Diffusion of a New Product under Network Effects: The Case of U.S. DVD Player Market

Firat Inceoglu and Minsoo Park*

Boston University

December 5, 2003

Abstract

Indirect network effects arise when the benefit of using a product increases in the use of a complementary set of goods. In this paper we formulate a nested logit demand model that captures the inter-dependence between hardware demand and software supply in the DVD industry. The identification of the network effect comes from the difference in software availability across two different formats: VHS and DVD. We find strong evidence for network effects: a 1% increase the number of DVD titles raises the demand for DVD players by 0.68%. Simultaneously, a 1% increase in the hardware installed base (DVD player sales) leads to a 0.15% increase in software availability (number of DVD titles). Our simulations show that there is a strong incentive for horizontal integration; hardware firms can internalize the network externality through horizontal integration.

*Department of Economics, Boston University. E-mail: inceoglu@bu.edu, msoopark@bu.edu. We would like to thank to our advisor Marc Rysman and dissertation committe members Victor Aguirregabiria and Kevin Lang for setting us in the right direction. This project would not have been finished without the support of Pierre Perron. We also received generous help from Martino De Stefano and Masashi Saito. This paper has benefited from the comments of the participants at the Graduate Student Workshop and Micro-lunch seminar series at Boston University. We are also indebted to Jeff Jordan at NPD Tech, and VSDA for their help in providing us with the data. Department of Economics Graduate Student Academic Research Grant is gratefully acknowledged.

1 Introduction

Network effects arise when the utility derived from the consumption of a good depends on the number of people using it. Most common examples are phone lines and fax machines. A phone line is of no use unless others are also connected to the network. Indirect network effects arise when the benefit of using a product is increasing in the use of a complementary set of goods. Typical examples of products exhibiting indirect network effects are computers (complementary with software) and CD-players (complementary with CD's).

In the last few years, the Industrial Organization literature has increasingly explored indirect network effects. The theory of network effects has wide applicability. Indeed, it has fundamental importance for competition policy, regulation, business strategy, intellectual property, and technical change in a wide range of industries, from hi-tech such as IT industries (Gandal 1994) to traditional ones such as the yellow page industry (Rysman 2003).[1] Developments in these industries cannot be fully understood without an understanding of network effects. Equally important is the fact that network effects, under very general conditions, will give rise to network externalities (Church, Gandal, and Krause 2002). The externality occurs when the private benefit of the marginal hardware purchaser is less than the social benefit. The marginal hardware purchaser does not internalize the welfare improving response of the software industry, where an increase in software variety will benefit the inframarginal users.

Indirect network externalities create a two-way contingency between the hardware demand and the software supply. In this paper we develop a model that captures the inter-dependence between hardware demand and software supply in a differentiated goods market. Specifically, we apply this model to the U.S. DVD player market using price and sales data on DVD players and data on DVD titles. We assume that the utility a consumer receives from owning a DVD player increases in the number of movie titles available in the DVD format. In return, an increase in the number of DVD players leads to a greater demand for software, thus completing the cycle. We use data on hardware sales and software availability to measure the extent and the size of the network effects in the DVD player market. The dataset used in this study is unique; we have almost complete coverage of the

[1] There exist indirect network effects in yellow pages because consumers value a directory more when there are more listings (information), while businesses place more ads if there are many users of a directory.

hardware market with detailed information on each separate product. Additionally, we observe all movies released on DVD, along with their features. This extra information is essential for getting an unbiased estimate of the network effect.

We focus on the DVD market because of the relatively recent introduction (April 1997) of the product, its emergence as the main home-video system in a remarkably short period of time, and the oligopolistic structure of the market. Since hardware prices decline and software availability increases steadily over time, it is difficult to identify the effect of increased software availability on hardware adoption using time series data alone. The identification of the network effect comes from the differences in the availability of new titles on two different formats: VHS and DVD. Until very recently the number of new titles released was greater for the VHS format than for DVD.[2] Differences in software availability and the resulting differences in adoption rates allow us to estimate the network effect accurately.

Our study is motivated by the recent growth of the DVD player market. Why was the DVD so successful, whereas similar products, such as laser discs, failed to establish a significant consumer base? An indisputable factor is the rapid drop of prices. There are several forces that may have driven down prices. Entry into the DVD player market being one of them. Hardware manufacturers might have strategically lowered prices to exploit the network effect. Or maybe the fall in prices simply reflects a reduction in the costs of production. Another factor behind the success of DVD is the rapid increase in the DVD titles variety. We are especially interested in measuring the strength of network effects in the DVD market. The ability to measure network effects is crucial to firms competing in these industries. It is also important from a social welfare point of view. Under certain conditions, the good or service exhibiting network effects will be underprovided or underconsumed, leading to market inefficiency. In the context of inter-dependence, vertical integration of hardware and software industries, or horizontal integration of hardware firms could be socially desirable as the integrated body will be able to internalize the network externality. In order to be able to analyze all these issues we construct a discrete choice model of a differentiated products market with an explicit structure of supply and demand.

