How Video Rental Patterns Change as Consumers Move Online

Alejandro Zentner[†], Michael D. Smith[‡], Cuneyd Kaya[†]

azentner@utdallas.edu, mds@cmu.edu, cckaya@utdallas.edu

This Version: October 2012

Acknowledgements: The authors thank an anonymous video chain for generously providing data to support this research and Venkata Tumuluri, Samita Dhanasobhon, and Ketaki Potkar for providing outstanding research assistance. Smith acknowledges generous support from Carnegie Mellon's iLab and Zentner acknowledges generous support from the Center for the Analysis of Property Rights and Innovation (CAPRI). The authors are listed in reverse alphabetical order.

[†] University of Texas at Dallas
[‡] Heinz College, Carnegie Mellon University

How Video Rental Patterns Change as Consumers Move Online

ABSTRACT

How will consumption patterns change when consumers move from brick-and-mortar to Internet markets? If consumers purchase more niche products online than at brick-and-mortar stores, does this have something to do with the channel or is this solely due to selection effects from the types of consumers who decide to use the Internet channel or the types of products that consumers select to purchase online?

We address this question using customer-level panel data obtained from a national video rental chain as it was closing many of its local stores. This allows us to observe how individual consumers change their behavior when they are forced to move from brick-and-mortar stores to online consumption.

Our results suggest that when consumers move from brick-and-mortar to online channels they are significantly more likely to rent "niche" titles relative to "blockbusters." This suggests that a significant amount of niche product consumption online is due to the influence of the channel on consumers, not just due to selection effects.

Keywords: *long tail, DVD rentals, natural experiment, empirical estimation.*

1. Introduction

A variety of papers and articles have documented large differences between the types of products purchased online and offline (e.g., Brynjolfsson, Hu, and Smith 2003; Anderson 2006; Brynjolfsson, Hu, and Simester 2011). However, these differences in purchase patterns across channels could be solely due to selection effects: heterogeneous consumers may sort into channels based on their tastes, or consumers may choose the channel based on the types of products they want to purchase. Our objective in this paper is investigating whether observed differences in consumption patterns between online and offline markets are solely due to selection, or whether the nature of the channel has some impact on consumer's choices.

Answering this question requires the use of customer-level panel data on online and offline purchases and an exogenous shock affecting consumers channel choices. In this paper we use just such a dataset for DVD rentals. We base our empirical analysis on household-level panel data from a large video rental chain that closed many physical locations during our study period. The market for DVD rentals has traditionally exhibited "superstar" effects, where a few top-selling products take the majority of all revenues (Rosen 1981); our focus is to empirically examine how the introduction of online commerce has changed the share of transactions taken by superstar versus niche DVDs.

In terms of descriptive statistics, we note that the top 100 most popular DVDs make up 85% of in-store rentals for our focal company but account for only 35% of the company's online rentals. However, we cannot use these simple statistics to conclude that online markets change consumer behavior because these statistics may be solely due to selection effects. Our approach to study whether online markets change consumer behavior is to examine how household-level rental patterns for popular and niche titles change when the exit of physical stores forces consumers to move from offline to online channels. Specifically, we use the exit of physical stores as an instrumental variable for the online versus offline channel choice. This instrument exploits transportation cost changes experienced by individuals located near the exiting stores. Our findings indicate that

characteristics of the online channel cause superstar DVD titles to take a smaller share of the market as consumers shift from offline to online marketplaces.

From the standpoint of theory, online markets may (or may not) transform markets that have traditionally exhibited "superstar" effects. Various supply-side or demand-side mechanisms can cause long tail or superstar markets (Brynjolfsson, Hu, and Smith 2006; Brynjolfsson, Hu, and Smith 2010). For example, the selection of products available from the Internet channel is often much wider than the selection available at physical stores. Online marketplaces can offer a larger selection of products than traditional physical stores can because the online channel has lower storage and inventory costs and there are no shelf space limitations. As a consequence, the concentration of overall sales across products may tend to decrease as transactions shift from offline to online channels.

Online channels could also change consumers' product choices even when the sets of products offered online and offline are identical. In part, this might happen because the ways consumers search for products online and offline are fundamentally different. At physical stores, finding a popular product may be easier than finding a niche product, even when both are available. Popular products typically occupy more prominent shelf space in physical stores versus niche products that are relegated to less visible positions. In online marketplaces, search tools may be used to promote the discovery of niche titles tailored to individual customers' preferences. However, personalization and recommendation engines and other search tools could also increase the concentration of product sales. For example, top 10 seller lists may tend to reinforce the popularity of already popular products. Similarly, recommendation systems may increase the concentration of product sales because they base their recommendations on actual sales and there is limited data for products that have low historical sales (Fleder and Hosanagar 2009; Oestreicher-Singer and Sundararajan 2011).

The answer to how consumption patterns change when consumers move online is important for both the academic literature and for managerial practice. If the observed differences between online and offline markets are solely due to selection, without an effect on aggregate consumption, then there is no need for producers to change their

behavior—and in our context in particular, motion picture studios should continue to focus on producing blockbuster titles. However, if using online markets changes consumer behavior, then producers may wish to reexamine their current strategies and shift their production toward more "long tail" products.

2. Literature

Our results contribute most directly to a small empirical literature studying the effect of information technology on sales concentration patterns. While the "Long Tail" was considered one of the best ideas of 2005 by industry observers (Businessweek 2005), it is important to note that there is no general agreement in the academic literature regarding how online commerce will affect the concentration in product sales. In this literature, Elberse and Oberholzer-Gee (2007) use aggregate data by title to study how online commerce affected the distribution of sales in the United States' home video industry from 2000 to 2005. They find that, although the number of product choices increases, by the end of their study period superstar products comprised a larger proportion of sales than before.[1] Brynjolfsson, Hu, and Simester (2011) examine the concentration of product sales for a retailer of women's clothing selling through both Internet and catalog channels. Using cross sectional data on sales, aggregated by item and channel, they find that the concentration of product sales is lower for the Internet channel than for the catalog channel.

Our paper is also related to the research examining grocery shopping using household-level data for households that shop interchangeably at online and offline stores from the same grocery chain (e.g., Chu, Chintagunta, and Cebollada 2008). In this literature, our paper is most closely related to Pozzi (2012), who finds that brand exploration is more prevalent in physical stores than online. As consumption goods, however, groceries are substantially different than DVD rentals because groceries are typically consumed more repeatedly as compared to DVDs.[2]

[1] Partly motivated by this result, Bar-Isaac, Caruana, and Cuñat (forthcoming) formulated a model in which a reduction in search costs generates both superstar and long tail effects.

[2] Our results also contribute to a growing literature on the impact of popularity and recommendation information on online sales of niche and popular (Tucker and Zhang 2011, Fleder and Hosanagar 2009, and

Our examination of superstar and long tail effects uses individual-level panel data including information on consumers' transactions from both online and brick-and-mortar channels. We use these data to analyze how individuals change their consumption patterns when they are induced by store closures to move from in-store to online consumption.

3. Data and Setting

Our data come from a large video rental company that operates both brick-and-mortar stores and online DVD rental channels. For a monthly flat rate subscription, customers in our data can rent DVDs online and receive them in the mail, and then exchange these DVDs either through the mail or at a physical store.

The selection of DVD titles available for rental at these physical stores is a subset of the selection of titles available for rental online. While a typical store has a rotating selection of approximately 2,000 titles, the online channel has over 100,000 titles. The Internet channel has a much larger DVD selection than the selection available at physical stores because the online channel has lower storage and inventory costs.[3] Due to these storage capacity limitations, our focal company's physical stores stock more copies of new releases than of older titles. Inventory costs are also lower online than in physical stores because the company we study ships DVDs to its customers from a small number of centralized warehouses, compared with a substantially larger set of physical stores. Thus, as these shipping locations reach a much larger number of consumers than a physical store would, the law of large numbers indicates that the company can reduce inventory costs by more accurately predicting demand from the online channel.

Our data cover DVD rental activity from both the online and in-store channels for all subscribing customers, and include more than 49 million rental transactions for the thirty

---

Oestreicher-Singer and Sundararajan 2011). In contrast to our study, these studies do not examine sales from physical stores or cross channel choices. Our paper is also related to Waldfogel (2012) who documents a decrease in the degree of music sales concentration in a few artists.

[3] Storage costs are even lower for video streaming services; however, video streaming was in its infant stages of development during our study period, and our focal company did not offer a video streaming service during our period of analysis.

week period from October 2, 2009 through April 29, 2010. Although consumers without a monthly subscription can rent DVDs from our company's physical stores, our data only include the information from consumers with a monthly subscription that allows the rental of an unlimited number of DVDs.[4] Our customers maintain an online queue of DVD titles they wish to watch, and when they return a DVD, the company sends the next DVD title from that queue to the subscriber's home.

Our data include the renting subscriber, DVD title, transaction date, and whether the DVD was delivered by mail or exchanged at a physical store. In addition, we have the zip code for each subscriber, the address for each physical location operated by this firm, and the closing date for the locations that were closed during our study period. For approximately 56% of the subscribers we also have exact addresses, which, when combined with the address for each store in our sample, allows us to calculate the distance between each of these customers and their closest store.[5]

Table 1 presents the summary statistics for our data. The DVD rental market has experienced important changes during the last decade. Industry trends show that traditional physical stores have been displaced by online DVD rental services, and more recently by video streaming services and by physical kiosks. We do not know exactly how the number of subscribers changed during our period of analysis, because subscribers may not rent every week and we do not have a list of subscribers indicating when they signed up for service or canceled their subscriptions. But the decrease in the number of subscribers renting DVDs by the end of our study period shown in Table 1 suggests that the number of subscribers decreased during this period. Our data come from a company that closed 15.2% of its physical stores during our period of analysis (see the last column of Table 1). Our focal company did not open new physical locations during the thirty-week study period, and thus the number of physical store locations is entirely driven by store exit. The substantial change in the number of physical rental stores will

[4] These subscription plans include a restriction on how many DVDs a subscriber can have out at any given time.

[5] We have exact addresses for all consumers who signed up for service after January 1, 2008. While this is not a random sample of the subscribers, using these addresses we are able to provide a finer-grained examination of how transportation costs influence channel selection than using the zip codes for all the subscribers.

play a central role in our identification strategy since we will use store exit as an instrumental variable for online versus offline channel selection by consumers.

Rentals via mail represent 68.3% of all rentals, and DVD exchanges at physical stores represent the remaining 31.7%. Averaging our information across subscriber-week observations with positive DVD rentals (subscribers may not rent every week and our data only record the rental instances), subscribers rented an average of 2.25 DVDs per week: 0.71 DVDs from the store and 1.54 DVDs by mail.

Table 1: Summary Statistics

Week	Number of Transactions	Number of Subcribers with Positive Rentals	Number of Stores
Week 1	1,547,158	739,696	3,356
Week 2	1,511,517	721,494	3,356
Week 3	1,520,978	727,174	3,356
Week 4	1,580,138	743,421	3,354
Week 5	1,526,952	690,849	3,325
Week 6	2,212,001	798,357	3,324
Week 7	1,764,180	735,643	3,300
Week 8	1,590,305	721,258	3,072
Week 9	1,736,078	786,002	3,066
Week 10	1,607,267	735,969	3,066
Week 11	1,660,696	750,181	3,066
Week 12	1,725,464	764,054	3,065
Week 13	1,661,685	740,716	3,047
Week 14	1,744,056	770,986	3,047
Week 15	1,765,719	775,946	3,046
Week 16	1,693,175	748,035	3,046
Week 17	1,732,115	755,332	3,033
Week 18	1,733,123	758,485	3,033
Week 19	1,629,584	725,281	3,033
Week 20	1,620,146	726,576	3,032
Week 21	1,633,609	717,592	3,032
Week 22	1,637,397	721,227	2,860
Week 23	1,648,239	722,855	2,859
Week 24	1,625,214	713,490	2,858
Week 25	1,656,037	715,472	2,858
Week 26	1,617,684	702,461	2,844
Week 27	1,478,658	662,682	2,844
Week 28	1,441,504	650,747	2,844
Week 29	1,442,232	648,772	2,843
Week 30	1,432,050	645,485	2,843

Importantly, a monthly subscription fee allows the rental of a certain number of DVDs at a time, but subscribers do not pay a price each time they rent a DVD from either the online or offline channel. Thus, for subscribers, the DVD rental price is neither a driver of the online versus offline channel selection, nor of the specific DVD title choice.

