate	85.828% (0.007%)	152,523
	Fewer Donors for Top Candidate	85.694% (0.009%)	79,394
	Difference	0.135%** (0.012%)	

This table reports the average success rates of new products purchased at MassStore when grouping the observations according to *More $ for Top Candidate* and *More Donors for Top Candidate*. The unit of analysis is a zip code x election cycle x congressional district. Standard errors are in parentheses.

We have conducted this analysis using a zip code x election cycle x congressional district as the unit of analysis. As a robustness check we can also aggregate up to the zip code x election cycle level. These results are summarized in the Appendix. The dependent variables are now averages of the binary variables. Therefore, we also report the pair-wise correlation between these averages and the average success rate of new products purchased at MassStore. Reassuringly, this analysis replicates our earlier findings.

Contributions to Winning vs. Losing Candidates

An alternative approach to analyzing the political contributions data is to ask: are harbinger zip codes less likely to contribute to the winning candidate? To investigate this question we construct a binary variable Win_{cdt} indicating whether a candidate won the election for that congressional district in that electoral cycle. We then calculate the "Winning Percentage" as a weighted average of *Win*. In particular, we average across candidates and congressional districts to calculate a weighted average for each zip code in each election cycle. We weight by either the number of donors to each candidate (*Donor Wtd Win %*) or by the total dollar contribution from that zip code to each candidate (*Amount Wtd Win %*).

In Table 6.3 and Figure 6.2 we report these two outcome measures for the four groups of zip codes. We used the same approach to group the zip codes. The zip codes that purchased new product from MassStore that had the highest average success rate are in Group 1, with Group 4 containing the harbinger zip codes. We also report the pair-wise correlation between each measure and the average success rate of new products purchased at MassStore from customers in each zip code.

Table 6.3. % of Winning Candidates

	Group 1	**Group 2**	**Group 3**	**Group 4**	**Correlation**
Amount Wtd Win %	63.45% (0.27%)	62.24% (0.28%)	62.17% (0.28%)	61.61% (0.31%)	0.0251**
Donor Wtd Win %	63.20% (0.26%)	62.13% (0.26%)	62.07% (0.27%)	61.43% (0.30%)	0.0252**
Sample Size	12,332	12,342	12,322	12,331	49,327

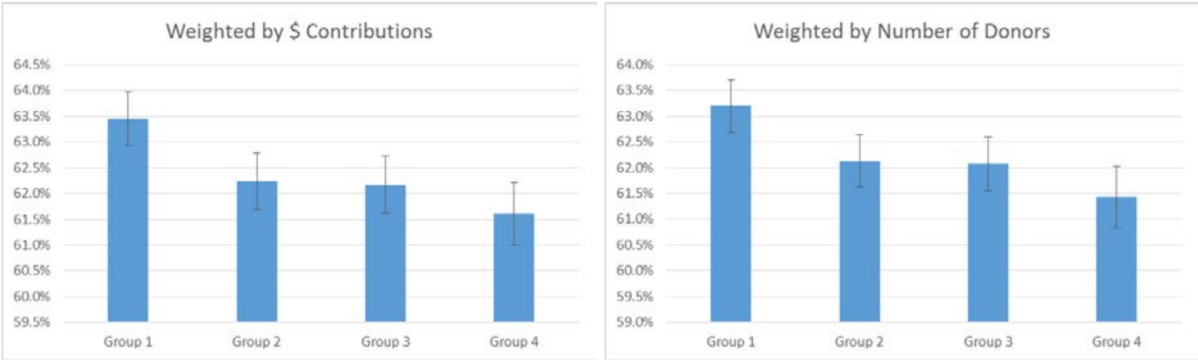
The table reports the proportion of donations that were contributed to candidates that won their congressional elections. The unit of analysis is a zip code x election cycle. We report the average proportion when weighting by either $ contributions or number of donors in that zip code. The zip codes are grouped according to the success or failure of MassStore's new product transactions. Standard errors are in parentheses. We also report the pair-wise correlations between the winning percentages and the new product success rate at MassStore.

