Online pricing dynamics in Internet retailing: The case of the DVD market

Baibing Li a, *, Fang-Fang Tang b, c

a Business School, Loughborough University, Loughborough LE11 3TU, United Kingdom
b Experimental Economics Lab, Southwest University of Finance and Economics, Chengdu, PR China
c National Policy Lab, China Center for Economic Research, National School of Development, Peking University, Beijing, PR China

Abstract

The explosive growth of Internet retailing offers an excellent opportunity to collect online prices at a disaggregate (e.g., individual store and/or individual product) level over time and to investigate the evolution of Internet markets. In this paper, we generalize the results obtained in existing static analyses and develop two random coefficient regression models to capture the dynamics of prices in the US online DVD market. On the basis of the models, we test hypotheses to compare the rates of change in price levels and in price dispersion at both pure dotcoms and online branches of multichannel retailers in the DVD market. The results, based on the analysis of 6759 price quotes over a 12-month period, suggest that multichannel retailers effectively differentiated themselves from pure dotcoms on nonprice dimensions so that they charged higher prices and maintained the difference in price levels throughout the time period of the study. Head-to-head price competition within pure dotcoms tended to be more severe. Our results also suggest that there is a sign of maturity in the current online TVD market.

1. Introduction

Although in the early years of Internet retailing it was widely predicted that online marketing would lead to frictionless e-commerce (Alba et al. 1997, Bakos 1997), a considerable number of recent studies have overwhelmingly shown that this is not true (see e.g., Lal and Sarvary 1999, Brynjolfsson and Smith 2000, Pan et al. 2004). This study shifts the research focus and by investigating pricing behavior aims to gain a better understanding of how different types of retailer compete with each other in online markets.

The issue of online pricing is of particular importance in the online marketing research. This is because pure dotcoms tend to differentiate themselves from other types of retailer via flexibly pricing their products. In addition, the competition among dotcoms also tends to be on the price dimension. Such competition leads to substantial price dispersion in the Internet markets. It is thus crucial for researchers to understand the characteristics of online pricing behavior and how it evolves over time. The results on this research issue also have important managerial implications.

Most of the earlier empirical studies performed a static analysis where price competition was measured in terms of price levels and price dispersion. From the perspective of marketing research, both price levels and price dispersion are summaries of the price distribution in a market that reflects how retailers interact with each other. In these empirical studies price levels and price dispersion were compared between bricks-and-mortar (traditional) and online retailers (Bailey 1998, Brynjolfsson and Smith 2000, Erevelles et al. 2001, Clay et al. 2002). Further studies focused on comparisons of various retailing channels. They compared pure Internet retailers (hereafter dotcoms) and online branches of multichannel (hereafter multichannel) retailers (Tang and Xing 2001); or traditional retailers, dotcoms, and multichannel retailers (Ancarani and Shankar 2004). These studies resulted in some interesting findings that suggest substantial differences in pricing behavior among different retailing channels.

It has also been recognized that the stage of development of Internet retailing has a substantial influence on the pricing behavior of retailers. In the early stage, for instance, online retailers priced products at a higher-level than traditional retailers (Bailey 1998, Erevelles et al. 2001). As Internet markets developed in the early years of this century, online retailers substantially lowered their prices. During the transition period there was a mixture of findings, some of which contradicted each other. For example, Clay et al. (2002) did not find any significant difference in prices between online retailers and traditional retailers, whereas Brynjolfsson and Smith (2000) compared prices of CDs and books and found that online retailers had a lower price level than traditional retailers. There were also conflicting results on price dispersion. See Pan et al. (2004) for a comprehensive review.

Internet markets are now more mature. It is thus of interest to investigate which of the earlier findings on Internet retailing can be generalized to the current online markets. In addition, since the majority of the existing researches were carried out at a fixed
time-point, it is of interest to investigate which of the findings in these static analyses can be generalized to a longer time period so that the evolution of online prices over time can be investigated.

The emergence of Internet data sources offers an impetus to the development of dynamic models that capture price dynamics (Dekimpe and Hanssens 2000, Pauwels et al. 2004). Consequently, recent studies on online pricing have used more sophisticated dynamic approaches. As Pauwels et al. (2004) have pointed out, however, few existing studies in marketing research recognize that the neglect of heterogeneity across the entities over which the data are averaged is a serious issue in dynamic modeling. For instance, in the recent analysis in Xing et al. (2006), cross-sectional heterogeneity was absent and the correlation of the prices posted at different retailers for the same product (e.g., a particular DVD title) during the same time period was ignored. Statistically, when aggregation bias is not addressed properly, it may result in parameter estimates being inconsistent, inefficient, and/or biased (Pauwels et al. 2004).

This paper incorporates a sophisticated statistical technique to address these econometric issues. On the basis of our models, we focus on the pricing dynamics in online market evolution and investigate how different types of retailer compete with each other in an online market, the US DVD market.

The US DVD market is chosen for several reasons. First, it is generally considered that DVDs are relatively homogeneous goods and thus likely to experience strong price competition given the characteristics of Internet channels (see e.g., Bakos 1997, Lal and Sarvary 1999, Brynjolfsson and Smith 2000, Harrington 2001, Tang and Xing 2001, Iyer and Pazgal 2003, Xing et al. 2006). Secondly, there is a rich literature on the US online DVD market so it is easy to compare and contrast the findings of this study with other results, and in particular to compare the current price dynamics with those presented in Xing et al. (2006). In addition, it is more straightforward to compare DVDs because they are relatively homogeneous. For instance, prices of identical DVDs at different retailers can be compared directly. This is not the case for goods such as clothes, shoes, and electronics where there are many styles and/or models, and similar products may differ from each other to a considerable extent. Finally, the US online DVD market has a long history of Internet retailing and is likely to be more mature than other markets.