Our results point to the existence of strong network effects in the DVD market. A 1% increase

[2] By 'titles' we mean movies produced for theatrical and/or home-video screening. Today, all new movies that appear in the video rental market are released in both formats. But since VHS is a much older product, the accumulated number of VHS titles surpasses that of DVD titles by far.

in the number of DVD titles raises the demand for DVD players by 0.68%. On the other hand, a 1% increase in hardware installed base leads to a 0.18% increase in the number of titles released on DVD. Using our parameter estimates we run simulations that confirm our a priori belief that horizontal integration will enable the hardware industry to internalize the network externality.

The endogeneity of software availability complicates the estimation of hardware demand. Any unobserved factor making DVD a more popular product will lead to greater hardware sales and increased software availability. To correct this endogeneity we use an instrumental-variables approach We need instruments that are uncorrelated with hardware demand but correlated with software supply . We believe that the characteristics of titles released at the box office and on VHS are good instruments.

The next section reviews the related literature. In section 3, we briefly describe the evolution of DVD as a new home video format. Section 4 outlines the estimation methodology and section 5 describes the data. In section 6 we provide the estimation results, and then experiment with a counterfactual exercise to show the value of consolidation to hardware firms. The last section concludes.

2 Related Literature

Starting in the 1980's a small group of economists began to show interest in markets with network externalities. Initially, the focus of these mostly theoretical studies was on direct network externalities. Katz and Shapiro (1985) develop an oligopoly model to analyze a market with network effects. Farrell and Saloner (1985) focus on the role of firms as these decide sequentially whether to adopt the new technology (network). In the early nineties Chou and Shy (1990) and Church and Gandal (1992) extended the analysis to markets of complementary products.

Empirical studies on the subject started to appear only very recently. Gandal, Kende, and Rob (2000) is the first to investigate empirically the complementaries between hardware and software products. They use aggregate level data to estimate a dynamic model of the CD player industry. Ohashi (2003) estimates network effects in the U.S. VCR market, but he lacks information on software availability and approximates it with lagged installed base.

The study closest to our paper is Nair, Chintagunta, and Dube (2003). They analyze the market for Personal Digital Assistants (PDA) and show the existence of strong network effects. Although the methodology of their paper is similar to ours, they do not control for the endogenity of software availability in their estimation. This casts some doubt on the validity of their estimates.

This paper follows the track of the literature and improves on it in a couple of ways. First, we have an extensive, firm-level data set that allows us to take into account the differentiated products nature of the market. The importance of this becomes apparent when one formulates hypotheses about cooperative supply side behavior or about welfare effects of integration.

Additionally we have a more elaborate description of the software side with full information on software availability plus characteristics of software. This allows us to identify the network effect in the DVD market without having to make strong assumptions

3 The U.S. DVD Industry[3]

3.1 Hardware

In April 1997, the Digital Versatile Disc (DVD) was introduced by a consortium of hardware makers and motion picture studios as an affordable, yet markedly superior, replacement for videotapes. At the time of the introduction of the first DVD player, there were only 40 movie titles available as studios were reluctant to enter the market. Early sales of DVD players exceeded expectations and, after a brief standards war that resulted in the clear victory of the DVD, sales reached unprecedented levels. As a result, the DVD player became the fastest growing electronics product. In 2001, monthly shipments of DVD players surpassed those of VCRs. Today, accumulated sales of DVD players in the U.S. have reached 42 million units.

The radical expansion of the format was partly stimulated by a continuous decline in hardware prices. The average price of a DVD player fell from $491 in 1997 to $118 today. Strikingly, during this period the average quality steadily increased as hardware manufacturers added new features to their products. The fall in prices is, in part, due to the entry of a significant number of low-cost manufacturers such as Apex Digital Inc. which, established by the end of 1999, possesses the largest

[3]The chronology of the DVD industry, especially the software side history, is taken from Taylor(2000). Also see the webpage version (www.dvddemystified.com/dvdfaq.html) for comprehensive information on the DVD format.