3.1 Popular and Niche Product Definitions

A stream of prior research, while focusing on examining the potential of information technologies to transform the distribution of sales across products and channels, has classified products as either niche or popular. Products are typically classified as niche when they are less likely to be stocked in physical stores, or are only available after incurring a high search cost. In spite of this definition, because of data restrictions, prior studies classified products as niche or popular based primarily on product sales from the online channel, and not on product sales from the offline channel. Our data have the advantage that they allow us to define the popularity of DVD titles during a week using information from both online and offline rental channels.

Classifying goods as either niche or popular based on online sales might be problematic if the distribution of sales across products online and offline are different. We know that firms choose which products to stock online and offline, and furthermore know that product availability by channel may influence consumers' channel choices. As an example of the possible problems that can occur when classifying products based solely on online sales, suppose that consumers buy a product online only when this product is not available at the physical store. If this situation is common for a given product then this product could be classified as popular using online sales, even though many consumers are buying it online precisely because it is not available in physical stores and therefore could be considered a niche product. Additionally, some online retailers, because they face low competition from physical stores, may specialize in selling only niche products that are less likely to be available at physical stores. For these retailers classifying top-selling products as popular and the remaining products as niche may be incorrect.

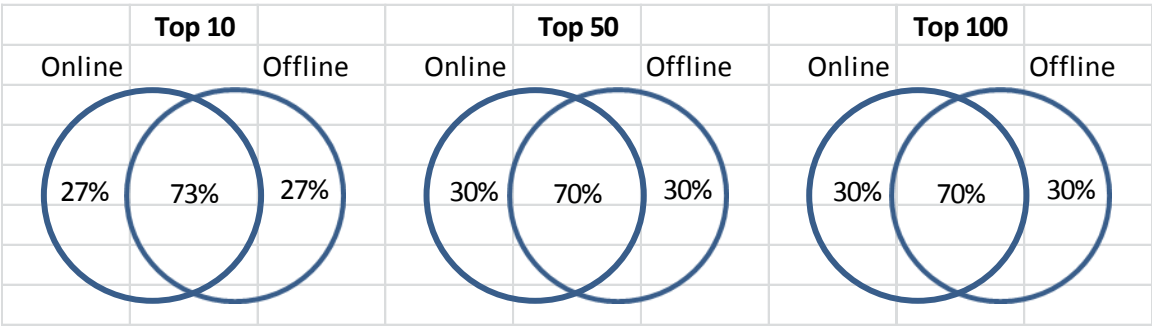
Using our data, we can only provide comparisons of transactions online and offline for a single product category (DVD rentals) and within a single firm. However, within this major firm, we can compare the extent to which the selection of DVDs rented online is different than the selection of DVDs rented offline. We do this by using online and offline data aggregated at the national level, and computing the total number of rentals

for each DVD title and each channel during each week. We then rank DVDs by popularity, computing two separate weekly ranks of DVDs using either online or offline rental information.

Our definition also allows for the popularity of a DVD to vary from week to week: a DVD that is popular at the beginning of the study period can become niche by the end the study period since DVDs have short-lived popularity cycles. For example, 91 (228, 347) different DVD titles are among the top 10 (top 50, top 100) DVD titles for at least one week during our thirty-week study period.

Figure 1 shows that the rankings of DVD rentals computed using online rentals are somewhat different than the rankings of DVD rentals computed using offline rentals. For example, comparing the top 10 DVD titles based on either online and offline rental information shows that a weekly average of 7.3 DVD titles are included in both rankings, but 2.7 DVD titles are included in only one ranking. Similarly, for the top 50 and top 100 DVDs approximately 70% of the titles are included in both rankings and the remaining 30% of the DVD titles are included in only one ranking.

Figure 1: Commonality Between Online and Offline Popular Titles



Differences between the online and offline rankings of DVD titles may be due to selection effects and may also be due to other cross-channel differences on the demand side or on the supply side. For example, these differences could be driven by preference heterogeneity between consumers who disproportionately choose the online versus the offline channel, different display and promotional activities across channels, or differences in the selection and inventory of titles available from the offline and online channels. Using our data it is difficult to disentangle the degree to which each factor may

contribute to the differences between online and offline rankings. This is partly because we do not observe inventory or title assortments online and offline, and moreover title assortments may vary across our company's physical stores.[6]

In this section we have noted that the specific titles that are popular in the online channel are somewhat different than the titles that are popular in the offline channel. Using our data we can generate popularity lists based on both online and offline rentals, and we will examine whether popular titles defined in this way change as consumers move online.

3.2 Superstars: Online versus Offline DVD Rental Distributions

In Table 2 we see that superstar DVD titles take a substantially larger share of total rentals offline than they do online. For example, the top 100 DVD titles in our sample represent 84.6% of in-store rentals, but only 35.1% of online rentals.

Table 2: Cumulative Share of Weekly Top Ranked DVD Rentals by Channel

	Rented Online and Delivered by Mail	Rented from the Physical Store
Top 10	11.0%	47.7%
Top 50	26.4%	77.6%
Top 100	35.1%	84.6%

Although the statistics in Table 2 may be suggestive of how consumption patterns change when consumers move from offline to online markets, we must be cautious when interpreting Table 2. From these statistics alone we cannot conclude that online commerce decreases the superstar nature of the DVD rental market, since these differences in rental patterns across channels could be solely due to selection effects. For example, different rental concentrations may be due to heterogeneous tastes of the consumers who rent primarily online versus consumers who rent primarily offline. These differences could also be explained by selection effects due to consumers' impatience. If consumers are impatient regarding watching a newly-released DVD and do not wish to

[6] Having acknowledged the limitations of our data for distinguishing between alternative explanations for the differences in the titles included in the rankings online versus offline, Table EA1 in the extended online appendix may provide some relevant information regarding the selection of titles available from each channel.

wait for the DVD to arrive in the mail, then consumers who typically rent from both channels may select to rent from the physical store more often than from the online store when they wish to watch popular versus non-popular DVDs.

Our objective in this paper is to examine whether consumption patterns change as consumers move online, or whether the observed differences in online and offline consumption are primarily due to selection effects. Although in this paper we do not seek to identify why consumers change consumption patterns when they move online, beyond selection effects, the differences in the statistics in Table 2 may be an effect of the channel. We argued in the introduction that the literature has identified a variety of long tail effects arising from both demand and supply factors. Some demand and supply factors are more specific to our context. For example, the focal company's different display and promotional activities across channels may partly explain the statistics in Table 2. Popular products occupy a disproportionate amount of prominent shelf space in the company's physical stores as compared to the online channel. In addition, the queue system for video rentals may also partly explain the statistics in Table 2. Consumers who do not frequently update their online queue of DVDs may end up watching older and less popular titles when they move to the online channel.[7]

[7] Two other possible "channel related" explanations for our results are worth mentioning. First, one might wonder whether our results are influenced by stock-outs being more common online than in stores, driving customers to long tail titles because of the unavailability of (otherwise preferable) popular titles online. To partially test for this supply side mechanism, we monitored stock-outs for a matched set of newly released "popular" titles at three physical stores and through the company's online channel. Our brief examination showed that, if anything, stock-outs of popular titles are somewhat more common in physical stores than online, suggesting that our results are not due to this supply side mechanism. Similarly, one might wonder whether the increase in online subscribers over time as stores closed increased the number of stock-outs for popular titles online relative to what these customers would have experienced in physical stores, leading to a similar shift in consumption unrelated to customer preferences. While we do not observe inventory levels in our data and thus cannot strictly rule out this possibility, we confirmed that our focal company's policy was to shift physical inventory of popular titles from closing stores to the online channel (while selling off some "long tail" inventory when a store closed). This shift, combined with the increased availability associated with a single centralized warehouse/queue versus multiple queues in physical stores should, if anything, have increased the relative availability of popular titles online as stores closed. Moreover, even if the increase in online usage led to a temporary reduction in the availability of popular titles, our focal company's access to wholesale DVD suppliers should have allowed them to respond to this shift quickly by ordering more DVDs. Finally, as we noted above, our data suggest that some customers may have cancelled their subscription when their local stores closed which, combined with the shift of popular DVDs from physical stores to online inventory should have further increased the relative availability of popular DVDs online relative to in stores

We now present our empirical approach for examining whether consumers change rental patterns as they move online, or whether the differences in online versus offline consumption are primarily due to selection effects.

4. Econometric Model

As noted above, our objective is to study whether changes in which rental channel consumers use affects their selection of DVD rental titles, and establishing whether online markets affect the consumption of superstar versus long tail DVD titles. We showed in Table 2 that the distributions of DVD rentals online and offline are quite different. Superstar DVD titles in particular take a substantially larger share of all rentals made in physical stores than they do online. However, while these distributions are suggestive about what would be expected when consumers move from offline to online markets, we cannot use these statistics alone to conclude that the rental channel changes a household's selection of DVD rental titles. Specifically, the different rental distributions online and offline in Table 2 could be explained solely by selection effects. Cross-section regressions would suffer from a similar problem, because these regressions obtain empirical identification from comparing DVD rental selections across heterogeneous consumers.

Our empirical approach then is to control for unobserved heterogeneity using panel data, exploiting changes in DVD rental activity across time and across rental channel for each household. For each household i in each week t we define the following variables: $Share\ Popular_{it}$ is the share of the number of superstar DVD title rentals (weekly top 10, top 50, and top 100) divided by the total number of rentals, and $Share\ Offline_{it}$ is the share of the number of rentals made offline, divided by the total number of rentals made both online and offline.

We then use these variables to estimate the following fixed effect model:

$$(1)\quad Share\ Popular_{it}=\alpha+\beta\ Share\ Offline_{it}+\gamma\ Total\ Rentals_{it}+\phi_i+\psi_t+\varphi_z\times t+u_{it}$$

The variable $Total\ Rentals_{it}$ in Model (1) represents the total number of DVD rentals made by household i in week t. The coefficient β in Model (1) measures how weekly changes in the share of DVDs rented from the physical store relate to weekly changes in the share of popular DVD rentals. We control for the weekly total DVD rentals from both online and offline channels because our objective is to examine the effect of channel choice conditional on the total amount of rental consumption (the online appendix presents results not controlling for weekly total DVD rentals). The model includes fixed effects for each household ϕ_i and for each week ψ_t, and includes zip code-specific trends $\varphi_z \times t$.

By using a longitudinal model we can "difference out" the time invariant unobserved characteristics of each household (for example household fixed effects capture income levels or household sizes that are unlikely to change substantially during a seven month period). The week fixed effects capture aggregate changes over time, such as changes in DVD rental consumption that can be caused by school breaks or seasons. To account for pre-existing trends at the level of the zip code, Model (1) also includes zip code-idiosyncratic trends. For example, these idiosyncratic trends may account for market-level changes, such as changes in Internet or cable television usage that might have affected rental consumption patterns during our study period. Identification in Model (1) arises from deviations from zip code-level trends in changes in the DVD rental selection and the rental channel within households from week to week.