The existing static analyses have revealed some interesting results on online marketing. Tang and Xing (2001) found that prices at dotcoms were significantly lower than prices at multichannel retailers. In addition, the corresponding price dispersion was much lower among dotcoms than among multichannel retailers. Contrary to Tang and Xing (2001), Pan et al. (2003) found that multichannel retailers generally had smaller price dispersion than did dotcoms. Ancarani and Shankar (2004) argued that multichannel retailers can combine the benefits of online shopping with physical inspection, pickup, and return of merchandise via support from their offline stores. In their static analysis they suggested that multichannel retailers may effectively differentiate themselves from dotcoms on nonprice dimensions and charge higher prices. Recently Xing et al. (2006) have investigated the dynamics of online prices in the US DVD market. On the basis of the online price data in the US DVD market collected during years 2000–2001, they have found that multichannel retailers charge higher prices than dotcoms and prices go up with time for both multichannel and dotcom retailers. In addition, prices of dotcoms go up faster than those of multichannel retailers.

In this paper we shall investigate which of these earlier findings can be generalized to the current online DVD market and can be generalized from a given time-point to a longer time period. More importantly, if there exists a difference in price levels between different types of retailer at a given time-point, we shall investigate whether the difference is maintained across the time period.

To reveal the competitive pricing behavior of retailers, two dynamic models will be built at an individual product level, one model for price levels and the other for price dispersion. The nature of the data collected in this study raises several challenging issues for dynamic modeling, including extremely high dimensionality, and cross-sectional heterogeneity and the associated random effects. As indicated in Dekimpe and Hanssens (2000) and Pauwels et al. (2004), it is difficult to address these issues in the framework of the widely used VARX approach. Hence, in this paper we shall consider an alternative approach, random coefficient regression models, to analyze pricing dynamics at an individual product level where the issues of time correlation and cross-sectional heterogeneity can be easily dealt with. We can also link marketing characteristics directly to the rate of change in price levels and in price dispersion so that the research issues of interest can be investigated.

The next section is devoted to data collection and summary statistics. We then develop our main research questions. Then we build econometric models and test the formulated hypotheses. Finally we summarize the main results and discuss the managerial implications.

2. Data collection and summary statistics

2.1. Data collection

In this study, we investigate the US online DVD market. Two types of online retailers are included in this study: dotcoms and multichannel retailers. All the included retailers sell a general selection of DVD titles and the prices are posted on their websites. Following Brynjolfsson and Smith (2000), Tang and Xing (2001), Ancarani and Shankar (2004), and Xing et al. (2006), the retailers in this study were chosen by their high ranks in the PowerRanking for Movies by Forrester Research and the DVD Talk Online store listings (http://www.dvdtalk.com). In total five dotcoms and five multichannel retailers were selected. Together, the market share of these online retailers is substantial, ensuring that their pricing behavior represents the US online DVD market.

With regard to DVD titles, we selected an even mix of bestsellers and non-bestsellers. The bestsellers (“popular titles”) were selected from the lists of bestsellers available when the study was initiated. Non-bestsellers were chosen by randomly selecting pages from an English dictionary and finding a title starting with a word on the page (“random titles”).

In the beginning of the study, two lists of DVD titles were considered, one for the popular titles and the other for the random titles, each included 26 DVD titles. During the study period some of the selected popular titles became non-bestsellers and some of the selected random titles were no longer posted by the retailers. Whenever this happened, another DVD title was selected from the corresponding categories (popular or random) to replace the one that had disappeared to ensure the total of 26 DVD titles for each list was retained. In the end, there were 32 different popular titles and 29 different random titles on the two lists across the whole time period, although the total number of the titles on each list was kept to be 26 at any time-point of the study period.

For the 61 selected titles, we collected online price quotes weekly at the selected dotcoms and multichannel retailers from April 3, 2004 to March 5, 2005. This resulted in a total of 29,796 price quotes (excluding missing values) over the time span of the year.
2.2. Summary statistics

The means and standard deviations of the prices by retailer type and DVD type are displayed in Table 1. It can be seen that on average dotcoms have lower prices than multichannel retailers, overall and for both popular and random title categories. The shipping costs per item were calculated on the basis of shipping rate tables for various baskets of typical purchases. It was found that there was no significant difference in shipping costs between the two retailer types. This paper will therefore focus on the posted prices and omit further consideration of the shipping costs.

It is evident from Table 1 that the magnitudes of the standard deviations of the prices are related to the magnitudes of the means of the prices. This suggests that the logarithm transformation is a suitable choice. We thus consider log-prices in the remainder of this paper. It can also be seen that the prices vary across individual retailers within each retailer type. For each DVD title, we will investigate both the means and standard deviations of the log-prices across individual retailers within each retailer type.

Further, we carried out a pilot analysis based on the basic descriptive statistics with graphical presentation of the data. It turned out retailers hardly changed their prices within a month. This is in line with the findings of Xing et al. (2006). Hence, monthly prices—recorded during the first week of each month from April 3, 2004 to March 5, 2005—were used in this study. After the missing values were excluded, there were 6759 price quotes.

The individual monthly log-prices observed over the year are displayed in Fig. 1. Overall, it seems that prices posted at multichannel retailers were higher than those at dotcoms over the year, and prices for popular titles were higher than those for random titles. However, it is hard to detect if there were any differences in the rate of change in prices posted at different retailer types.

To reveal patterns in the price data, we calculated the mean and standard deviation of the log-prices of each title across all dotcom (or multichannel) retailers in each month, and then averaged them across DVD titles by DVD type and retailer type. The results are displayed in Fig. 2a and b, respectively.

For the mean log-prices shown in Fig. 2a, the plot shows strong downward trends over time for both retailer types and both DVD types, clearly indicating that overall prices went down gradually over time. It is worth noting that the differences in price levels between dotcoms and multichannel retailers were maintained over time.

On the other hand, the standard deviations of the log-prices shown in Fig. 2b appear to exhibit different patterns. The spreads of the log-prices among multichannel retailers became stable or reduced gradually, indicating that the prices at multichannel retailers might be starting to converge; whereas the spreads of the log-prices among dotcoms show a clear increasing trend, indicating that dotcoms tended to post prices at more diverse levels. The two graphs have thus provided us with insightful information for further econometric modeling and these observations will be tested formally in the following analysis.

3. Hypothesis development

In the literature, the comparison of pricing behavior between different retailer types is normally based on two measures: price level and price dispersion. In this study we shall compare price levels at dotcoms and multichannel retailers to investigate the competition across different channels in the US online DVD market.