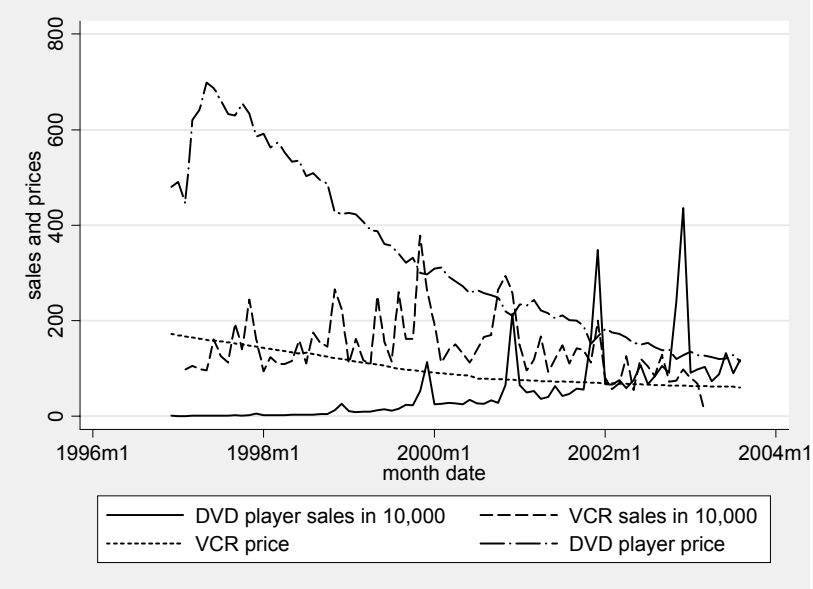

Figure 1: Monthly sales and prices of DVD players and VCRs

market share in units sold today. The number of manufacturers increased from 16 in 1997 to 51 in 2003. We present a summary of the evolution of the DVD player market in table A1, in appendix A.

The DVD market has not grown without a threat of other competing standards. In September 1997, Circuit City, one of the biggest consumer electronics chain store in the U.S., announced the introduction of a new format (DVIX) with the support of a few movie studios. DVIX is a pay-per-view version of DVD, and DVIX players can play normal "open" DVD titles, but DVD players cannot play DVIX discs. Rumors and the late announcement of the new format confused possible consumers for a while until it quickly became clear that DVIX would not replace the DVD format. Dranove and Gandal (2003) report that the installed base of DVD players was approximately 1.9 million units through mid-1999 while that of DIVX at the same time period was only 165,000. Several explanations are suggested for the failure of the DIVX format. One unambiguous reason was the lack of support from the software providers. According to Dranove and Gandal (2003), as of May 1999, there were 3,317 DVD titles and only about 100 DVIX titles available.[4]

[4]For the same period, the number of DVD titles in our dataset is 1,917. See Dranove and Gandal (2003) for more details about the DVD-DVIX standard war.

3.2 Software

The first feature films on DVD appeared in Japan in December 1996 (*The Assassin, Blade Runner, Eraser*, and *The Fugitive* from Warner Home Video). In the U.S. Warner Brothers US launched DVD titles on March 24, 1997, but the launch was limited to seven cities. Nevertheless, almost 19,000 discs were purchased in the first two weeks of the U.S. launch, a number much greater than that expected by the studio. By December 1997, over 1 million DVD discs were shipped while the number of titles climbed up to 530. Today, all movies that appear in theaters are released on DVD as well as on VHS. By the end of 2002 there were about 23,000 titles available in the U.S. Due to the late response from the video rental market – it was September, 1999 when Blockbuster began DVD rentals in 500 stores – growth of DVD rentals was slower compared to the growth of sell-through sales. But also the rental market transited quickly to the DVD and DVD rentals outpaced videocassette rentals for the first time in June 2003.[5]