While our panel data approach allows us to control for the time invariant tastes of each household, and therefore accounts for the sorting of heterogeneous consumers into channels, Ordinary Least Squares estimates of Model (1) may still provide a misleading measurement of how the rental channel affects the selection of DVD rentals when a household's desire for popular versus non-popular DVDs changes over time. For example, consumers may choose to rent a popular DVD title from the physical store in weeks when they feel impatient about watching a popular newly released title and do not wish to wait for the DVD to arrive in the mail. The rental channel is a choice; and individuals' changes in their desires to watch popular versus non-popular DVDs may influence their channel selection, creating an endogeneity problem. In order to identify

how changes in the rental channel affect the overall selection of DVD rental titles we need to observe changes in individuals' shares of offline rentals that are not caused by weekly changes in the desire to watch popular versus non-popular DVDs.

To break this endogeneity problem, we use the exit of physical stores as an instrumental variable. The rationale for using the exit of physical stores as an instrument is that the exit of a store, by changing the transportation cost of traveling to the store for the individuals that previously rented DVDs from the closing store, increases the relative cost of renting DVDs from the physical channel. In turn, the increase in the relative cost of renting DVDs from the physical channel may induce consumers to shift their rentals from the offline to the online channel. Our instrument is valid as long as it affects channel selection and can be excluded from Model (1). Specifically, the exit of a physical store is a valid instrument even when store closures are not random and are possibly related to a decrease in the local aggregate demand for DVD rentals, as long as store closures are unrelated to relative rental demands for popular versus non-popular DVD titles.

We will use two alternative models to test whether households change the share of transactions made from physical stores when the stores in their geographical market exit. First, following Brynjolfsson, Hu, and Rahman (2009), we assume that the transportation cost of traveling to the physical store increases when the number of physical stores in the zip code decreases. Brynjolfsson, Hu, and Rahman (2009), however, treat zip codes as isolated markets. By computing distances in miles among the zip codes' centroids using data from the United States Census we can extend Brynjolfsson, Hu, and Rahman (2009) to account for changes in the number of physical stores located in adjacent zip codes.

Specifically, we estimate the following first stage model:

$$(2)\quad \textit{Share Offline}_{it}=\delta+\sum_{j=1}^{6}\theta_j\,\textit{Number of Stores in Zip Code}_{ijt}+\rho\,\textit{Total Rentals}_{it}+\vartheta_i+\tau_t+\zeta_z\times t+e_{it}$$

where j equal to 1 represents the zip code where household i resides, j equal to 2 (3, 4, 5, and 6) represents zip codes with centroids located between zero and five (five and ten, ten

and fifteen, fifteen and twenty, and twenty and thirty) miles away from the centroid of the zip code where household *i* resides. We also note that the number of stores in a zip code changes over time through store exit.

We obtained the latitude and longitude for all physical stores and for the fraction of all consumers where we have the consumer's address. For these consumers we can compute the distance to the closest store in each week, and estimate the following model:

$$(3)\ Share\ Offline_{it}=\omega+\epsilon\ Distance\ to\ the\ Closest\ Store_{it}+\mu\ Total\ Rentals_{it}+\pi_i+\sigma_t+\eta_z\times t+v_{it}$$

where $Distance\ to\ the\ Closest\ Store_{it}$ represents the geodesic distance between the location of household *i* and the closest physical store in week *t*. Note that the distances to the closest store change over time for households living near closing stores.

The focal company may naturally close its least successful stores, and the selection of which stores to close may be related to local demographic characteristics or to changes in the local market environment. However, we believe that the exit of stores is unlikely to be affected by individuals' high frequency changes in their relative desires to rent popular versus non-popular DVDs, in which case our instrument is orthogonal to the error. Moreover, the zip code-specific trends in our regression control for pre-existing trends at the level of the zip code that might have induced stores closures. These trends may include trends induced by demographics, Internet or cable connectedness, or the local market environment. In sum, our instrument is valid if the high-frequency timing of store closure is unrelated to the relative desire to rent popular versus non-popular titles. We also show in online appendix EA2 that store closures are unrelated to the number of high-speed Internet providers in each zip code.[8]

It is important to note that our company faces competition from other DVD rental companies, and during our study period other companies rented DVDs exclusively "by

[8] The exit of physical stores may have potentially increased stock-outs of popular titles from surviving physical stores, leading to an increase in niche title consumption from physical stores. However, aggregate rentals of niche titles from physical stores do not show an upward time trend during the period of analysis.

mail," from kiosks, and from physical stores.[9] The entry of DVD rental kiosks (and even the mere existence of rental kiosks prior to our study period) and competition from other physical DVD rental stores might be thought to represent a challenge to our identification strategy because households living near closing stores may rent popular DVD titles from other companies while they continue to rent niche DVD titles from our focal company. If this happens we would observe a change in the relative demand for popular versus niche titles associated with the exit of stores, when the unobserved consumption bundle of niche and superstar titles from both the focal company and the competitors might remain unchanged. This would invalidate the use of our instrument.

However, as noted above, our data only include information from consumers with a "Rent by Mail" subscription, and these subscribers do not pay a separate price each time they rent a DVD from either the online or offline channel. Although in theory the consumers in our data may simultaneously have subscriptions with our focal company and may be willing to pay extra to have a separate subscription with other competing companies, or to rent DVDs from competitive outlets, we are doubtful that this is a common practice. In this regard, we note that the market leader in rentals via kiosks, Redbox, considers that, "people who use the kiosks tend to be casual viewers who don't want to be tied down to subscriptions or membership fees" (Green 2009). More importantly, in the Appendix, we use historical data on Redbox locations to show that our conclusions in the main text are robust to restricting the analysis to locations without a local Redbox kiosk.

Of course, it is also true that households living near closing stores may decide to cancel their subscriptions from the focal company, and begin renting from other companies (e.g., from Netflix) in which case their rentals will not be recorded in our data. For this reason, in the results below, we analyze the sensitivity of our results to attrition using a balanced panel of consumers.

[9] The market could also be defined more broadly as consumers may consider DVD rentals as a substitute for cable television, DVD purchases, movie theaters, or other entertainment options such as playing games or using the Internet.

5. Summary Statistics and Results

5.1 Summary Statistics

Table 3 presents summary statistics for our data, computed using household-week observations. Consumers in our data rent on average 2.25 DVDs per week and they rent 28.4% of these DVDs from physical stores. On average, top 10 (top 50, top 100) DVD titles represent 20.6% (39.6%, 47.9%) of all transactions. The average number of stores per zip code is 0.52. We will use these statistics below in interpreting our regression results.

Table 3: Additional Summary Statistics

	Mean	Standard Deviation	5%	95%
Total DVD Rentals per Week	2.255	1.391	1	5
Share of Offline Rentals	0.284	0.373	0	1
Share of Rentals Taken by Top 10 Titles	0.206	0.324	0	1
Share of Rentals Taken by Top 50 Titles	0.396	0.406	0	1
Share of Rentals Taken by Top 100 Titles	0.479	0.419	0	1
Stores per Zip Code†	0.525	0.612	0	2

These statitics are computed using 20,249,021 observations.

We explain below that some regressions do not use all the observations.

†In the data, zip codes have between zero and three stores.

5.2 Ordinary Least Square Results

Table 4 presents the OLS estimation results for Model (1). The standard errors are clustered at the household level to allow for the possibility of serial correlation over time. The regressions include fixed effects for each week, over a million and a half fixed effects at the household level, and approximately twenty five thousand zip code-specific trends. The results show that households increase the fraction of rentals of popular DVDs when they rent more DVDs from the physical store. The coefficient estimates on the fraction of offline rentals are both statistically and economically significant.

Table 4: Share of Popular Rentals - OLS Estimates			
	I	**II**	**III**
	Top 10	Top 50	Top 100
Share of Offline Rentals	0.3650***	0.4967***	0.4682***
(mean 0.28)	(0.0003)	(0.0004)	(0.0004)
Total DVD Rentals	-0.0120***	-0.0090***	-0.0066***
(mean 2.25)	(0.0001)	(0.0001)	(0.0001)
Constant	-0.0142***	-0.0202***	-0.0118***
	(0.0004)	(0.0004)	(0.0005)
Observations	20,249,021	20,249,021	20,249,021
R-squared	0.3251	0.4248	0.4268
Includes fixed effects for both weeks (30) and individuals (1,529,028), and ZIP code-specific trends (24,648).			
Standard errors in parentheses are clustered by household.			
The mean of the dependent variable is 0.20 in Column I, 0.39 in Column II, and 0.47 in Column III.			
* significant at 10%; ** significant at 5%; *** significant at 1%			

The sizes of the coefficient estimates on the fraction of offline rentals indicate that a household that decreases the fraction of DVDs rented from the physical store from twenty eight percent to zero, as might be the case when physical stores are eliminated as a choice for consumers (note that the mean of the share of offline rentals in Table 4 is 0.28), would decrease the fraction of top 10 (top 50, top 100) DVD rentals by 10.2 (13.9, 13.1) percentage points. These effects are substantial. For example, Table 3 above shows that the top 10 DVD titles represent twenty percent of all rental transactions. Given this, the coefficient estimate on the share of offline rentals in Column I of Table 4 indicates that top 10 DVD titles would take approximately ten percent of all transactions if all physical stores go out of business, which represents a fifty percent decrease in the share of transactions taken by the top 10 DVD titles. Columns II and III of Table 4 indicate that when consumers move all their transactions to the online channel the share of transactions taken by the top 50 and top 100 titles would decrease by thirty five and twenty seven percent respectively.[10]

[10] Although the coefficient is greater for the top 50 and top 100 titles than for the top 10 titles, the impact is smaller as a percentage change because the size of the dependent variable is also greater for the top 50 and top 100 titles compared to the top 10 titles.

The sign of the coefficient estimates on total DVD rentals per week is negative and economically small. The negative sign may be unsurprising because individuals may tend to rent top DVDs first, and rent DVDs that are further down the popularity distribution during weeks when they increase the number of DVD rentals. Renting an additional DVD in a week reduces the fraction of top 10, top 50, or top 100 DVD rentals by between 0.6 and 1.2 percentage points.

5.3 Instrumental Variable Results

The regressions in Table 5 present our first stage results examining how channel choice is affected by the exit of stores. The results in Column I of Table 5 show that when one store exits from a zip code, consumers living in that zip code decrease their share of offline rentals by an average of 2.2 percentage points (or approximately 7.8% of the transactions made at physical stores). This result is expected because, by increasing the transportation cost, the exit of a store from a zip code increases the relative cost of renting from the physical store for households that reside in that zip code.

Our first stage results are consistent with the prior literature showing that the likelihood of purchasing products online decreases as the number of stores in the zip code increases (Brynjolfsson, Hu, and Rahman 2009). However, Brynjolfsson, Hu, and Rahman (2009) treat zip codes as isolated geographic markets, while we can also provide information regarding the size of the geographic market.

Column I of Table 5 shows how the impact of the exit of a store on channel choice dissipates for households living further away from the closing store. The results in Column I of Table 5 indicate that the closure of a store within the zip code where a household resides has an impact on the household's channel choice that is seven times larger than the impact of the closure of a store in other zip codes with centroids located less than five miles away from the centroid of the zip code where the household resides. The results in Column I of Table 5 also show how the exit of stores in zip codes located further away have no impact on households' channel choices.