Price dispersion, on the other hand, is a measure of the variability in prices and it characterizes the alternative offerings in the market (Pan et al. 2004). The price dispersion within each retailer type reflects the price competition within that channel (Ancarani and Shankar 2004). Hence, we shall also investigate price dispersion within dotcoms and within multichannel retailers.

3.1. Competition between dotcoms and multichannel retailers

In the initial stage of Internet marketing, research focused on the comparison between traditional retailers and online retailers (Bailey 1998, Brynjolfsson and Smith 2000, Erevelles et al. 2001, Clay et al. 2002). Compared with the traditional commerce, there are some exclusive factors that may affect the pricing behavior in online markets. For instance, online price levels may depend on channel-specific price sensitivity, stage of development of the Internet channels, Internet reach, extent of digital attributes in the product, service levels, and search costs (Lal and Sarvary 1999, Ancarani and Shankar 2004, Xing et al. 2006).

When turning to the comparison between dotcoms and multichannel retailers, however, the above factors affect both retailer types in a similar manner since both are online retailers. So given the fact that they are both online retailers, it is of interest to investigate if there is any difference between multichannel retailers and dotcoms in terms of pricing. If the answer is yes, then what other factors influence their pricing behavior? This issue has been explored from both theoretical and empirical aspects in the existing literature.

From the perspective of marketing research, multichannel retailers have some advantages over dotcoms that they may exploit to effectively differentiate themselves from dotcoms. First, as argued by Lal and Sarvary (1999), Ancarani and Shankar (2004), and Xing et al. (2006), multichannel retailers can combine the benefits of online shopping with support from their offline stores such as physical inspection, pickup, and return of merchandise. Secondly, multichannel retailers are usually long established with greater market power and customer loyalty. Hence, multichannel retailers may translate their market power and brand names from offline to online modes and thus charge higher prices without losing all their customers to dotcom competitors (Xing et al. 2006). Dotcoms, on the other hand, do not maintain stores and thus their operating costs are much lower. In addition, multichannel retailers may wish to coordinate prices across their different channels to prevent destructive internal competition (Pan et al. 2004, Xing et al. 2006). Consequently, dotcoms have more flexibility in pricing strategy. Thus, the two retailer types may use different pricing strategies. This has been documented in many static analyses (Tang and Xing 2001, Pan et al. 2002, 2003; Ancarani and Shankar 2004) and in the dynamic analysis in Xing et al.
It is also suggested by the preliminary analysis in the previous section (see Fig. 1). Thus, we propose:

**Hypothesis 1.** On average, the price level of DVD titles at multichannel retailers is higher than that at dotcoms at any time-point in the US online market.

Although some previous studies did not consider the impact of different DVD types, we expect that the dynamics of pricing behavior for popular titles may differ from that for non-popular titles. This is based on the observation that the demand for popular titles is high when they are newly launched and will diminish as their rankings drop and new titles become more popular. The demand for non-popular titles may have different dynamics although their prices may also decline over time. Hence, we follow Tang and Xing (2001), Clay et al. (2002), and Xing et al. (2006) and distinguish popular titles from non-popular titles. This will also make the definitions of mean and standard deviation more appropriate, and the corresponding results more relevant from the perspective of marketing research.

It seems reasonable to expect that both dotcoms and multichannel retailers charge higher prices for popular titles. This is also suggested by the preliminary analysis in the previous section (see Fig. 1). Thus, we propose:

**Hypothesis 2a.** On average, at dotcoms the price level of popular DVD titles is higher than that of random titles at any time-point in the US online market.

**Hypothesis 2b.** On average, at multichannel retailers the price level of popular DVD titles is higher than that of random titles at any time-point in the US online market.

Next we investigate whether or not the difference in price levels hypothesized at a given time-point can be maintained over time. As mentioned earlier, existing static analyses have shown that long-established multichannel retailers have greater market power than dotcoms and their online branches can get support from their offline stores. Hence, they may charge higher prices. The question is whether this difference will decrease or even disappear because of low search costs in the online market. If the difference in price levels does not decrease, will it be maintained at the same level? Xing et al. (2006), based on the data collected in an earlier year, predict that the difference in price levels will disappear and the two retailer types will have similar pricing behavior. On the basis of the results in Xing et al. (2006), we tentatively propose:

**Hypothesis 3.** Although some previous studies did not consider the impact of different DVD types, we expect that the dynamics of pricing behavior for popular titles may differ from that for non-popular titles. This is based on the observation that the demand for popular titles is high when they are newly launched and will diminish as their rankings drop and new titles become more popular. The demand for non-popular titles may have different dynamics although their prices may also decline over time. Hence, we follow Tang and Xing (2001), Clay et al. (2002), and Xing et al. (2006) and distinguish popular titles from non-popular titles. This will also make the definitions of mean and standard deviation more appropriate, and the corresponding results more relevant from the perspective of marketing research.
Hypthesis 3. The average difference in the log-prices of DVD titles between dotcoms and multichannel retailers will disappear in the US online market.

3.2. Competitions among dotcoms and among multichannel retailers

As mentioned earlier, multichannel retailers may differentiate themselves from dotcoms on nonprice dimensions and price their products at a higher-level. Given the severe competition from multichannel retailers on the nonprice dimensions, we are concerned with how dotcoms compete with each other in the online market.

Research in the early stage of Internet marketing focused on the comparison between traditional retailers and online retailers. The literature shows conflicting results on price dispersion. Some studies show that price dispersion was higher online than that among traditional retailers (e.g., Bailey 1998, Erevelles et al. 2001, Clay et al. 2002) but others show the opposite (Brynjolfsson and Smith 2000, Pan et al. 2003). Pan et al. (2004) have provided a comprehensive review of price dispersion in Internet retailing and showed substantial price dispersion observed on the Internet.

Tang and Xing (2001) showed that price dispersion was lower for dotcoms but Pan et al. (2003) found that the standard deviation of prices among multichannel retailers was higher. Overall it seems that the results are inconclusive (Pan et al. 2004).