The findings indicate that harbinger zip codes are systematically less likely to donate to winning candidates. Compared to other zip codes a higher proportion of donors in these zip codes give money to congressional candidates that lose. A higher proportion of total dollar contributions are also given to candidates that lose (compared to other zip codes). The relationships are monotonic, with the percentage of winning candidates declining consistently from Group 1 to Group 4.

Summary

We have shown that harbinger zip codes identified from new product purchases at MassStore tend to support congressional candidates that neighboring zip codes are less likely to support. They also support congressional candidates that are less likely to be elected than other zip codes. This confirms that harbinger zip codes can explain variation in choices that extend beyond retail settings. In the next section we compare changes in house prices in harbinger zip codes and neighboring zip codes.

Figure 6.2. % of Winning Candidates



The figures report the proportion of donations that were contributed to candidates that won their elections. The zip codes are grouped according to the success or failure of MassStore's new product transactions. The unit of analysis is a zip code x election cycle. Error bars indicate 95% confidence intervals.

6. Change in House Prices

In Section 4 we used ApparelCo customers that changed zip codes to investigate whether harbinger zip codes result from harbinger households choosing to cluster together, or from households changing their preferences when they move into harbinger zip codes. The findings strongly supported the first explanation. Households that move from a harbinger zip code tend to move to another harbinger zip code, while non-harbinger zip codes do the reverse. We illustrated how this could contribute to the formation of harbinger zip codes using an example of ocean views. If non-harbingers like ocean views but harbingers do not then we would expect to see clusters of harbingers in zip codes without ocean views.

By choosing where they live, households also determine (in part) how much their home values will increase. In this section we investigate whether harbinger zip codes experience larger or smaller house price increases. In particular, we investigate the relationship between the average success rate of new products purchased at MassStore and the average change in house prices (by zip code).

The decision to live in a harbinger zip code may reflect more than just a preference for the location. It may also reflect a preference for the housing stock within that location, and this could also contribute to the rate at which house prices change. For example, a recent study by Trulia (reported in Bloomberg 2016) discovered that customers who buy unusual houses suffer smaller house price increases. Not only are "McMansions" unattractive to look at, their prices increases at slower rates than other houses. If unusual preferences lead customers to purchase unusual houses, then these preferences may also impact home values.

The findings in this section are different in several respects than the findings in the previous sections. First we focus on the housing market rather than product choices or political decisions. Second, while the datasets used in the previous sections document explicit customer decisions, the connection between changes in house prices and household decision-making is less direct. Changes in house prices are measured using transactions, and so changes in house prices are related to the pricing decisions of some households.[19] However, the decision that is perhaps most relevant to this study is the choice of which zip code to live in. Households make this decision when they first enter the zip code (unless they inherit the property), and at least implicitly again when they decide to continue to live there.

A third difference between the findings in this section and the results in the previous section is that it is difficult to argue that the results are solely attributable to unusual customer preferences. We do not have any data describing features of the location or housing stock that reflect unusual preferences. This means that there are a range of intervening factors that could contribute to the findings, including for example the earlier evidence that harbinger zip codes tend to be more rural. Therefore, rather than trying to establish the mechanism that drives our results, we focus instead on documenting a surprising relationship between the adoption of successful (or unsuccessful) new products at MassStore and changes in house prices. We also conduct extensive robustness checks to determine the limits of this relationship.

We begin by illustrating the general trend in house prices using Zillow's US median price index.[20] We use the Zillow data to calculate the year-on-year *Price Change* in zip code z in year t:

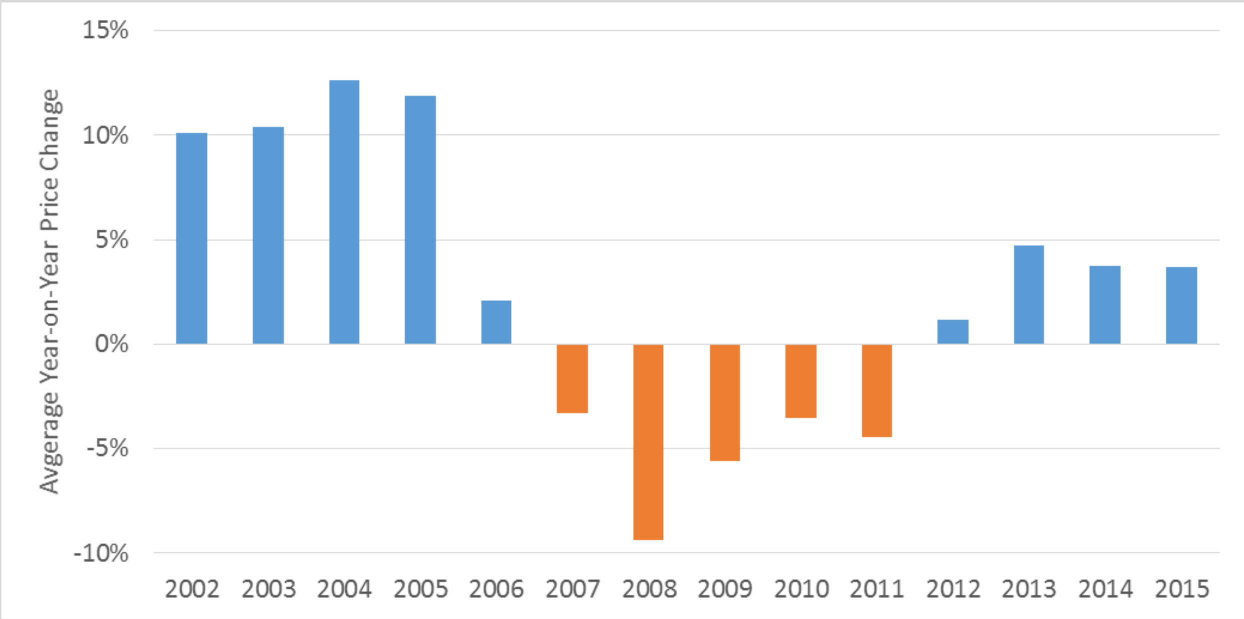
$$Price\ Change_{zt} = \frac{Price\ End_{zt} - Price\ Start_{zt}}{0.5 * Price\ End_{zt} +\ 0.5 * Price\ Start_{zt}}$$

Using the average of the two prices in the denominator ensures that price increases and decreases are treated symmetrically. A positive (negative) value of *Price Change* indicates a price increase (decrease). We then calculate the average of this *Price Change* for each year by averaging across the zip codes. The averages are calculated using a common sample of 4,291 zip codes for which both Zillow data was available every year and the MassStore data provided an average success measure. The average year-on-year *Price Change* is reported in Figure 7.1.

[19] Obviously many households in a zip code do not participate in the transactions that determine the change in house prices. However, it is also true that many of the households in each zip code do not purchase from MassStore and ApparelCo, and do not donate to congressional election campaigns.

[20] Available at: http://files.zillowstatic.com/research/public/State/State_Zhvi_Summary_AllHomes.csv

Figure 7.1. Year-on-Year Change in Zillow Median House Price Index



The figure reports the year-on-year Price Change averaged across zip codes. The averages are calculated using a common sample of 4,291 zip codes for which both Zillow data was available every year and the MassStore data provided an average success measure.

We see that 2002 and 2006 was a period during which house prices increased on average, before the fall in prices from 2007 to 2011. House prices then recovered between 2012 and 2015. We might expect that the zip codes in which house prices increased the most between 2002 and 2006 are also the zip codes in which house prices decreased the most between 2007 and 2011. Therefore, we will investigate changes in house prices across three separate periods: January 2002 to December 2006, January 2007 to December 2011, and January 2012 to December 2015. For each zip code we construct *Price Change*$_{zt}$ using prices at the start and end of these multi-year periods and use them as dependent variables to re-estimate Equation 5.1 (which we re-state below):

$$\text{Price Change}_z = \alpha + \beta_1 \text{ Average Success Rate}_z + \boldsymbol{\beta} \textbf{ 3-Digit} + \varepsilon_z \qquad (5.1)$$