Table 5: Share of Offline Rentals - First Stage		
	I	**II**
Number of Stores in the Zip Code Where the	0.02239***	na
Household Resides (mean 0.52)	(0.00078)	na
Number of Stores in Zip Codes between Zero	0.00304***	na
and Five Miles Away (mean 2.17)	(0.00031)	na
Number of Stores in Zip Codes between Five	-0.00004	na
and Ten Miles Away (mean 5.23)	(0.00019)	na
Number of Stores in Zip Codes between Ten	0.00004	na
and Fifteen Miles Away (mean 6.56)	(0.00017)	na
Number of Stores in Zip Codes between Fifteen	-0.00004	na
and Twenty Miles Away (mean 6.89)	(0.00017)	na
Number of Stores in Zip Codes between Twenty	-0.0001	na
and Thirty Miles Away (mean 12.80)	(0.00011)	na
Distance to the Closest Store in Miles	na	-0.00898***
(mean 3.6)	na	(0.00035)
Squared Distance to the Closest Store in Miles	na	0.00012***
	na	(0.00001)
Total DVD Rentals	0.02605***	0.02518***
(mean 2.25 in Columns I and II)	(0.00008)	(0.00011)
Constant	-0.02364***	-0.02841***
	(0.00042)	(0.00062)
Observations	20,249,021	9,141,694
F-test (global test for the excluded instruments in the second stage)	155.5	399.7
R-squared	0.2935	0.3199
Includes fixed effects for both weeks (30) and individuals (Column I 1,529,028; Column II 855,996), and		
ZIP code-specific trends (Column I 24,648; Column II 19,233). Standard errors in parentheses are clustered by household.		
The mean of the dependent variable is 0.28 in Column I and 0.25 in Column II.		
* significant at 10%; ** significant at 5%; *** significant at 1%		

We also use the geodesic distance from consumers' locations to the closest physical stores as an alternative instrument. We acknowledge that some consumers may use stores that are not the closest to their home address (e.g., stores nearby their working location or in the way when running errands), but we still believe that using the closest store to the home address is useful as an approximation for the transportation costs of using the offline channel. Comparing unconditional means, households living less than one mile away from a physical store make 29.4% of their rentals offline and households living more than 20 miles away from physical stores make 10.3% of their rentals offline. Column II of Table 5 shows the estimates for Model (3) using the square of the distance to the closest store as an additional covariate to account for non-linear transportation

costs.[11] For example, the results in Column II of Table 5 indicate that households residing near a closing physical store will decrease average rentals from physical stores by 8.8 percentage points (or approximately 35.2% of the transactions made from physical stores) when the new closest store for these households is ten miles away. Moreover, the sizes of the coefficients indicate that households that reside near a closing physical store will decrease the transactions made from physical stores to approximately zero when the new closest store for these households is thirty miles away.

The results for the second stage of Model (1) in Table 6 still show that individuals increase the fraction of popular DVD rentals when they rent more DVDs from the physical store. The first three columns use Column I in Table 5 for the first stage regression and the last three columns use Column II in Table 5 for the first stage regression. In the first three regressions in Table 6 the sizes of the coefficient estimates on the fraction of offline rentals indicate that when a household decreases the fraction of DVDs rented from physical stores from twenty eight percent to zero, the fraction of top 10 (top 50, top 100) DVD rentals decreases by 10.7 (10.0, 7.3) percentage points. In the last three regressions in Table 6 the sizes of the coefficient estimates on the fraction of offline rentals indicate that when a household decreases the fraction of DVDs rented from the store from twenty five percent to zero (note that the mean of the share of offline rentals for the last three columns of Table 6 is 0.25), the fraction of top 10 (top 50, top 100) DVD rentals decreases by 10.4 (9.9, 8.9) percentage points.

Table 6: Share of Popular Rentals - IV Estimates - Second Stage

	I	II	III	IV	V	VI
	Top 10	Top 50	Top 100	Top 10	Top 50	Top 100
Share of Offline Rentals	0.3856***	0.3597***	0.2638***	0.4184***	0.3969***	0.3594***
(mean 0.28 in Columns I through III	(0.0372)	(0.0457)	(0.0473)	(0.0391)	(0.0511)	(0.0524)
and 0.25 in Columns IV through VI)						
Total DVD Rentals	-0.0125***	-0.0054***	-0.0013	-0.0129***	-0.0068***	-0.0044***
(mean 2.25 in Columns I through VI)	(0.0010)	(0.0012)	(0.0012)	(0.0010)	(0.0013)	(0.0013)
Constant	-0.0137***	-0.0234***	-0.0166***	-0.0092***	-0.0177***	-0.0086***
	(0.0010)	(0.0012)	(0.0012)	(0.0013)	(0.0016)	(0.0017)
Observations	20,249,021	20,249,021	20,249,021	9,141,694	9,141,694	9,141,694

Includes fixed effects for both weeks (30) and individuals (Columns I through III 1,529,028; Columns IV through VI 855,996), and ZIP code-specific trends (Columns I through III 24,648; Columns IV through VI 19,233). Standard errors in parentheses are clustered by household.

The mean of the dependent variable is 0.20 in Column I, 0.39 in Column II, 0.47 in Column III, 0.19 in Column IV, 0.38 in Column V, and 0.46 in Column VI.

* significant at 10%; ** significant at 5%; *** significant at 1%

[11] We note that excluding the square of the distance from the first stage (or alternatively including higher order polynomials) causes no substantial change in the second stage results.

Comparing the Instrumental Variables results in Table 6 with the OLS results in Table 4, we observe that the size of the coefficient estimates on the fraction of offline rentals are similar for top 10 titles and smaller, but still significant both economically and statistically, for top 50 and top 100 titles.

As in Table 4, the coefficient estimates on total DVD rentals per week in Table 6 are negative and economically small.

5.4 Sensitivity of Results to Attrition

Table 1 suggests that our focal company lost subscribers during our study period. Additionally, the customer base exhibits a high churn rate likely fueled by free of charge trial period offers. Our panel of data is therefore unbalanced. An unbalanced panel does not generate biased estimates when the reason for having missing observations is not correlated with the regression error term. For example, subscribers that rent DVDs in only one week during our study period do not bias the estimates since time demeaning for such observations yields all zeros. But in order to examine whether or not our previous empirical results are driven by changes in the customer base, in this section we rerun our regressions using a balanced panel of subscribers.

Table 7 presents OLS results analogous to those in Table 4, but only including the information from subscribers who rented at least one DVD in both the initial and final four weeks of our study period. There are 764,875 such subscribers, or approximately half of the total number of subscribers compared with Table 4. In Table 7 the coefficient estimates on both the fraction of offline rentals and the total DVD rentals per week variables are very similar to those in Table 4.

The similarity of the results in Tables 4 and 7 indicates that the effects of the online versus offline channel choice on superstar DVD rentals are not significantly different for subscribers who maintained subscriptions during the entire length of our study period versus subscribers who did not.

Table 7: Share of Popular Rentals - OLS Estimates - Balanced Panel			
	I	**II**	**III**
	Top 10	Top 50	Top 100
Share of Offline Rentals	0.3713***	0.5003***	0.4701***
(mean 0.30)	(0.0003)	(0.0004)	(0.0004)
Total DVD Rentals	-0.0123***	-0.0093***	-0.0070***
(mean 2.33)	(0.0001)	(0.0001)	(0.0001)
Constant	-0.0079***	-0.0161***	-0.0089***
	(0.0004)	(0.0005)	(0.0005)
Observations	15,239,005	15,239,005	15,239,005
R-squared	0.3105	0.4094	0.4075
Includes fixed effects for both weeks (30) and individuals (764,875), and ZIP code-specific trends (21,396).			
Standard errors in parentheses are clustered by household.			
The mean of the dependent variable is 0.22 in Column I, 0.41 in Column II, and 0.49 in Column III.			
* significant at 10%; ** significant at 5%; *** significant at 1%			

However, the value of a subscription may be greater as the distance to an offline store decreases, since having a physical store nearby provides the additional value of exchanging DVDs at the store. Since attrition in our data is likely correlated with the exit of physical stores, attrition might bias our instrumental variable results. Tables 8 and 9 present first and second stage regressions analogous to those in Tables 5 and 6, but using the balanced sub-sample of our data.

The results for both the first stage and second stage regressions in Tables 8 and 9 using the balanced sub-sample are similar than those in Tables 5 and 6 using the entire sample. This similarity suggests that our previous results using the entire sample are not driven by changes in the profile of customers over time.

Table 8: Share of Offline Rentals - First Stage - Balanced Panel		
	I	II
Number of Stores in the Zip Code Where the	0.02114***	na
Household Resides (mean 0.53)	(0.00085)	na
Number of Stores in Zip Codes between Zero	0.00277***	na
and Five Miles Away (mean 2.17)	(0.00034)	na
Number of Stores in Zip Codes between Five	-0.00009	na
and Ten Miles Away (mean 5.23)	(0.00020)	na
Number of Stores in Zip Codes between Ten	0.00013	na
and Fifteen Miles Away (mean 6.58)	(0.00018)	na
Number of Stores in Zip Codes between Fifteen	0.00002	na
and Twenty Miles Away (mean 6.93)	(0.00018)	na
Number of Stores in Zip Codes between Twenty	-0.00011	na
and Thirty Miles Away (mean 12.88)	(0.00012)	na
Distance to the Closest Store in Miles	na	-0.00863***
(mean 3.49)	na	(0.00043)
Squared Distance to the Closest Store in Miles	na	0.00011***
	na	(0.00001)
Total DVD Rentals	0.02518***	0.02406***
(mean 2.33 in Column I and 2.34 in Column II)	(0.00009)	(0.00014)
Contant	-0.00741***	-0.00026
	(0.00046)	(0.00073)
Observations	15,239,005	5,667,199
F-test (global test for the excluded instruments in the second stage)	114.5	256.7
R-squared	0.2737	0.2888
Includes fixed effects for both weeks (30) and individuals (Column I 764,875; Column II 288,845), and		
ZIP code-specific trends (Column I 21,396; Column II 15,770). Standard errors in parentheses are clustered by household.		
The mean of the dependent variable is 0.30 in Column I and 0.28 in Column II.		
* significant at 10%; ** significant at 5%; *** significant at 1%		

Table 9: Share of Popular Rentals - IV Estimates - Second Stage - Balanced Panel						
	I	II	III	IV	V	VI
	Top 10	Top 50	Top 100	Top 10	Top 50	Top 100
Share of Offline Rentals	0.4129***	0.3147***	0.2042***	0.4125***	0.3800***	0.3546***
(mean 0.30 in Columns I through III	(0.0443)	(0.0539)	(0.0556)	(0.0502)	(0.0650)	(0.0658)
and 0.28 in Columns IV through VI)						
Total DVD Rentals	-0.0133***	-0.0047***	-0.0003	-0.0129***	-0.0068***	-0.0047***
(mean 2.33 in Columns I through III	(0.0011)	(0.0014)	(0.0014)	(0.0012)	(0.0016)	(0.0016)
and 2.34 in Columns IV through VI)						
Constant	-0.0076***	-0.0174***	-0.0108***	-0.0044***	-0.0128***	-0.0051***
	(0.0005)	(0.0007)	(0.0007)	(0.0007)	(0.0008)	(0.0009)
Observations	15,239,005	15,239,005	15,239,005	5,667,199	5,667,199	5,667,199
Includes fixed effects for both weeks (30) and individuals (Columns I through III 764,875; Columns IV through VI 288,845), and						
ZIP code-specific trends (Columns I through III 21,396; Columns IV through VI 15,770). Standard errors in parentheses are clustered by household.						
The mean of the dependent variable is 0.22 in Column I, 0.41 in Column II, 0.49 in Column III, 0.21 in Column IV, 0.40 in Column V, and 0.48 in Column VI.						
* significant at 10%; ** significant at 5%; *** significant at 1%						

Regressions in the online appendix show various other sensitivity tests for our results including results that exclude total DVD rentals as a covariate in Tables 4, 5, and 6 (Appendix EA3, Tables EA3, EA4, and EA5), results that only use subscribers where we have the consumer's physical address in Table 4, Column I of Table 5 and Columns I through III of Table 6 (Appendix EA3, Tables EA6, EA7, and EA8), and results examining rentals of "top 2,000" titles (those that are likely to be stocked in both physical and online channels) (Appendix EA4). Our main results are robust to each of these considerations.