We argue that the mixed findings on price dispersion may be a result of the immaturity of Internet markets when these studies were carried out. It is clear that Internet markets have been gradually evolving. For instance, Pan et al. (2003) found that online price dispersion decreased from 2000 to 2001 but increased from 2001 to 2003. Xing et al. (2006) showed that at the beginning of their study, price dispersion at dotcoms was smaller but it increased and tended to have a similar level to that at multichannel retailers.

We note that the analysis of Xing et al. (2006) was based on the data collected in 2001. After several years of severe price competition, it is likely that dotcoms are becoming more and more heterogeneous and thus for the current DVD market the price dispersion within dotcoms may have become even larger. In particular, some dotcoms have established their reputation and obtained a considerable market share. They have thus gained substantial market power and can raise prices without losing all their customers to their competitors. On the other hand, new dotcoms are launched every year because online markets have easier entry than offline markets (Brynjolfsson and Smith 2000, Pan et al. 2004). We also argue that multichannel retailers can differentiate themselves from dotcoms on nonprice dimensions whereas dotcoms have to differentiate themselves mainly on price. Thus we propose:

Hypothesis 4a. The spread measured by the standard deviation of log-prices of DVD titles among dotcoms is increasing over time in the US online market.

We believe that multichannel retailers are much more homogeneous. They are long established and have a considerable market share. Their reputations may help them to retain customers and they can differentiate themselves from dotcoms on nonprice dimensions. Thus we propose:

Hypothesis 4b. The spread measured by the standard deviation of log-prices of DVD titles among multichannel retailers is decreasing or stable over time in the US online market.

Lee and Gosain (2002) compared online and offline markets and showed that the degree of price dispersion depends on the product type. Clay et al. (2002) showed that relative to random books, best-sellers had a higher standard deviation because retailers chose to aggressively discount bestsellers but not random books. On the other hand, Tang and Xing (2001) focused on dotcoms and multichannel retailers. They found that for multichannel retailers the price dispersion of popular DVD titles was higher than that of random titles, whereas for dotcoms the difference between the two categories was indistinguishable. Following the results in Tang and Xing (2001), we tentatively propose:

Hypothesis 5a. At any time-point the spread measured by the standard deviation of log-prices among dotcoms for popular titles is the same as that for random titles in the US online market.

Hypothesis 5b. At any time-point the spread measured by the standard deviation of log-prices among multichannel retailers for popular titles is higher than that for random titles in the US online market.

4. Econometric models

4.1. Features of the collected data

As described earlier, the data in this study were collected at the individual retailer level and at the individual title level. Clearly the price of a title posted at a retailer during a certain time period is correlated with the price of the same title posted at the same retailer a few periods later (time correlation). It is also correlated with the prices of the same title posted at other retailers in the same time period (spatial correlation or intra-class correlation). Fig. 3 illustrates these two types of correlation. Cross-sectional heterogeneity and aggregation bias therefore become important issues in the modeling. Unfortunately, this was ignored in most existing research (Pauwels et al. 2004).

4.2. Random coefficient regression models

In this study we use the approach of two-level random coefficient regression (Hsiao 2003, Fitzmaurice et al. 2004) to capture the dynamics of price levels and price dispersion. In this approach, repeated measurements over time (i.e., price levels and price dispersion) are modeled using a regression equation on time, and the time correlation is captured via an autoregressive error structure. Further, to take cross-sectional heterogeneity into account, the regression coefficients are allowed to vary from one title to another, and are assumed to be drawn randomly from the entire population of titles. The title-specific random coefficients thus reflect the heterogeneity in the population of titles.

Specifically, for each title, we consider the means and standard deviations of the log-prices calculated across individual retailers within dotcoms and within multichannel retailers. Let \( y_{ijklm} \) denote the means or standard deviations of the \( k \)th title \((k = 1, \ldots, 61)\) measured in the \( m \)th time period (month) \((m = 1, \ldots, 12)\). The index
i is an indicator for the retailer type with 1 for multichannel retailers and 2 for dotcoms; the index j is an indicator for the DVD type with 1 for random titles and 2 for popular titles.

In the conventional linear trend regression, log-price = \( \alpha + \beta \times \text{time} + \epsilon \), the natural log of price varies systematically with the variable time, and the variation that cannot be accounted for by time is captured with a random variable \( \epsilon \) that is assumed to be white noise. The two coefficients \( \alpha \) and \( \beta \) are the population intercept and slope parameter, respectively, and are assumed to be constant across all individual titles.

It is clear from Fig. 1 that the assumption of constant intercept and slope parameters is problematic since the titles are obviously not homogeneous. We incorporate a two-level random coefficient regression model, where the lower level characterizes the dynamic behavior of the means or standard deviations of the log-prices over time via regressing \( y_{ijklm} \) on a linear model of time \( t_{mk} \):

\[
y_{ijklm} = \alpha_{ij} + \beta_{ij}t_{mk} + \epsilon_{ijklm}, \quad i, j = 1, 2; \quad k = 1, \ldots, 61; \quad m = 1, \ldots, 12.
\]  

(1)

The error terms \( \epsilon_{ijklm} \) are assumed to be normally distributed, \( N(0, \sigma^2) \). To capture the variability caused by time, we assume an autoregressive of order 1 covariance structure for the error terms \( \epsilon_{ijklm} \) with a parameter \( \rho \) that characterizes the correlation of the prices of a title in two adjacent time periods at a retailer.

In contrast to the conventional linear regression, the coefficients in Eq. (1), \( \alpha_{ij} \) and \( \beta_{ij} \), are title specific, i.e., they are related to a specific title in a specific category of marketing characteristics via the indicators \( i \) and \( j \). The intercept \( \alpha_{ij} \) is the initial value of the mean or standard deviation of the log-prices of the first title at the beginning of the investigation. The slope parameter, \( \beta_{ij} = dy_{ijklm} / dt_{mk} \), represents the rate of change over time.