The unit of analysis is a zip code and *Average Success Rate*$_z$ is the average success rate of new products purchased from MassStore by households in zip code *z*. The **β 3-Digit** term denotes fixed effects identifying each 3-digit zip code. This controls for regional changes in house prices. Under this specification, the coefficient of interest is β_1, which measures how changes in house prices vary within a 3-digit zip code according to the *Average Success Rate* of new product purchases at MassStore.[21]

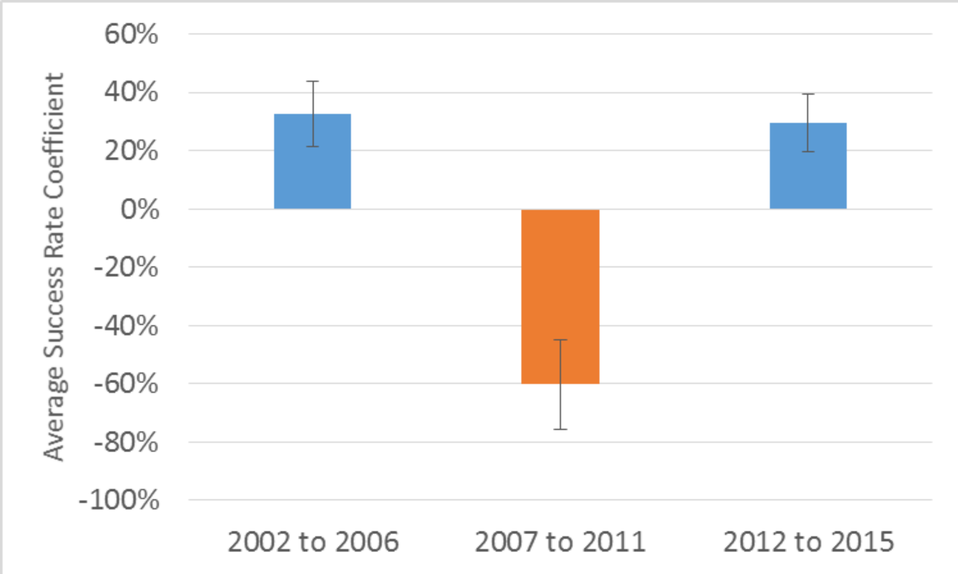
We re-estimate Equation 5.1 separately using the *Price Change* measures calculated for each of the three multi-year periods. Because each model focuses on a single multi-year period, the coefficient

[21] The inclusion of the fixed effects again ensures that β_1 is only identified by variation in house prices within each 3-digit zip code (and not by regional variation across 3-digit zip codes).

of interest (β_1) is only identified by cross-sectional variation in the *Average Success Rate* (i.e. these are not time series models). All three models are estimated using the same sample of 4,291 zip codes.

The coefficients of interest are summarized in Figure 7.2 and Table 7.1. They indicate that during periods in which house prices increased on average, the harbinger zip codes tended to have smaller house price increases. In particular, a 10% smaller *Average Success Rate* is associated with a 3.3% smaller house price increase between 2002 and 2006, and a 2.9% smaller increase between 2012 and 2015.

Figure 7.2. Relationship Between House Price Changes and Average Success Rates at MassStore



The figure reports the *Average Success Rate* coefficients when estimating Equation 4 using three different time periods. The unit of analysis is a zip code and the sample size in all three models is 4,291. The error bars indicate 95% confidence intervals.

Not surprisingly these zip codes also experienced a smaller house price decrease when average house prices fall from 2007 to 2011. This is consistent with these zip codes experiencing a smaller run up in house prices between 2002 and 2006. We will focus our attention on price changes during the two periods that average house prices increased, and investigate the robustness of our finding that harbinger zip codes enjoy smaller house price increases during these periods.