6. Discussion

As the proportion of commerce conducted online increases, will producers and retailers need to re-evaluate their investment and inventory choices? Answering this question is complicated by selection effects surrounding the types of consumers who purchase online and the types of products that consumers choose to purchase online. While early research has observed a large proportion of sales online in niche products, it is unclear whether this observation is merely a reflection of the characteristics of the consumers who select the channel, of the types of products that consumers select to purchase online versus offline, or whether it might reflect a change in consumption patterns caused by the characteristics of the Internet channel.

Breaking this endogeneity requires an exogenous shift in the cost of purchasing online, and the ability to observe customer-level purchase decisions by channel before and after the shift. Our data provide us with just such an opportunity. Our data document customer-level rental decisions before and after a customer's local video rental store closes, and our empirical analysis suggests that when consumers move online they are much less likely to rent blockbuster titles than they were previously.

While our objective in this paper has been to examine how channel selection affects consumption patterns, our results showing how the impact of store exit on channel choice varies depending on where consumers live relative to the closing store, also complement and extend the prior literature on transportation costs and channel selection (e.g.,

Brynjolfsson, Hu, and Rahman 2009; Forman, Ghose, and Goldfarb 2009). Specifically, our data are substantially more granular than those used in previous work studying this question. Although our results extend the prior literature on channel selection and transportation costs, our examination of how changes in home-store distances affect channel selection in this paper is presented in the context of a first stage regression and is not the main research focus. Conducting a more detailed investigation of how transportation costs influence the online versus offline channel selection using our data is a potential avenue for further research.

Our main result, indicating that when consumers move to online channels they decrease their likelihood of renting popular titles, is of course not without limitations. Importantly, while we examine how consumption changes when consumers move online we do not examine why this change in consumption patterns occurs. Moreover, our results only provide evidence concerning a specific market, and are not necessarily generalizable to other environments. Online commerce could have heterogeneous impacts across industries, and transform different markets into either a "long tail" or "superstar" market based on the specific nature of each industry. For example, a mechanism that may partly underlie our results is that our focal company does not display or promote popular products as heavily in the online channel as it does in the brick-and-mortar channel. While this characteristic is typical across various online versus brick-and-mortar channels, it might be more pronounced in the specific market we study than in other markets.

The queue system for online consumption is also specific to our setting. If consumers were more likely to experience stock-outs in the online channel relative to the physical channel, or if the relative prevalence of stock-outs increased over time as stores closed, our results could be explained by supply-side effects related to stock availability versus demand-side effects related to customer preferences. However, our limited checks of stock levels in both channels described above suggest that, if anything, popular titles have a higher availability online than in physical channels. Similarly, as we noted above, our focal company's practice of moving inventory for popular titles from physical stores to the online channel as stored closed, combined with the increased efficiency of a single

online warehouse queue relative to multiple queues in physical stores and with the observed reduction in total subscribers as stores closed should, if anything, increase the relative availability of popular titles online as stores close.[12] Nonetheless, our inability to observe channel-specific stock levels in our data represents a limitation of our analysis. Identifying how our observed effects are determined by supply versus demand side factors is an important avenue for future research.

Similarly, even focusing on a single market, our results could vary over time. For example, early adopters of "Rent by Mail" subscriptions might be more interested in niche DVD titles than late adopters are. In this regard, we believe that our analysis of a mature market provides a more useful examination of the market-level impact of online commerce on product concentration than an analysis of a nascent market would. Although our period of analysis is too short to examine how the results change over time, this examination is also a potential avenue for future research.

To summarize, our results show that there is a change in consumption patterns caused by the characteristics of the Internet channel. As a consequence, our finding that online channels may shift DVD consumption away from blockbuster titles and toward more niche titles may have implications for movie studios and movie producers. Specifically, movie studios have typically faced a market where a small number of hits made up the vast majority of industry profits. Our results suggest that this historical pattern of highly concentrated transactions in a handful of titles might have been driven by the characteristics of the offline channel, and that studios may wish to shift their resources relatively toward more "long tail" titles as consumers move online. There is obviously a need for more research to be in a position to predict the degree to which Internet markets change the incentives of movie producers and allow for the production of more niche titles; and we believe this paper is a first step in that direction.

[12] We also note that our results in the extended appendix show that our long tail effects are robust to excluding data for the top 5 or top 10 most popular titles (Appendix AE5).

References

Anderson, Chris. 2006. The Long Tail: Why The Future of Business Is Selling Less of More. New York, NY: Hyperion.

Bar-Isaac, Heski and Caruana, Guillermo and Cuñat, Vicente. Forthcoming. Search, design, and market structure. *American Economic Review*. Forthcoming.

Brynjolfsson, Erik, Yu Hu, Michael Smith. 2003. Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety. *Management Science* **49**(11) 1580-1596.

Brynjolfsson, Erik, Yu Hu, Michael D. Smith. 2006. From Niches to Riches: Anatomy of the Long Tail. *Sloan Management Review,* **47 4**(Summer) 67-71.

Brynjolfsson, Erik, Yu Hu, Mohammad 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 Hu, Michael D. Smith. 2010. Long Tails vs. Superstars: The Effect of Information Technology on Product Variety and Sales Concentration Patterns. *Information Systems Research* **4**(21) 736-747.

Brynjolfsson, Erik, Yu Hu, Duncan Simester. 2011. Goodbye Pareto Principle, Hello Long Tail: the Effect of Search Costs on the Concentration of Product Sales. *Management Science* **57**(8) 1373-1386.

BusinessWeek, 2005. http://images.businessweek.com/ss/05/12/bestideas/index_01.htm

Chu, Junhong, Pradeep Chintagunta, and Javier Cebollada. 2008. A Comparison of Within-Household Price Sensitivity across Online and Offline Channels. *Marketing Science* 27(2), March–April 2008, pp. 283–299.

Elberse, Anita, Felix Oberholzer-Gee. 2007. Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales. *Marketing Science Institute* 4 (2007): 49-72.

Elberse, Anita. 2008. Should You Invest in the Long Tail? *Harvard Business Review* **86**(7/8) 88-96.

Fleder, Daniel, Kartik Hosanagar. 2009. Blockbuster culture's next rise and fall: The impact of recommender systems on sales diversity. *Management Science* **55**(5) 697–712.

Forman, Chris, Anindya Ghose, Avi Goldfarb. 2009. Competition between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where you Live. *Management Science* **55**(1) 47-57.

Green, Chris. 2009. http://www.independent.co.uk/arts-entertainment/films/news/dvd-rental-machines-to-be-launched-in-uk-supermarkets-1782926.html

Oestreicher-Singer, Gal, Arun Sundararajan. 2011. Recommendation Networks and the Long Tail of Electronic Commerce. *MIS Quarterly*. Forthcoming.

Pozzi, Andreas. 2012. Shopping Cost and Brand Exploration in Online Grocery. *American Economic Journal: Microeconomics*. Forthcoming.

Rosen, Sherwin. 1981. The Economics of Superstars. *The American Economic Review*, **71**(5) 845-858.

Tucker, Catherine, Juanjuan Zhang. 2011. How does popularity information affect choices? A field experiment. *Management Science* **57**(5) 828-842.

Waldfogel, Joel. 2012. And the Bands Played On: Digital Disintermediation and the Quality of New Recorded Music. Working Paper, University of Minnesota. Minneapolis, Minnesota.

Appendix: Sensitivity of Main Results to Redbox Kiosk Locations

In this appendix we use historical data on the location of Redbox kiosks as of April 28, 2010 (the end of our study period is April 29 2010). We obtained these data from AggData LLC, a company that provides data to businesses and organizations.[13] Tables A1 and A2 show that the conclusions in the main text are not affected when using only the information from individuals living in zip codes where Redbox was not present during our study period. During our study period, approximately one fourth of the focal company's consumers lived in zip codes without a Redbox's kiosk; and these consumers made on average 26% of their transactions from physical stores (compared to 28% for all consumers) and rented a similar fraction of top 10, top 50, and top 100 DVD titles as other consumers do (comparing the mean of the dependent variables presented at the bottom of Tables 6 in the main text and A2). Table A1 shows that Redbox had substantially more presence in zip codes where our focal company also had a physical presence (the variable "Mean Number of Stores in the Zip Code Where the Household Resides" is 0.52 in Table 5 in the main text compared to 0.20 in Table A1).

[13] http://www.aggdata.com/

Table A1: Share of Offline Rentals - First Stage	
	I
Number of Stores in the Zip Code Where the	0.02727***
Household Resides (mean 0.20)	(0.00244)
Number of Stores in Zip Codes between Zero	0.00494***
and Five Miles Away (mean 2.15)	(0.00064)
Number of Stores in Zip Codes between Five	-0.00103***
and Ten Miles Away (mean 5.01)	(0.00036)
Number of Stores in Zip Codes between Ten	-0.00027
and Fifteen Miles Away (mean 6.33)	(0.00034)
Number of Stores in Zip Codes between Fifteen	-0.00064*
and Twenty Miles Away (mean 6.57)	(0.00034)
Number of Stores in Zip Codes between Twenty	-0.00030
and Thirty Miles Away (mean 12.56)	(0.00022)
Total DVD Rentals	0.02604***
(mean 2.23)	(0.00016)
Constant	-0.02346***
	(0.00082)
Observations	4,897,442
F-test (global test for the excluded instruments in the second stage)	30.6
R-squared	0.3026
Includes fixed effects for both weeks (30) and individuals (378,160), and ZIP code-specific trends (16,861).	
Standard errors in parentheses are clustered by household.	
The mean of the dependent variable is 0.26.	
* significant at 10%; ** significant at 5%; *** significant at 1%	

Table A2: Share of Popular Rentals - IV Estimates - Second Stage			
	I	**II**	**III**
	Top 10	Top 50	Top 100
Share of Offline Rentals	0.4471***	0.3042***	0.3202***
(mean 0.26)	(0.0819)	(0.1021)	(0.1061)
Total DVD Rentals	-0.0139***	-0.0039	-0.0026
(mean 2.23)	(0.0021)	(0.0027)	(0.0028)
Constant	-0.0135***	-0.0239***	-0.0140***
	(0.0021)	(0.0026)	(0.0027)
Observations	4,897,442	4,897,442	4,897,442
Includes fixed effects for both weeks (30) and individuals (378,160), and ZIP code-specific trends (16,861).			
Standard errors in parentheses are clustered by household.			
The mean of the dependent variable is 0.20 in Column I, 0.39 in Column II, and 0.47 in Column III.			
* significant at 10%; ** significant at 5%; *** significant at 1%			

Extended Appendix

Appendix EA1: Popularity Rankings Based on Online and Offline Information

In Section 3.1 we explained the convenience of computing rankings using information from both the online and offline channels in examining consumption patterns when consumers move from offline to online. Using our data on DVD rentals, we noted that the specific titles that are popular in the online channel are somewhat different than the titles that are popular in the offline channel. We also explained that the differences in the titles making up the rankings online versus offline may occur due to selection effects and other cross-channel differences on the demand-side or on the supply-side. However, partly because we do not observe inventory or title assortments online and offline, we have limited ability to measure the degree to which each factor contributes to these differences. Although the focus of this paper is not on identifying the factors explaining the differences in online versus offline title popularity (nor is our focus on examining the specific demand or supply factors underlying the changes in the concentration of rental transactions when consumers move online), Table EA1 may provide some useful information regarding the selection of titles available from each channel.