We assume that both \( \alpha_{ij} \) and \( \beta_{ij} \) are affected by the retailer type and the DVD type. Hence, at the higher-level of the model for the \( k \)th title we postulated an analysis-of-covariance model, relating the title-specific regression coefficients, \( \alpha_{ij} \) and \( \beta_{ij} \), to the marketing characteristics that may affect the pricing parameters. Specifically, for the intercept we assume

\[
\alpha_{ij} = \theta_{11}MC_i + \theta_{12}DC_i + \theta_{13}RT_j + \theta_{14}PT_j + \theta_{15}MC_i \times RT_j + \theta_{16}MC_i \times PT_j + \theta_{17}DC_i \times PT_j + u_{ik},
\]  

(2)

and for the slope parameter we assume

\[
\beta_{ij} = \theta_{21}MC_i + \theta_{22}DC_i + \theta_{23}RT_j + \theta_{24}PT_j + \theta_{25}MC_i \times RT_j + \theta_{26}MC_i \times PT_j + \theta_{27}DC_i \times PT_j + \theta_{28}DC_i \times PT_j + v_{ik},\]

(3)

where \( u_k \) (or \( v_k \)) is the random error component associated with each title \( k \); it is assumed to be normally distributed with zero mean and variance \( \sigma^2_u \) (or \( \sigma^2_v \)). The intra-class correlation is thus characterized by the two random effects \( u_k \) and \( v_k \). The parameters \( \theta_{11}, \ldots, \theta_{28} \) are the coefficients to be estimated. MC (DC) is an indicator of the retailer type, defined to be one if a title is priced at a multichannel (dotcom) retailer and zero otherwise. RT (PT) is an indicator of the DVD type, defined to be one if a DVD is random (popular) and zero otherwise.

The equations at the higher-level of the model, (2) and (3), simply say that the title-specific coefficients \( \alpha_{ij} \) and \( \beta_{ij} \) are related to the retailer type, the DVD type, and their interactions via an analysis-of-variance model. Consequently, each title has its own dynamics characterized by the title-specific coefficients \( \alpha_{ij} \) and \( \beta_{ij} \).

4.3 Properties of the random coefficient regression model

The model of two-level random coefficient regression specified in Eqs. (1)–(3) can accommodate the important features of the data collected in this study. First, on the basis of the observations from the preliminary analysis (e.g., Fig. 1), we have assumed a linear model at the lower level, Eq. (1), to capture the trends of price levels and price dispersion. The error terms \( \epsilon_{ijklm} \) in Eq. (1) are assumed to have an autoregressive of order 1 covariance structure with parameter \( \rho \). Consequently, for any title, say the kth, the prices posted at two adjacent time periods \( m \) and \( m + 1 \) are correlated, i.e.,

\[
\text{cov}(y_{ijklm}, y_{ijklm+1}) = \text{cov}(\epsilon_{ijklm}, \epsilon_{ijklm+1}) = \rho \sigma^2 \neq 0.
\]

Hence the time correlation can be captured in the model.

Secondly, Fig. 1 shows substantial variability among titles so the intercept and slope of the linear trend are assumed to vary from one title to another. Hence, we have specified a random effect model at the higher-level, Eqs. (2) and (3). Now consider a title, say the kth, posted in a time period \( m \) at two different retailer types, 1 and 2. We have

\[
\text{cov}(y_{ijklm}, y_{ijklm+2}) = \sigma^2_u + \sigma^2_v + \rho^2 \sigma^2 \neq 0.
\]

Hence, the price of a title posted at one retailer is correlated with the prices of the same title posted at other retailers so that intra-class correlation is captured in the model.

Statistically, the specified structure of the covariance in model (1)–(3) will be taken into account when the parameters are estimated using the two-level random coefficient regression model. Compared with a model where the structure of the covariance is ignored, the estimates obtained are more efficient and the potential aggregation biases are avoided. In a nutshell, the random coefficient regression approach provides a powerful means to model the variability at each level and allows us to simultaneously take into account of the spatial and time correlations. It is the standard approach to the analysis of repeated measurements in statistics and econometrics (Hsiao 2003, Fitzmaurice et al. 2004).

5. Empirical results

In this section, we will discuss the empirical results obtained using the econometric model (1)–(3) for which there are several important assumptions: (a) time correlation and intra-class correlation; (b) linear trend in Eq. (1); (c) normal distributions in Eqs. (1)–(3). We note that the time correlation and intra-class correlation are evident due to the nature of the data, and the linear trend assumption is justified from Figs. 1 and 2. In addition, because of the large sample size, the hypothesis testings are not sensitive to the normality assumption due to the asymptotic normality theory.

5.1 Analysis of price levels

On the basis of the two-level random coefficient model (1)–(3), a statistical analysis was carried out for the means of log-prices across individual retailers within each retailer type via the restricted maximum likelihood method. After having removed the fixed effects that were not statistically significant, we obtained the following model given by equations (1), (4), and (5):

\[
\alpha_{ij} = \theta_{11}MC_i + \theta_{12}DC_i + \theta_{13}RT_j + \theta_{14}PT_j + \theta_{15}MC_i \times RT_j + \theta_{16}MC_i \times PT_j + u_{ik},
\]  

(4)

\[
\beta_{ij} = \theta_{21}MC_i + \theta_{22}DC_i + v_{ik}.
\]  

(5)

It is worth noting that the dummy variables RT and PT for the DVD types do not appear in Eq. (5) because they were not significant and thus had been removed from the model. From the perspective of marketing research, this means that for both dotcoms and multichannel retailers, the rate of change in price levels is not related...
Hence, initially there was a difference in price levels between the upper right and lower right.

The estimated individual price levels, $x_{ijk} + \beta_{ijk}t_m$ in Eq. (1), for all titles are displayed in Fig. 4. For instance, the estimated price levels versus months for all 32 popular titles at dotcoms are displayed in Fig. 4 (upper right). It is assumed in the model that different titles have different initial prices (intercepts) and different rates of change in price levels (slopes), thus they all have different profiles. It can be seen that overall the price levels in Fig. 4 (upper left) for popular titles at multichannel retailers are lower than the price levels in Fig. 4 (upper left) for popular titles at dotcoms. Further, the individual straight lines in Fig. 4 are not parallel because the model is able to accommodate the fact that different pricing strategies were adopted for different titles.

The estimates of the parameters at the higher-level of the model, Eqs. (4) and (5), are displayed in Table 2. All of these estimates are statistically significant at the 1% level, except $\theta_{15}$ which is significant at the 10% level.