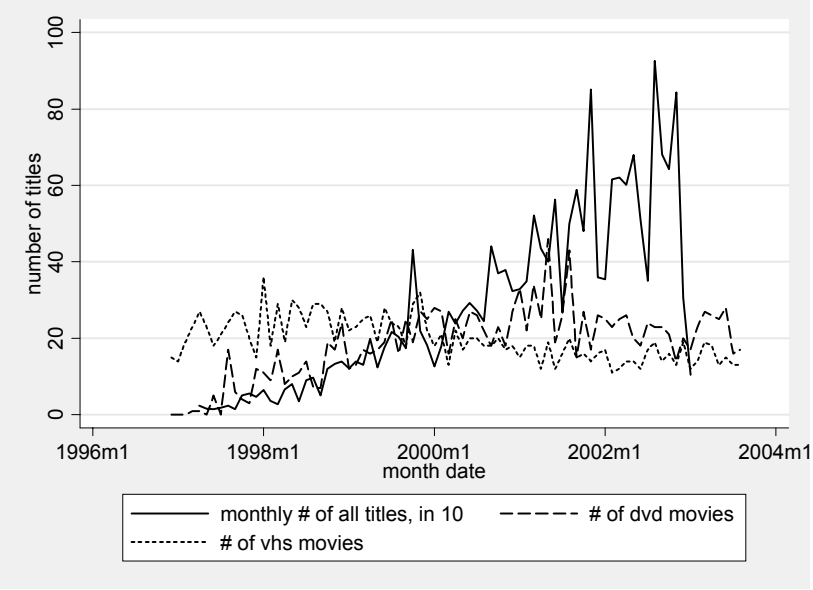

Figure 2: Number of monthly titles releases

Among software providers, the initial supporters of the DVD format were Warner, Columbia/TriStar, Polygram (music recording), and MGM. After observing the success of DVD, other movie studios began to enter the DVD market. Universal started to release their titles on DVD starting from July

[5] *The Washington Times*, June 21, 2003

1997, followed by Disney (Buena Vista) in Fall 1997. Paramount, Fox, and Dreamworks initially supported DVIX only, but in summer 1998 each announced its support for open DVD. The latest major studio to join the market was Dreamworks SKG, which entered in September, 1998.

One interesting fact is that, among the early backers of DVD, Columbia and Polygram were respectively owned by Sony and Phillips, two producers of hardware.[6] Universal Studios was previously owned by Matsushita but then sold to Seagram in 1995.

4 Data

4.1 Hardware Data

The structural estimation of the model necessitates the use of product level sales and prices, and information on the characteristics of DVD players. Monthly sales and list prices of individual models have been obtained from NPD Techworld. NPD collects point-of sale scanner data through retail channels which cover more than 80% of total sales in the U.S.. This unusually rich data set covers the time span from the introduction of the format in April 1997 till August 2003. From this dataset we generated aggregate price (LQDVDPRICE) and sales (LDSALES and LCDSALES) data for DVD players, as well as the product level variables. The characteristics data are manually collected from public sources. A study by the Bureau of Labor Statistics on the Consumer Price Index (Liegey 2001) finds the product characteristics used in our project to be empirically relevant. The descriptions of the characteristics and some basic statistics can be found in table A5 in appendix A.

We obtained total weekly sales of VCR decks from various trade magazines. These figures are collected and announced by the Consumer Electronics Association (CEA). In contrast to the DVD sales data, the sales figures reported by CEA are the number of shipments from manufacturers to dealers. Hence, actual sales to consumers occur one to one and a half months later. In the light of this, we construct monthly sales series. We have only annual average prices of VCRs. Since the VCR market is saturated we interpolate monthly prices, assuming prices fall at a constant rate over time.

[6] Polygram was then acquired by Universal in 1998

4.2 Titles Data

We employ two different datasets on software availability. The DVD titles data (LDVDTITLES) are obtained from the DVD Entertainment Group (DEG). DEG consists of major consumer electronics manufacturers, prominent film studios and, music labels. The data set includes all DVD titles, including music videos. It has detailed information on release date, genre, MPAA rating, and prices. In our data period, on average about 60 new titles were released per month. The data are available on the DEG website.[7]

We obtained data on VHS titles from Video Software Dealers Association (VSDA). VSDA has the archives of VHS titles that are released since September 1998. The number of VHS titles at the time of the introduction of the DVD format (April 1997) was not available, but currently one can search from about 32,000 titles on Amazon.com. Therefore, we assume that there were about 28,000 VHS titles available in Spring 1997. We later check if the estimation results are sensitive to this assumption on initial VHS availability.

Another software variety variable (LCMOVIES) includes only theatrical movies which are produced since 1996 and released in home video. In our sample there are about 1,800 movies. Only 1,200 are used in the estimation since we do not have complete information on all of them. This variable is our measure of software availability. The titles are comparable across the VHS and the DVD formats and there is complete information on their attributes. We also believe that for video watchers recently released theatrical movies are most relevant.

Previous studies on network effects have worked with either brand level information, but with no information on software availability, and thus were unable to show or measure the size of network effects, or with market level information on hardware sales, and hence were unable to analyze any supply side interaction. The data set we construct allows us to conduct a more in-depth analysis than any previous project. Using this dataset, we can measure the size and importance of network effects in the DVD market and potentially test several hypotheses related to supply side behavior, such as collusion of hardware manufacturers.[8]

[7] http://www.dvdinformation.com/titles/index.html

[8] One of our aims is to test the hypotheses of cooperative behavior among hardware producers, and between hardware and software producers. This is an issue we plan to investigate further in future research.