Table EA1: Popularity Online and Offline

	I	II	III		IV	V	VI
	Top 10 Online	Top 50 Online	Top 100 Online		Top 10 Offline	Top 50 Offline	Top 100 Offline
Top 10 Offline	0.73	n.a	n.a	Top 10 Online	0.73	n.a	n.a
Top 20 Offline	0.92	n.a	n.a	Top 20 Online	0.82	n.a	n.a
Top 50 Offline	0.95	0.70	n.a	Top 50 Online	0.90	0.70	n.a
Top 100 Offline	0.95	0.89	0.70	Top 100 Online	0.97	0.86	0.70
Top 200 Offline	0.95	0.90	0.84	Top 200 Online	0.98	0.94	0.84
Top 1,000 Offline	0.95	0.90	0.85	Top 1,000 Online	0.99	0.99	0.99
Top 5,000 Offline	0.95	0.90	0.85	Top 5,000 Online	1.00	1.00	1.00

Column I (II, and III) in Table EA1 shows the offline rankings of titles that are ranked in the top 10 (50, 100) in the online channel. Consistent with Figure 1 in the main text, Column I indicates that a weekly average of 7.3 DVD titles that are top 10 in the online channel are also top 10 in the offline channel. Column I of Table EA1 also shows that in any given week most but not all titles that are top 10 online are ranked among the top 5,000 in the offline channel. This might happen, for example, when a title becomes

available online before it is available offline. On the other hand, Column IV in Table EA1 shows that all titles that are top 10 offline are ranked among the top 5,000 online.

Column I (Column IV) of Table EA1 also shows that approximately two thirds (one third) of the 2.7 titles that are top 10 online (offline) but not offline (online) are ranked between the top 10 and top 20 offline (online).[14] While the different rankings online and offline may arise from selection effects or other demand differences online versus offline, they may also be explained by stock-outs for specific titles in the online or offline channels. For example, a top 10 title in the online channel may be ranked between top 10 and top 20 in the offline channel when there are fewer copies of this title in the offline channel leading to stock-outs at physical stores. In addition, a title may be available from some but not all brick-and-mortar stores; for example, a title may become available from various brick-and-mortar stores on different dates. Similarly, a top 10 title in the offline channel may be ranked between top 10 and top 20 online when there are stock-outs online.

Columns II and III of Table EA1 show that in any given week there are top 50 or top 100 titles online are not ranked top 5,000 in the offline channel. Conversely, Columns V and VI of Table EA1 show that all top 50 and top 100 titles online are ranked among the top 5,000 in the offline channel.

Appendix EA2: Online Connectedness and Store Closures

We collected data from various sources in order to investigate whether store closures are more likely to occur in places with higher online connectedness. We obtained data on the number of high-speed Internet service providers at the zip code level from the Federal Communications Commission (FCC).[15] We constructed a variable measuring average

[14] The result of 0.92 minus 0.73 (0.19) is approximately two thirds of 0.27 (1-0.73); 0.82 minus 0.73 (0.09) is approximately one thirds of 0.27.

[15] We use data from June 2008 because the FCC discontinued the reporting of these data at the zip code level after this date. http://transition.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-State_Link/IAD/hzip0608.pdf

income by zip code using data from tax returns from the Internal Revenue Service.[16] We also collected data at the zip code level on average family size, median age, race, population, and area size from the 2010 U.S. Census.[17]

Using our data on physical stores' locations and closing dates we constructed a dummy variable equal to one for stores that closed during our study period and zero for stores that remained open at the end of our study period. In Table EA2 we use a logit regression model to investigate the factors affecting the probability of store closures. These results show that the effect of the number of high-speed Internet providers on the likelihood of a store closure is statistically insignificant. This result may give more confidence regarding the exclusion restriction in our instrumental variable regressions.[18]

[16] http://www.irs.gov/uac/SOI-Tax-Stats---Individual-Income-Tax-Statistics---Free-ZIP-Code-data-(SOI)
[17] http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml
[18] Even if increases in Internet penetration had caused the exit of physical stores, omitting Internet penetration from our regressions in the main text might be a concern only if changes in Internet penetration change the relative desires to rent popular versus non-popular DVDs. Our regressions in the main text include zip code specific time trends that may account for zip code specific trends in Internet penetration. Our finding in this appendix indicating that the number of high-speed Internet providers does not affect the likelihood of store closures further strengthens the use of our instrumental variable.

Table EA2: Probability of Store Closure	
Cross Section - Logistic Estimates - Marginal Effects	
Number of ISP Providers in Zip Code	0.0026
	(0.0020)
Average Income	0.0064
	(0.0048)
Average Family Size	0.0165
	(0.0363)
Median Age	0.0027**
	(0.0015)
Percentage Hispanic	0.0665
	(0.0526)
Percentage African American	0.1454***
	(0.0486)
Percentage Asian	0.0906
	(0.0850)
Population	-0.0066
	(0.0052)
Area Size	-0.0093
	(0.0082)
Observations	3,356
Pseudo R-squared	0.0101
Robust standard errors in parentheses.	
The mean of the dependent variable is 0.15.	
* significant at 10%; ** significant at 5%; *** significant at 1%	

Appendix EA3: Robustness

Tables EA3, EA4, and EA5 present regression results analogous to those in Tables 4, 5, and 6 in the main text, but without including Total DVD Rentals as a covariate in the regressions. The results from Tables EA3, EA4, and EA5 are similar to the results in Tables 4, 5, and 6 in the main text.

Table EA3: Share of Popular Rentals - OLS Estimates			
	I	II	III
	Top 10	Top 50	Top 100
Share of Offline Rentals	0.3617***	0.4942***	0.4664***
(mean 0.28)	(0.0003)	(0.0004)	(0.0004)
Constant	-0.0126***	-0.0190***	-0.0109***
	(0.0004)	(0.0004)	(0.0005)
Observations	20,249,021	20,249,021	20,249,021
R-squared	0.3237	0.4243	0.4265
Includes fixed effects for both weeks (30) and individuals (1,529,028), and ZIP code-specific trends (24,648).			
Standard errors in parentheses are clustered by household.			
The mean of the dependent variable is 0.20 in Column I, 0.39 in Column II, and 0.47 in Column III.			
* significant at 10%; ** significant at 5%; *** significant at 1%			

Table EA4: Share of Offline Rentals - First Stage		
	I	II
Number of Stores in the Zip Code Where the	0.02312***	na
Household Resides (mean 0.52)	(0.00078)	na
Number of Stores in Zip Codes between Zero	0.00320***	na
and Five Miles Away (mean 2.17)	(0.00031)	na
Number of Stores in Zip Codes between Five	(0.00009)	na
and Ten Miles Away (mean 5.23)	-0.00019	na
Number of Stores in Zip Codes between Ten	0.00000	na
and Fifteen Miles Away (mean 6.56)	(0.00017)	na
Number of Stores in Zip Codes between Fifteen	(0.00008)	na
and Twenty Miles Away (mean 6.89)	-0.00017	na
Number of Stores in Zip Codes between Twenty	-0.00015	na
and Thirty Miles Away (mean 12.80)	(0.00011)	na
Distance to the Closest Store in Miles	na	-0.00919***
(mean 3.6)	na	(0.00036)
Squared Distance to the Closest Store in Miles	na	0.00012***
	na	-0.00001
Constant	-0.02737***	-0.03118***
	(0.00042)	(0.00062)
Observations	20,249,021	9,141,694
F-test (global test for the excluded instruments in the second stage)	162.7	403.9
R-squared	0.28842	0.31497
Includes fixed effects for both weeks (30) and individuals (Column I 1,529,028; Column II 855,996), and		
ZIP code-specific trends (Column I 24,648; Column II 19,233). Standard errors in parentheses are clustered by household.		
The mean of the dependent variable is 0.28 in Column I and 0.25 in Column II.		
* significant at 10%; ** significant at 5%; *** significant at 1%		

Table EA5: Share of Popular Rentals - IV Estimates - Second Stage						
	I	II	III	IV	V	VI
	Top 10	Top 50	Top 100	Top 10	Top 50	Top 100
Share of Offline Rentals	0.3691***	0.3500***	0.2611***	0.4065***	0.3906***	0.3554***
(mean 0.28 in Columns I through III	(0.0361)	(0.0443)	(0.0458)	(0.0382)	(0.0499)	(0.0512)
and 0.25 in Columns IV through VI)						
Constant	-0.0124***	-0.0229***	-0.0165***	-0.0082***	-0.0171***	-0.0082***
	(0.0011)	(0.0013)	(0.0013)	(0.0013)	(0.0017)	(0.0018)
Observations	20,249,021	20,249,021	20,249,021	9,141,694	9,141,694	9,141,694
Includes fixed effects for both weeks (30) and individuals (Columns I through III 1,529,028; Columns IV through VI 855,996), and						
ZIP code-specific trends (Columns I through III 24,648; Columns IV through VI 19,233). Standard errors in parentheses are clustered by household.						
The mean of the dependent variable is 0.20 in Column I, 0.39 in Column II, 0.47 in Column III, 0.19 in Column IV, 0.38 in Column V, and 0.46 in Column VI.						
* significant at 10%; ** significant at 5%; *** significant at 1%						

Table EA6 presents OLS regression results analogous to those in Table 4 in the main text, but limiting the sample to the observations used in the regressions where we have consumers' actual addresses (the observations used in Columns IV through VI of Table 6 in the main text). The results from Table 4 in the main text using the entire sample are similar to the results in Table EA6 using the restricted sample.[19]

Table EA6: Share of Popular Rentals - OLS Estimates			
	I	II	III
	Top 10	Top 50	Top 100
Share of Offline Rentals	0.3595***	0.4854***	0.4559***
	(0.0005)	(0.0006)	(0.0006)
Total DVD Rentals	-0.0114***	-0.0091***	-0.0068***
	(0.0001)	(0.0001)	(0.0001)
Constant	-0.0155***	-0.0187***	-0.0092***
	(0.0005)	(0.0007)	(0.0007)
Observations	9,087,726	9,087,726	9,087,726
R-squared	0.3344	0.4307	0.4358
Standard errors in parentheses are clustered by household.			
* significant at 10%; ** significant at 5%; *** significant at 1%			

Table EA7 presents first stage regression results analogous to those in Column I of Table 5 in the main text, but limiting the sample to the observations used in the regressions

[19] Note that the number of observations in Table EA6 is similar but not identical to the number of observations in Columns IV to VI of Table 6 in the main text; the explanation is that centroids' latitudes and longitudes from the Census are missing for a few zip codes.

where we have consumers' actual addresses (the observations used in Column II of Table 5 in the main text). The results in Column I of Table 5 in the main text using the entire sample are similar to the results in Table EA7 using the restricted sample. Table EA8 presents second stage regression results analogous to those in Columns I through III of Table 6 in the main text, but limiting the sample to the observations used in the regressions where we have consumers' actual addresses. The results in Columns I through III of Table 6 in the main text using the entire sample are similar to the results in Table EA8 using the restricted sample.