Note that parameters $\theta_{11}$ and $\theta_{12}$ are the effects of multichannel retailers and dotcoms, respectively, on the intercept parameter $x_{ijk}$. Thus, a test of the following statistical null hypothesis

$$\text{H}_1 : \theta_{11} - \theta_{12} = 0$$

shows on average whether or not dotcoms and multichannel retailers charged prices at the same level at the beginning of the study period. The resulting $p$-value is less than 0.0001, thus $\text{H}_1$ is rejected. Hence, initially there was a difference in price levels between the two retailer types.

On the other hand, parameters $\theta_{13}$ and $\theta_{14}$ are the effects of multichannel retailers and dotcoms, respectively, on the slope parameter $\beta_{ijk} = \frac{dy_{ijk}}{dx_m}$. Thus, a test of the following statistical null hypothesis

$$\text{H}_2 : \theta_{21} = \theta_{22}$$

shows on average whether or not dotcoms and multichannel retailers changed prices at the same rate. The resulting $p$-value is 0.2350, thus $\text{H}_2$ is not rejected. Statistically this suggests that multichannel retailers and dotcoms reduced their prices at the same rate. Hence, contrary to Hypothesis 3, it appears that the average difference in the log-prices of DVD titles between dotcoms and multichannel retailers was maintained at a constant level over time in the US online market.

Furthermore, by combining the results of statistical hypothesis testing for $\text{H}_1$ and $\text{H}_2$, we conclude that the initial difference in price levels between dotcoms and multichannel retailers was maintained, thus supporting Hypothesis 1.

Now we consider Hypothesis 2a. We note that Eq. (5) for the slope parameter $\beta_{ijk}$ does not include the dummy variables $PT$ and $RT$, thus $\beta_{ijk}$ does not depend on the DVD type. Hence, to test Hypothesis 2a we need to consider only the intercept $x_{ijk}$ in Eq. (4). For dotcoms, the average intercept for random titles is $\theta_{12} + \theta_{13} + \theta_{16}$ and the average intercept for popular titles is $\theta_{12} + \theta_{14} + \theta_{16}$. Hence, a test of the statistical null hypothesis

$$\text{H}_3 : \theta_{12} + \theta_{13} + \theta_{16} - (\theta_{12} + \theta_{14} + \theta_{16}) = 0$$

shows on average whether or not dotcoms set popular and non-popular titles at the same price level, and thus further shows whether or not the difference in price levels was maintained across

### Table 2

| Effect | Estimate | Standard error | $t$ | $Pr > |t|$ |
|--------|----------|----------------|-----|---------|
| $\theta_{11}$ | 3.0746 | 0.04454 | 69.04 | <0.0001 |
| $\theta_{12}$ | 2.9290 | 0.04454 | 65.77 | <0.0001 |
| $\theta_{13}$ | -0.2775 | 0.06289 | -4.41 | <0.0001 |
| $\theta_{14}$ | 0.0000 | - | - | - |
| $\theta_{15}$ | -0.0539 | 0.03154 | -1.71 | 0.0879 |
| $\theta_{16}$ | 0.0000 | - | - | - |
| $\theta_{17}$ | 0.0000 | - | - | - |
| $\theta_{18}$ | 0.0000 | - | - | - |
| $\theta_{21}$ | -0.0140 | 0.002459 | -5.51 | <0.0001 |
| $\theta_{22}$ | -0.0167 | 0.002465 | -6.79 | <0.0001 |

* Statistically the parameters $\theta_{14}$ and $\theta_{16}$ are not simultaneously estimable. One of them (say $\theta_{14}$ here) has to be set to zero so that essentially it is the differences $\theta_{13} - \theta_{14}$ that is estimated. For a similar reason only $\theta_{15}$ is estimated but $\theta_{16}$, $\theta_{17}$, and $\theta_{18}$ are set to zero.

### Fig. 4

The estimated price level versus month at multichannel retailers for popular titles (upper left) and for random titles (lower left), and at dotcoms for popular titles (upper right) and for random titles (lower right).
the whole time period of the study. The resulting p-value is less than 0.0001, thus H$_3$ is rejected. Statistically this suggests that dotcoms priced popular and non-popular titles at different levels at the beginning of the study. Now given the fact that $\beta_{ijk}$ does not depend on the DVD type, such a difference was maintained throughout the whole study period. Hence, Hypothesis 2a is supported.

Similarly, the statistical null hypothesis H$_4$ in Table 3 shows on average whether or not multichannel retailers set popular and non-popular titles at the same price level. H$_4$ is rejected with a p-value less than 0.0001 and thus Hypothesis 2b is supported.

Table 3 displays a summary of the hypothesis testing for price levels.

Finally, we consider the random effects. All the estimates of the covariance parameters are significant at the 1% level. In particular, the estimated standard deviation $\sigma_\epsilon$ of $u_k$ is 0.2235, which is much larger than the estimated standard deviation $\sigma_\epsilon$ of the residuals $\epsilon_{ijkm}$ of 0.1213. So the variability of the mean log-prices from one title to another is large. This suggests that different titles were priced differently at the beginning of the study period. Hence, enforcing a constant intercept in regression equation (1) is problematic. In addition, the estimate of the time correlation coefficient $\rho$ is 0.8407, indicating that the price of a title is highly correlated to the price of the same title in the previous month. Hence, ignoring the time correlation in the longitudinal study is not acceptable.

Turning to the random effects on the slope parameters, we note that the estimated standard deviation $\sigma_\epsilon$ of $h_k$ is 0.0167. Hence, the variability of the slope parameters from one title to another is quite high compared to the mean slopes, i.e., –0.0140 for multichannel retailers and –0.0167 for dotcoms. Statistically, this suggests that there were many retailers at which the prices were stable rather than declining, although on average there were downward trends for both multichannel retailers and dotcoms.