One explanation for this finding is that these are the zip codes that had lower home values at the start of this period. This turns out to be correct: the house prices of harbinger zip codes at the start of 2002 were lower on average than prices in other zip codes. However, lower prices at the start of 2002 does not imply larger price increases between 2002 and 2006. To understand this relationship, we included starting prices as a control variable in the regressions.[22] The pattern of findings is unchanged (the coefficients of interest are reported in Table 7.1). Moreover, during periods average prices increased, the increases were actually larger in regions with lower initial prices (the *Starting*

[22] Starting prices appear in both the left and right hand side variables. This is not unusual in a difference model.

Price coefficients are negative). This indicates that the smaller increase in house prices in harbinger zip codes occurs despite their lower starting values, not because of it.

Table 7.1. Relationship Between House Price Changes and Average Success Rates at MassStore

	2002 to 2006	2007 to 2011	2012 to 2015
Base Model	32.61%**	-60.22%**	29.39%**
	(5.72%)	(7.82%)	(5.08%)
Robustness Checks			
4-digit zip Fixed Effect	26.27%**	-61.14%**	25.19%**
	(6.62%)	(8.96%)	(5.91%)
Controlling for Starting Price	21.46%**	-44.57%**	31.52%**
	(5.47%)	(7.47%)	(5.10%)
Adding Demographic Controls	16.90%**	-36.72%**	17.81%**
	(5.80%)	(8.12%)	(5.50%)
All	19.93%**	-40.80%**	13.76%*
	(6.67%)	(9.07%)	(6.25%)

The table reports the *Average Success Rate* coefficients when estimating Equation 4 using three different time periods and control variables. The unit of analysis is a zip code and the sample size for all of the models is 4,291, except for the models including demographic controls where missing observations reduce the sample sizes to 4,085. Standard errors are in parentheses.

Our second robustness check replaces the fixed effects identifying each 3-digit zip code with fixed effects identifying each 4-digit zip code. This reduces the price variation to even smaller regions, with at most 10 (5 digit) zip codes within each 4-digit zip code. The pattern of findings is again replicated (the coefficients of interest are also reported in Table 7.1).

Finally, we recognize that the relationship between new product failure rates at MassStore and changes in house prices is very unlikely to be causal. Smaller house price increases is unlikely to have caused these households to purchase new product flops at MassStore. Instead, there is almost certainly an unobserved intervening variable(s) that explains which zip codes are harbinger clusters, and which zip codes had the smallest house price increases. As we discussed, in the previous sections of this paper we have interpreted this unobserved variable as "unusual preferences". However, with this house price data it could represent a range of different underlying explanations, including perhaps household income. To investigate this possibility we add the demographic variables as control variables to our model. The coefficients of interest are reported in Table 7.1, and complete findings are reported in the Appendix. The pattern of findings survives with the inclusion of these controls. This suggests that the effect is unlikely to be solely attributable to income, age, the degree of urbanization, or any of the other control measures.

Summary

In this section we showed that during periods in which house prices were generally increasing harbinger zip codes experienced smaller house price increases than other zip codes. This is a robust finding that survives controlling for starting prices, controlling for a wide range of demographic variables, and limiting identification to variation within 3-digit and 4-digit zip codes.

We attributed the findings in the previous sections to unusual preferences; households in harbinger zip codes appear to have preferences that are different from households in neighboring zip codes. This may also contribute to the findings in this section. Harbingers may have different preferences for houses (the "McMansion" effect identified by Trulia) and may also have different preferences for house locations (they may not like ocean views). However, we also recognize that the findings in this section remain open to a broader set of interpretations. While it is possible that unusual preferences contribute to variation in house prices, we recognize that there are other unobservable differences that could contribute to the findings.

7. Conclusions

Using data from multiple sources, we have shown that the phenomenon of harbingers is surprisingly widespread. We begin by showing that there exist harbinger zip codes. Households in these zip codes are more likely than households in other zip codes to purchase new products that fail. Their adoption of a new product is a signal that the new product will fail. We interpret this finding as evidence that households in these zip codes have preferences that are not representative of households in other zip codes. We then show that the evidence of unusual preferences extends across retail product categories and across retailers. We also extend the findings to very different domains, including donations to federal election campaigns and variation in house prices.