Table EA7: Share of Offline Rentals - First Stage

	I
Number of Stores in the Zip Code Where the Household Resides	0.01945***
	(0.00119)
Number of Stores in Zip Codes between Zero and Five Miles Away	0.00225***
	(0.00047)
Number of Stores in Zip Codes between Five and Ten Miles Away	0.00024
	(0.00028)
Number of Stores in Zip Codes between Ten and Fifteen Miles Away	0.00002
	(0.00026)
Number of Stores in Zip Codes between Fifteen and Twenty Miles Away	-0.0001
	(0.00026)
Number of Stores in Zip Codes between Twenty and Thirty Miles Away	-0.00007
	(0.00017)
Total DVD Rentals	0.02512***
	(0.00011)
Constant	-0.06015***
	(0.00062)
Observations	9,087,726
F-test (global test for the excluded instruments in the second stage)	49.1
R-squared	0.3243
Standard errors in parentheses are clustered by household.	
* significant at 10%; ** significant at 5%; *** significant at 1%	

Table EA8: Share of Popular Rentals - IV Estimates - Second Stage			
	I	**II**	**III**
	Top 10	Top 50	Top 100
Share of Offline Rentals	0.3795***	0.4231***	0.3115***
	(0.0655)	(0.0821)	(0.0854)
Total DVD Rentals	-0.0120***	-0.0075***	-0.0032
	(0.0016)	(0.0021)	(0.0021)
Constant	-0.0143***	-0.0224***	-0.0179***
	(0.0040)	(0.0050)	(0.0052)
Observations	9,087,726	9,087,726	9,087,726
Includes fixed effects for both weeks and individuals, and ZIP code-specific trends.			
Standard errors in parentheses are clustered by household.			
* significant at 10%; ** significant at 5%; *** significant at 1%			

Appendix EA4: Top 2,000 Titles

The focus of this paper is not on examining whether changes in rental concentrations are caused by demand versus supply side long-tail effects. However, in this section we attempt to provide some clues regarding the extent to which differences in title selection across channels causes a change in rental concentrations when consumers move online. In Tables EA9, EA10, and EA11 we present regressions analogous to those in Tables 4, 5, and 6 in the main text, but controlling for the share of rentals taken by the titles below (and also above) the top 2,000 titles—the top 2,000 titles are more likely to be available from both channels. These regressions seek to account for the possibility that consumers may change the proportion of rentals of top 2,000 versus below top 2,000 titles as they move online. Specifically, Tables EA9 and EA11 report estimates of the following model

$$(A)\quad \textit{Share Popular}_{it}=\alpha+\beta\,\textit{Share Offline}_{it}+\gamma\,\textit{Total Rentals}_{it}$$
$$+\theta\textit{Share of Below Top 2000}_{it}+\phi_i+\psi_t+\varphi_z\times t+u_{it}$$

Model (A) is identical to Model (1) in the main text but controls for the variable $\textit{Share of Below Top 2000}_{it}$, which represents the rentals of DVD titles below the top 2,000 as a proportion of all rentals made by household i in week t. The coefficient β in Model (A) measures how weekly changes in the share of DVDs rented from the physical

store relate to weekly changes in the share of popular DVD rentals, while holding fixed the share of rentals of DVD titles below (and also above) the top 2,000 titles.

The results in Tables EA9 and EA11 are similar to those in Tables 4 and 6 in the main text, and show that the long tail effects persist when holding fixed the share of rentals of DVD titles below the top 2,000 titles—when consumers do not change the proportion of rentals of top 2,000 versus below top 2,000 titles as they move online.

While the results from this section may suggest that the long tail effects that we find are not primarily caused by differences in the selection of titles available online versus offline, in this paper we do not focus on examining whether our results are caused by demand versus supply side effects. In the main text we listed several factors preventing us from separating demand and supply side effects (e.g., display of popular products at the brick-and-mortar channel versus the online channel; inventory and title selection online versus offline; online and offline stock-outs; the queue system for online consumption). Identifying how the long tail effects are determined by supply versus demand side effects is an important avenue of future research.

Table EA9: Share of Popular Rentals - OLS Estimates			
	I	**II**	**III**
	Top 10	Top 50	Top 100
Share of Offline Rentals	0.33526***	0.42062***	0.36405***
(mean 0.28)	(0.0003)	(0.0004)	(0.0004)
Total DVD Rentals	-0.01214***	-0.00947***	-0.00726***
(mean 2.25)	(0.0001)	(0.0001)	(0.0001)
Share Below Top 2,000 DVD Rentals	-0.11846***	-0.30290***	-0.41498***
	(0.0002)	(0.0003)	(0.0003)
Constant	-0.01426***	-0.02036***	-0.01202***
	(0.0004)	(0.0004)	(0.0004)
Observations	20,249,021	20,249,021	20,249,021
R-squared	0.3344	0.4637	0.4954
Includes fixed effects for both weeks (30) and individuals (1,529,028), and ZIP code-specific trends (24,648).			
Standard errors in parentheses are clustered by household.			
The mean of the dependent variable is 0.20 in Column I, 0.39 in Column II, and 0.47 in Column III.			
* significant at 10%; ** significant at 5%; *** significant at 1%			

Table EA10: Share of Offline Rentals - First Stage

	I	II
Number of Stores in the Zip Code Where the	0.02168***	na
Household Resides (mean 0.52)	(0.0008)	na
Number of Stores in Zip Codes between Zero	0.00263***	na
and Five Miles Away (mean 2.17)	(0.0003)	na
Number of Stores in Zip Codes between Five	0.00011	na
and Ten Miles Away (mean 5.23)	(0.0002)	na
Number of Stores in Zip Codes between Ten	0.0001	na
and Fifteen Miles Away (mean 6.56)	(0.0002)	na
Number of Stores in Zip Codes between Fifteen	-0.00015	na
and Twenty Miles Away (mean 6.89)	(0.0002)	na
Number of Stores in Zip Codes between Twenty	-0.00013	na
and Thirty Miles Away (mean 12.80)	(0.0001)	na
Distance to the Closest Store in Miles	na	-0.00840***
(mean 3.6)	na	(0.0004)
Squared Distance to the Closest Store in Miles	na	0.00011***
	na	(0.00001)
Total DVD Rentals	0.02347***	0.02307***
(mean 2.25 in Columns I and II)	(0.0001)	(0.0001)
Share Below Top 2,000 DVD Rentals	-0.31690***	-0.28279***
	(0.0004)	(0.0006)
Constant	-0.02194***	-0.02798***
	(0.0004)	(0.0006)
Observations	20,249,021	9,141,694
F-test (global test for the excluded instruments in the second stage)	143.5	333.0
R-squared	0.3497	0.3669

Includes fixed effects for both weeks (30) and individuals (Column I 1,529,028; Column II 855,996), and
ZIP code-specific trends (Column I 24,648; Column II 19,233). Standard errors in parentheses are clustered by household.
The mean of the dependent variable is 0.28 in Column I and 0.25 in Column II.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table EA11: Share of Popular Rentals - IV Estimates - Second Stage

	I	II	III	IV	V	VI
	Top 10	Top 50	Top 100	Top 10	Top 50	Top 100
Share of Offline Rentals	0.36864***	0.31207***	0.20163***	0.39445***	0.31519***	0.24806***
(mean 0.28 in Columns I through III	(0.0383)	(0.0453)	(0.0454)	(0.0419)	(0.0530)	(0.0522)
and 0.25 in Columns IV through VI)						
Total DVD Rentals	-0.01292***	-0.00692***	-0.00345***	-0.01304***	-0.00724***	-0.00489***
(mean 2.25 in Columns I through VI)	(0.0009)	(0.0011)	(0.0011)	(0.0010)	(0.0012)	(0.0012)
Share Below Top 2,000 DVD Rentals	-0.10788***	-0.33730***	-0.46645***	-0.09844***	-0.32817***	-0.44280***
	(0.0121)	(0.0144)	(0.0144)	(0.0119)	(0.0150)	(0.0148)
Constant	-0.01353***	-0.02273***	-0.01556***	-0.00974***	-0.01949***	-0.01105***
	(0.0009)	(0.0011)	(0.0011)	(0.0013)	(0.0016)	(0.0016)
Observations	20,249,021	20,249,021	20,249,021	9,141,694	9,141,694	9,141,694

Includes fixed effects for both weeks (30) and individuals (Columns I through III 1,529,028; Columns IV through VI 855,996), and
ZIP code-specific trends (Columns I through III 24,648; Columns IV through VI 19,233). Standard errors in parentheses are clustered by household.
The mean of the dependent variable is 0.20 in Column I, 0.39 in Column II, 0.47 in Column III, 0.19 in Column IV, 0.38 in Column V, and 0.46 in Column VI.
* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix EA5: Popular, Fringe Popular, and Niche Titles

Our results in the main text show that consumers decrease their rentals of top 10, top 50, and top 100 DVD titles when they move to online markets. It may also be useful to examine in a more granular way whether movement to the online channel induces consumers to increase the rental of niche titles at the expense of the very top titles only, or whether the increase in rentals of niche titles when consumers move to the online channel is at the expense of fringe popular titles as well. We separate the titles into three categories (Popular, Fringe Popular, and Niche) and define the following variables for each household *i* in each week *t*: $Share\ Popular\ in\ Range\ [First,\ Top\ I]_{it}$ represents the number of rentals for titles ranked between the title ranked first and the title ranked *Top I* divided by the total number of rentals, $Share\ \ Fringe\ Popular\ in\ Range\ [Top\ I{+}1,\ Top\ II]_{it}$ represents the number of rentals for titles ranked between the title ranked *Top I+1* and the title ranked *Top II* divided by the total number of rentals, and $Share\ Niche\ in\ Range\ [Top\ II{+}1,\ Last]_{it}$ represents the number of rentals for titles ranked between the title ranked *Top II+1* and the title ranked last divided by the total number of rentals (*Top II>Top I)*. For example, when *Top I* is five and *Top II* is ten the variable $Share\ Popular\ in\ Range\ [First,\ Top\ I]_{it}$ represents the share of rentals taken by the titles ranked between the title ranked first and the title ranked fifth, $Share\ Fringe\ Popular\ \ in\ Range\ [Top\ I{+}1,\ Top\ II]_{it}$ represents the share of rentals taken by the titles ranked between the title ranked sixth and the title ranked tenth, and $Share\ Niche\ in\ Range\ [Top\ II{+}1,\ Last]_{it}$ represents the share of rentals taken by all remaining titles in the tail of the distribution beginning from the title ranked eleventh. We note that:

$$Share\ Popular\ in\ Range\ [First,\ Top\ I]_{it} +$$
$$Share\ Fringe\ Popular\ \ in\ Range\ [Top\ I{+}1,\ Top\ II]_{it}$$
$$+ Share\ Niche\ in\ Range\ [Top\ II{+}1,\ Last]_{it} = 1$$

We seek to examine whether consumers increase their rentals of $Share\ Fringe\ Popular\ in\ Range\ [Top\ I{+}1,\ Top\ II]_{it}$ and decrease their rentals of

$Share\ Niche\ in\ Range\ [Top\ II+1,\ Last]_{it}$ when they move to online channels. We estimate the following model

$$(B)\quad Share\ Fringe\ Popular\ in\ Range\ [Top\ I+1,\ Top\ II]_{it}=\alpha+\beta\ Share\ Offline_{it}+\gamma\ Total\ Rentals_{it}+\theta Share\ Popular\ in\ Range\ [First,\ Top\ I]_{it}+\phi_i+\psi_t+\varphi_z\times t+u_{it}$$

Model (B) controls for the variable $Share\ Popular\ in\ Range\ [First,\ Top\ I]_{it}$. We hold fixed the share of rentals taken by titles at the head of the distribution in Model (B) to identify movement away from the rental of fringe popular titles and toward the rental of titles in the tail of the distribution as consumers move online. The coefficient β in Model (B) measures how weekly changes in the share of DVDs rented from the physical store relate to weekly changes in the share of DVD rentals taken by the titles ranked between the title ranked *Top I+1* and the title ranked *Top II*, while holding fixed the share of rentals taken by the very top DVD titles.