To sum up, we have tested the statistical hypotheses H$_1$ to H$_4$ in this subsection. On the basis of that, the research issues, formulated as Hypotheses 1-3 in Section 3, are addressed. In a nutshell, Hypothesis 1 is supported and it shows that on average the price level of DVD titles at multichannel retailers was higher than that at dotcoms throughout the whole study period. In addition, both Hypotheses 2a and 2b are supported, thus on average the price level of popular DVD titles was higher than that of random titles at any time-point at both dotcoms and multichannel retailers. However, Hypothesis 3 is not supported. In fact, it was suggested from the analysis that the average difference in the log-prices of DVD titles between dotcoms and multichannel retailers was maintained at a constant level.

5.2. Analysis of price dispersion

Now we consider price dispersion in terms of the standard deviations of the log-prices across individual retailers within each retailer type. The statistical analysis was carried out using the two-level random coefficient regression model specified by Eqs. (1)–(3). After having removed the fixed effects that were not statistically significant, we obtained the final model given by (1), (6), and (7):

$$x_{ijk} = \theta_{11}MC_i + \theta_{12}DC_i + \theta_{13}RT_j + \theta_{14}PT_j + \theta_{15}MC_i \times RT_j + \theta_{16}DC_i \times RT_j + \theta_{17}MC_i \times PT_j + \theta_{18}DC_i \times PT_j + u_{it} = \beta_{ijk}$$

Here $\beta_{ijk}$ = $\theta_{11}MC_i + \theta_{12}DC_i + \theta_{13}RT_j + \theta_{14}PT_j + \theta_{15}MC_i \times RT_j + \theta_{16}DC_i \times RT_j + \theta_{17}MC_i \times PT_j + \theta_{18}DC_i \times PT_j + u_{it}$. Hence, a test of the statistical null hypothesis

$$H_5 : \theta_{21} = 0$$

shows on average whether or not the dispersion level of the log-prices within dotcoms remains stable. Table 4 shows that the estimate of $\theta_{21}$ is positive and significant with p-value 0.0012, thus H$_5$ is rejected. Hence, individual dotcoms tended to price their products differently over time, thus supporting Hypothesis 4a.

This result might be the consequence of several years’ competition where dotcoms have become more and more heterogeneous in terms of their brand names so that some of the well-established dotcoms such as Amazon have increased their prices. For instance, from Table 1 it can be seen that Amazon is the highest-priced dotcom with an average DVD price of US$16.61. This is even higher than the lowest-priced multichannel retailer, Tower, with an average DVD price of US$15.56.

On the other hand, a test of the statistical null hypothesis

$$H_6 : \theta_{22} = 0$$

shows on average whether or not the dispersion level of the log-prices within multichannel retailers remained stable. Table 4 shows that the estimate of $\theta_{22}$ is negative and not significant (with a p-value of 0.1657), indicating that multichannel retailers tended to maintain the price dispersion at the same level, thus supporting Hypothesis 4b. This shows that the difference in price levels among multichannel retailers tends to be consistent.

For dotcoms, the average intercept for random titles is $\theta_{12} + \theta_{13} + \theta_{14}$ and the average intercept for popular titles is $\theta_{12} + \theta_{14} + \theta_{18}$. Hence, a test of the following statistical null hypothesis

$$H_7 : (\theta_{12} + \theta_{13} + \theta_{14}) = (\theta_{12} + \theta_{14} + \theta_{18}) = 0$$

shows on average whether or not the price dispersion among dotcoms for random titles is the same as that for popular titles at the beginning of the study period. The resulting p-value is 0.0042; thus H$_7$ is rejected. Hence, initially the price dispersion differed among dotcoms for the two DVD types. Because Eq. (7) does not include the dummy variables RT and PT, the rate of change in Eq. (7), i.e., $\beta_{ijk}$,
does not depend on the DVD type. Hence, the initial difference was kept at the same level over time and Hypothesis 5a is not supported.

Likewise, a test of the following statistical null hypothesis

\[ H_8 : (\theta_{11} + \theta_{13} + \theta_{15}) - (\theta_{11} + \theta_{14} + \theta_{17}) = 0 \]

shows on average whether or not the price dispersion among multichannel retailers for random titles was the same as that for popular titles at the beginning of the study period. The resulting p-value is 0.9681, thus H8 is not rejected. Hence, it seems that initially the price dispersion was statistically identical among multichannel retailers for the two DVD types. Because Eq. (7) does not include the dummy variables RT and PT, the rate of change in Eq. (7), \( \beta_{ijk} \), does not depend on the DVD type. Hence, for multichannel retailers, the spread measured by standard deviation of log-prices for popular titles was maintained at the same level as that for random titles at any time-point. So Hypothesis 5b is not supported.

Table 5 displays a summary of the hypothesis testing for price dispersion.

<table>
<thead>
<tr>
<th>Null hypothesis</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_5 )</td>
<td>0.0142</td>
</tr>
<tr>
<td>( H_6 )</td>
<td>0.1657</td>
</tr>
<tr>
<td>( H_7 )</td>
<td>0.0042</td>
</tr>
<tr>
<td>( H_8 )</td>
<td>0.9681</td>
</tr>
</tbody>
</table>

Now we have tested the statistical hypotheses H5 to H8 in this subsection. On the basis of the analyses, the research issues, formulated as Hypotheses 4 and 5 in Section 3, have been addressed. Both Hypotheses 4a and 4b are supported. Hence, the spread among dotcoms increased but the spread among multichannel retailers decreased or was stable over time. So the competition among the dotcoms and the competition among the multichannel retailers show different patterns. We also note that both Hypotheses 5a and 5b are not supported. In fact, the data provide strong evidence that the spread for popular titles differed from that for random titles at dotcoms. In addition, it suggests that the spread of popular titles was similar to that for random titles at multichannel retailers.

6. Discussion and conclusions

In this study, we have investigated online pricing dynamics in the US online DVD market. Using the approach of two-level random coefficient regression models, we have tested hypotheses to compare the pricing behavior of dotcoms and multichannel retailers.

The data collected in this study raised several challenging modeling issues. At first glance it seems that the well-known VARX model would be an obvious choice. Because of the nature of the data, however, the research carried out at the disaggregate level would cause aggregation bias in the VARX model. In addition, the large number of commodities included in the study would have the problem of over-parameterization. To address these issues, we used an alternative approach to analyze the comprehensive dataset of online DVD prices, the two-level random coefficient regression models.