What makes these results particularly surprising is that while we measure the average outcome for a zip code, relatively few households in each zip code participate in each decision. Not every household purchases from the retailers that we study, and relatively few households contribute to congressional election candidates. Moreover, the households that participate will often be different households for each decision. It is unlikely that the households who purchase from one retailer are all the same households that purchase from the other retailer. They are also unlikely to all make donations to congressional election campaigns. Despite this, we observe significant similarities in zip code level decisions across these different purchasing contexts.

We explore two explanations for why households with unusual preferences cluster together. This analysis uses a sample of households that changed zip codes. The analysis reveals that households that moved from a harbinger zip code tended to move to another harbinger zip code. Similarly, households that started in a non-harbinger zip code generally moved to another non-harbinger zip code. This suggests that harbinger zip codes results at least in part from customers choosing to cluster with other households that have similar preferences. We did not find any support for the alternative explanation that customers learn their preferences when they move into a harbinger zip code. It appears instead that harbinger preferences are relatively sticky and that households bring their preferences when they change zip codes, rather than learning them when they get there.

The findings have important implications for the application of the harbinger effect in different settings. First, a lot more data is available at the zip code level than at the household level. This makes it much easier to apply the findings to different decision contexts. Second, the evidence that the harbinger zip codes identified through new product purchases at one firm explains variation in decisions at other firms and across housing and political settings suggests that we do not need to separately identify which zip codes are harbinger zip codes for each decision. Instead, identifying harbinger zip codes in one setting is sufficient to explain variation in other settings.

We have shown the predictive power of these harbinger zip codes using data spread over several years. Future research could investigate how the predictive power of this identification diminishes (or increases) over time as households move and environments change.

8. References

Anderson, Eric, Song Lin, Duncan Simester, and Catherine Tucker (2015), "Harbingers of Failure," *Journal of Marketing Research*, 52(5), 580-592.

Anderson, Eric, Chaoqun Chen, and Blake McShane (2017), "Harbingers of Failure … and Success," working paper, Kellogg School of Management.

Anderson, Chaoqun Chen, Caiyun Liu and Duncan Simester (2017), "Do Harbinger Products Signal Which New Customers Will Stop Purchasing?" working paper, MIT Sloan School of Management.

Anderson, Evan E. (1979), "An Analysis of Retail Display Space: Theory and Methods," *Journal of Business*, 52(1), 103-118.

Bass, Frank M. (1969), "A New Product Growth Model for Product Diffusion," *Management Science*, 15(5), 215-227.

Bloomberg (2016), "McMansions Define Ugly in a New Way: They're a Bad Investment," August 23, 2016, http://finance.yahoo.com/news/mcmansions-define-ugly-way-bad-142534896.html.

Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Mohammad S. Rahman (2009), "Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition," *Management Science*, 55(11), 1755–1765.

Brynjolfsson, Erik, Yu (Jeffrey) Hu and Duncan Simester (2011), "Goodbye Pareto Principle, Hello Long Tail: The Effect of Internet Commerce on the Concentration of Product Sales," *Management Science*, 57(8), 1373-1387.

Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Michael D. Smith (2003), "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science*, 49(11), 1580–1596.

Choi, Jeonghye and David R. Bell (2011), "Preference Minorities and the Internet," *Journal of Marketing Research*, XLVIII, August, 670-682.

Mahajan, Vijay, Eitan Muller and Frank M. Bass (1990), "New Product Diffusion Models in Marketing: A Review and Directions for Research," *Journal of Marketing*, 54(1), 1-26.

Pandya, Sonal S. and Rajkumar Venkatesan (2016), "French Roast: Consumer Response to International Conflict—Evidence from Supermarket Scanner Data," *Review of Economics and Statistics*, 98:1, 42-56.

Waldfogel, Joel (2003), "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated Product Markets," *RAND Journal of Economics*, 34(3), 557-568.