Tables EA12 through EA16 present OLS and first and second stage Instrumental Variables regression results, and in these tables Columns I through IV report estimates for DVD titles ranked between 6 and 10, 11 and 20, 21 and 50, and 51 and 100 respectively. When interpreting the size of the coefficients on the share of offline rentals it should be noted that the means of the dependent variables are different across columns and tables. For example, Column I of Table EA12 indicates that when consumers decrease the share of rentals made at physical stores from 0.28 to zero (when consumers move entirely to the online channel) they decrease the share of rentals of titles ranked between 6 and 10 by 4.0 percentage points (0.14 times 0.28). Noting that the mean of the dependent variable in Column I of Table EA9 is 0.069 then indicates that when consumers move entirely to the online channel there is a 58% percentage decrease in the rental of titles ranked between 6 and 10 (0.040/0.069). Columns II through IV of Table EA12 indicate that when consumers decrease the share of rentals made at physical stores from 0.28 to zero they decrease the share of rentals of titles ranked between 11 and 20, 21 and 50, and 51 and 100 by 59%, 38%, and 23% respectively.

The OLS regression results in Table EA12 show that for the various ranges of titles ranked among the top 100 that we considered there is a decrease in the share of fringe

popular title rentals and an increase of niche title rentals when consumers move online. Conversely, only the first two columns in Tables EA15 and EA16 presenting the Instrumental Variable second stage regression results show that fringe popular titles decrease at the expense of niche titles when consumers move online. Columns I and II in Table EA15 (Columns I and II in Table EA16) indicate that when consumers decrease the share of rentals made at physical stores from 0.28 (0.25)[20] to zero they decrease the share of rentals of titles ranked between 6 and 10, and 11 and 20 by 59% and 24% respectively (71% and 37% respectively). Columns III and IV in Table EA15 and EA16 do not show a statistically significant change in the share of rentals taken by the titles ranked between 21 and 50 or between 51 and 100 at the expense of the rentals of niche titles when consumers move online.

The Instrumental Variables results in this appendix section suggest that the long tail effects that we find are concentrated around the top 20 titles. These results are consistent with the results in the main text showing that the long tail effects are smaller when considering the top 50 and top 100 titles than when considering the top 10 titles (e.g., see page 17 in the main text). These results are also expected since the rentals from the focal company's stores are very concentrated at the top of the distribution: the top 20 titles represent 63.3% of all rentals from the store (the titles ranked between top 21 and top 100 represent 21.3% of all rentals from the store, and the titles ranked between top 51 and top 100 take 7.0% of the rentals from the store; see Table 2 in the main text).

Because our focal company does not display or promote the very popular products as heavily in the online channel as it does in the brick-and-mortar channel, the results in this section may provide some indication that promotional effects and the ways that people search at physical stores versus online are an important mechanism underlying our results. In addition, stock-outs of popular titles are frequent both online and offline for our focal company, and long tail effects might result if stock-outs online were more frequent than at brick and mortar stores. However, to the extent that stock-outs from the online channel are concentrated around the top 5 or top 10 most popular titles in every

[20] The mean of the share of offline rentals in Table EA16 is 0.25 as in the regressions in the main text using the sample of consumers where we have addresses.

week the results in this section may suggest that the long tail effects that we find are not primarily due to stock-outs.[21]

Table EA12: Share of Fringe Popular Rentals - OLS Estimates				
	I	**II**	**III**	**IV**
	Top 6 - Top 10	Top 11 - Top 20	Top 21 - Top 50	Top 51 - Top 100
Share of Offline Rentals	0.14609***	0.16868***	0.14918***	0.06836***
	(0.0002)	(0.0002)	(0.0003)	(0.0002)
Total DVD Rentals	-0.00333***	-0.00206***	-0.00021***	0.00063***
	(0.0000)	(0.0000)	(0.0001)	(0.0000)
Share Top 5	-0.15738***	na	na	na
	(0.0002)	na	na	na
Share Top 10	na	-0.19223***	na	na
	na	(0.0002)	na	na
Share Top 20	na	na	-0.25034***	na
	na	na	(0.0002)	na
Share Top 50	na	na	na	-0.19488***
	na	na	na	(0.0002)
Constant	-0.01224***	-0.01046***	-0.00375***	0.00448***
	(0.0002)	(0.0003)	(0.0003)	(0.0003)
Observations	20,249,021	20,249,021	20,249,021	20,249,021
R-squared	0.1694	0.1864	0.2162	0.1992
Includes fixed effects for both weeks (30) and individuals (1,529,028), and ZIP code-specific trends (24,648).				
Standard errors in parentheses are clustered by household.				
The mean of the dependent variable is 0.069 in Column I, 0.080 in Column II, 0.109 in Column III, and 0.082 in Column IV.				
* significant at 10%; ** significant at 5%; *** significant at 1%				

[21] In addition, our analysis in footnote 7 in the main paper suggests that stock outs are not more frequent online than offline.

Table EA13: Share of Offline Rentals - First Stage				
	I	II	III	IV
	Offline Rentals	Offline Rentals	Offline Rentals	Offline Rentals
Number of Stores in the Zip Code Where the	0.01974***	0.01875***	0.01875***	0.01963***
Household Resides (mean 0.52)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Number of Stores in Zip Codes between Zero	0.00263***	0.00254***	0.00230***	0.00212***
and Five Miles Away (mean 2.17)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Number of Stores in Zip Codes between Five	0.00028	0.00025	0.0002	0.00002
and Ten Miles Away (mean 5.23)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Number of Stores in Zip Codes between Ten	0	-0.00004	-0.00006	-0.00007
and Fifteen Miles Away (mean 6.56)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Number of Stores in Zip Codes between Fifteen	-0.0001	-0.00007	-0.00012	-0.00012
and Twenty Miles Away (mean 6.89)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Number of Stores in Zip Codes between Twenty	-0.00027**	-0.00023**	-0.00027***	-0.00024**
and Thirty Miles Away (mean 12.80)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Total DVD Rentals	0.02746***	0.02707***	0.02589***	0.02446***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Share Top 5	0.40856***	na	na	na
	(0.0005)	na	na	na
Share Top 10	na	0.41825***	na	na
	na	(0.0004)	na	na
Share Top 20	na	na	0.42935***	na
	na	na	(0.0004)	na
Share Top 50	na	na	na	0.40150***
	na	na	na	(0.0004)
Constant	-0.02017***	-0.01408***	-0.00950***	-0.01079***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Observations	20,249,021	20,249,021	20,249,021	20,249,021
F-test (global test for the excluded instruments in the second stage)	126.8	116.3	114.6	119.7
R-squared	0.3684	0.4013	0.4341	0.4343
Includes fixed effects for both weeks (30) and individuals (1,529,028), and ZIP code-specific trends (24,648).				
Standard errors in parentheses are clustered by household.				
* significant at 10%; ** significant at 5%; *** significant at 1%				

Table EA14: Share of Offline Rentals - First Stage				
	I	II	III	IV
	Offline Rentals	Offline Rentals	Offline Rentals	Offline Rentals
Distance to the Closest Store in Miles	-0.00794***	-0.00744***	-0.00727***	-0.00763***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Squared Distance to the Closest Store in Miles	0.00010***	0.00010***	0.00009***	0.00010***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Total DVD Rentals	0.02641***	0.02613***	0.02517***	0.02402***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Share Top 5	0.39650***	na	na	na
	(0.0007)	na	na	na
Share Top 10	na	0.39997***	na	na
	na	(0.0007)	na	na
Share Top 20	na	na	0.40390***	na
	na	na	(0.0006)	na
Share Top 50	na	na	na	0.36947***
	na	na	na	(0.0006)
Constant	-0.02613***	-0.01997***	-0.01578***	-0.01772***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Observations	9,141,694	9,141,694	9,141,694	9,141,694
F-test (global test for the excluded instruments in the second stage)	326.9	281.9	269.2	280.8
R-squared	0.3891	0.4176	0.4448	0.4417
Includes fixed effects for both weeks (30) and individuals (855,996), and ZIP code-specific trends (19,233).				
Standard errors in parentheses are clustered by household.				
* significant at 10%; ** significant at 5%; *** significant at 1%				

Table EA15: Share of Fringe Popular Rentals - IV Estimates				
	I	**II**	**III**	**IV**
	Top 6 - Top 10	Top 11 - Top 20	Top 21 - Top 50	Top 51 - Top 100
Share of Offline Rentals	0.14089***	0.06873**	0.03827	-0.03871
	(0.0276)	(0.0309)	(0.0359)	(0.0311)
Total DVD Rentals	-0.00318***	0.00065	0.00267***	0.00325***
	(0.0008)	(0.0008)	(0.0009)	(0.0008)
Share Top 5	-0.15525***	na	na	na
	(0.0113)	na	na	na
Share Top 10	na	-0.15043***	na	na
	na	(0.0129)	na	na
Share Top 20	na	na	-0.20272***	na
	na	na	(0.0154)	na
Share Top 50	na	na	na	-0.15189***
	na	na	na	(0.0125)
Constant	-0.01234***	-0.01186***	-0.00479***	0.00333***
	(0.0006)	(0.0005)	(0.0005)	(0.0004)
Observations	20,249,021	20,249,021	20,249,021	20,249,021
Includes fixed effects for both weeks (30) and individuals (1,529,028), and ZIP code-specific trends (24,648).				
Standard errors in parentheses are clustered by household.				
The mean of the dependent variable is 0.069 in Column I, 0.080 in Column II, 0.109 in Column III, and 0.082 in Column IV.				
* significant at 10%; ** significant at 5%; *** significant at 1%				

Table EA16: Share of Fringe Popular Rentals - IV Estimates				
	I	**II**	**III**	**IV**
	Top 6 - Top 10	Top 11 - Top 20	Top 21 - Top 50	Top 51 - Top 100
Share of Offline Rentals	0.18679***	0.11502***	0.02353	0.03565
	(0.0274)	(0.0318)	(0.0398)	(0.0348)
Total DVD Rentals	-0.00448***	-0.00088	0.00253**	0.0012
	(0.0007)	(0.0008)	(0.0010)	(0.0008)
Share Top 5	-0.16698***	na	na	na
	(0.0109)	na	na	na
Share Top 10	na	-0.16446***	na	na
	na	(0.0127)	na	na
Share Top 20	na	na	-0.19645***	na
	na	na	(0.0161)	na
Share Top 50	na	na	na	-0.18342***
	na	na	na	(0.0128)
Constant	-0.01100***	-0.01037***	-0.00316***	0.00587***
	(0.0008)	(0.0007)	(0.0008)	(0.0008)
Observations	9,141,694	9,141,694	9,141,694	9,141,694
Includes fixed effects for both weeks (30) and individuals (855,996), and ZIP code-specific trends (19,233).				
Standard errors in parentheses are clustered by household.				
The mean of the dependent variable is 0.065 in Column I, 0.077 in Column II, 0.110 in Column III, and 0.084 in Column IV.				
* significant at 10%; ** significant at 5%; *** significant at 1%				