The results of the empirical analysis have provided support for our hypothesis that multichannel retailers charge higher prices than dotcoms. This is true not only at a given time-point as shown in earlier static analyses (e.g., Tang and Xing 2001; Pan et al. 2002, Ancarani and Shankar 2004) but also is the case over the whole time period of this study. This result is consistent with the findings in Xing et al. (2006) but the conclusion was drawn on a more solid basis because the more advanced econometric model was used in the analysis so that the estimates were more efficient and potential biases were avoided.

Our results on the trend of the difference in price levels extend the existing research on the online DVD market. Based on the price data collected from 2000 and 2001, Xing et al. (2006) show that (a) price levels went up over time at both dotcoms and multichannel retailers; and (b) the prices at dotcoms increased faster than those at multichannel retailers. Xing et al. (2006) suggest that the difference in price levels tends to converge in the long term, and further predict that the two retailer types will have similar pricing behavior.

The results of this study, based on the more recent price data, suggest some interesting findings which have reflected the evolution from the early stage to a more mature US DVD market. First, contrary to the dynamics revealed in Xing et al. (2006) where overall prices increased over time, this study has exhibited declining trends in price levels. This is in line with our observation that demand is high when titles are launched and will diminish when their rankings drop and new titles become more popular. Consequently, prices reduce gradually over time. We note that the analysis in Xing et al. (2006) was based on the data collected during an earlier stage of Internet retailing. It is likely that their results reflected the fact that during the initial years of Internet retailing, online retailers charged lower prices at first to attract customers, and then gradually raised prices. When the DVD market became more mature, however, such a high price level was not sustainable due to competition. This is likely the reason that we observe a pattern of declining DVD prices in this study.

Next, in contrast to the prediction by Xing et al. (2006) that the two retailer types will have similar pricing behavior, this study shows that the difference in price levels is maintained over time. This seems logical for a mature market since there are strong theoretical reasons for different pricing behavior between the two retailer types: (a) multichannel retailers can combine the benefits of online shopping with support from offline stores; (b) multichannel retailers can translate market power from offline to online mode; and (c) dotcoms have lower operating costs. With all these fundamental retailing characteristics being maintained, it is unlikely that the difference in price levels will decrease in the long term.

From the stability in pricing behavior revealed in this study, there is a sign of maturity in the current US online DVD market. The results obtained in this study thus should not be generalized to immature online markets since it is unlikely that a market in a transition period will exhibit such stable pricing behavior.

Our results also suggest that on average the price levels of popular titles at both dotcoms and multichannel retailers are higher than those of random titles. This is consistent with our observation that the demand for popular titles is normally higher.

Our results are consistent with Lee and Gosain (2002) in that the degree of price dispersion depends on the product type. Contrary to Tang and Xing (2001) and Xing et al. (2006), however, our results show that for dotcoms the standard deviation was lower for popular titles, whereas for multichannel retailers there was no statistical difference in the price dispersion of the two product types.

Our results show that price dispersion among dotcoms increases over time. Contrary to the prediction in Xing et al. (2006) that price dispersion levels among dotcoms will converge, our results show that among dotcoms the price dispersion continues the increasing trend but among multichannel retailers the price dispersion seems to have become stable after several years of competition. Hence, there is no sign that the price dispersion levels for the two retailer types are similar.

We conjecture that the different patterns in price dispersion between dotcoms and multichannel retailers are fundamental
because of the following reasons. First, some dotcoms have established their reputation and have a considerable market share, whereas new dotcoms are being launched every year because online markets have easier entry than offline markets. Consequently, dotcoms tend to be more heterogeneous than multichannel retailers. Secondly, dotcoms tend to have head-to-head price competition, whereas multichannel retailers can differentiate themselves from dotcoms on nonprice dimensions.

The major findings of this paper include: (a) multichannel retailers charge higher prices than dotcoms throughout the whole time period of this study; (b) contrary to the prediction in the previous studies that the two retailer types will have similar pricing behavior, this study shows that the difference in price levels is maintained over time; and (c) contrary to their prediction in the previous studies that price dispersion levels among dotcoms and multichannel retailers will converge, our results show that the price dispersion among dotcoms continues the increasing trend but the price dispersion among multichannel retailers seems to have become stable. All these findings show that there exist some fundamental differences in pricing behavior between dotcoms and multichannel retailers.

Managers are interested in a better understanding of the pricing behavior in different channels. Our results have some important managerial implications. First, it is suggested that the two retailer types compete with each other in quite different ways. Multichannel retailers tend to differentiate themselves from dotcoms on nonprice dimensions by combining online shopping with support from their offline stores. They may also translate their market power and brand names from offline to online modes. They can charge higher prices without losing all their customers and this avoids severe price competition among them. It is thus essential for multichannel retailers to fully utilize their offline stores in retailing and to ensure that their brand names are successfully translated into the corresponding online markets.

Dotcoms tend to differentiate themselves from multichannel retailers through pricing, and competition among dotcoms also tends to be on the price dimension. Since online markets have easier entry than offline markets, direct price competition among dotcoms tends to be more intense. It is thus vital for dotcoms to use effective pricing strategies such as applying random pricing, adopting different shipping costs, and offering product bundles. To differentiate themselves from each other and to avoid direct price competition, dotcoms should also provide better service and communicate with customers effectively.

The findings of the present study must be considered in the light of the limitations in the data collection process. Because the sampling frame (i.e. the full list of the DVD titles in the market) was not available during the study period, probability-based sampling methods (such as simple random sampling and stratified sampling) could not be used in this study. Instead, in this study we followed the method used in the existing studies (e.g. Brynjolfsson and Smith 2000, Tang and Xing 2001, Ancarani and Shankar 2004, Xing et al. 2006) when selecting retailers and DVD titles. Although the market share of these online retailers was substantial, the data was not randomly selected and potentially it may lead to bias in the data collection process.

Acknowledgements

The authors thank the area editor and the two reviewers for their constructive comments on the earlier versions of this paper. The authors also thank Loretta Law and Lisa Wan for data collection assistance. The research of Fang-Fang Tang described in this paper was supported by a grant from China Natural Science Foundation (Project No. 70972014) for which we are grateful.

References