5 Identification of the network effects

5.1 Hardware Demand: Nested Logit Model

The simplest structure in which we can possibly describe the decision process of a consumer and identify the network effects is one that presents a choice between a VCR, a DVD player, and the no-purchase outside option in the upper nest and the choice of individual products in the lower nest. According to this nest structure consumers first choose between buying a VCR, a DVD player, or no video product at all. If they choose to buy a DVD player, then they decide on a specific model in the lower nest. The necessary variation for the identification of the network effect is observed in the upper nest. We expect DVD player sales to grow relatively faster as the number of DVD titles increases. The parameters on the characteristics of DVD players are identified in the lower nest. Although network effects are identified in the aggregate level, we still need to estimate a structural model to deal with the policy questions we raise later. Owing to the absence of product level data, we make a homogenous good assumption for VCR's.

The utility a representative consumer receives from purchasing a VCR or a DVD player is

$$U_{ijft} = X_{jt}\beta + \gamma P_{jt} + \eta f(N_{ft}) + \xi_{jt} + \xi_{ft} + \zeta_{ift} + \varepsilon_{ijt}$$

where the subscript $f \in (d, v, o)$ stands for the two video formats and the outside option, j for the jth product (let $j = 0$ for the outside option and $j = 1$ for VCR), and t for the time period. $f(N_{ft})$ is a function of the number of titles in each format f, at time t. ξ_{jt} is a unidimensional unobserved product characteristic. Format specific unobservables consist of ξ_{ft} and ζ_{ift}. We acknowledge the possible endogeneity of $f(N_{ft})$. Formally we have $E(\xi_{ift}|N_{ft}) \neq 0$. There might be unobserved attributes of format f,such as an unobserved taste for the specific technology, that are correlated with the software availability variable. To complete the nested logit structure we assume that $(\zeta_{ift} + \varepsilon_{ijt})$ has a type I Extreme Value distribution with scale parameter σ. Under this formulation the probability of choosing one model takes the well-known logit form.

To take into account the possible endogeneity of prices and the number of titles, we transform market shares into a linear combination of observed variables and unobserved variables following

Berry (1994). By normalizing the utility from the outside option to be $U_{i0t} = \zeta_{i0t} + \varepsilon_{i0t}$, we get

$$\ln s_{jt} - \ln s_{0t} = \delta_{jt} + \delta_{ft} + (1-\sigma) \ln s_{jft|B_f} \tag{1}$$

where $\delta_{jt} = X_{jt}\beta_1 + \gamma P_{jt} + \xi_{jt}$ is the mean value of indirect utility from product level characteristics plus the price of the product, $\delta_{ft} = \eta f(N_{ft}) + \xi_{ft}$ is indirect utility from the format level factor, and $s_{jft|B_f}$ is the within group market share of product j belonging to set B_f. The derivation of market shares and the linear estimation equation (1) in a nested logit model is provided in Appendix B.

There are two unobserved variables, ξ_{jt} and ξ_{ft}, in this equation. The estimation method depends on the correlation between the error terms and the explanatory variables. We believe that P_{jt} is correlated with unobserved product characteristics ξ_{jt}. Models with better unobserved features are more likely to have higher prices. This correlation between unobserved product attributes and prices leads to an upward bias of the price coefficient. Even though one can assume that there is no correlation between $f(N_{ft})$ and ξ_{jft}, $E(\xi_{ft}|N_{ft}) \neq 0$ is true since there is a two-way interaction between hardware and software. Therefore we need to instrument the software availability variable, $f(N_{ft})$. The within-nest share is instrumented for the same reason.

5.1.1 Instruments

If one is only interested in parameters β and γ, δ_{ft} can be treated as a group fixed effect that can be eliminated using a within-group estimation. However, since our main interest is to estimate η, this approach cannot be taken and we need to instrument for N_{ft}. There are a couple of candidates for valid instruments. Obviously, the best alternative is a cost shifter of producing a DVD title, but it is hard to obtain information on the costs of production. Prices of similar media products such as CDs and Video games, as far as their industries are competitive enough, make good instruments, since those media share some common features of production with the DVD. Unfortunately, we do not observe these prices. The lagged installed base, cumulative hardware sales up to $t-1$, will make a good instrument, since by assumption there are no direct network effects in the DVD market. But this makes a valid instrument only if there is no serial correlation within ξ_{ft}. Practically, it is notoriously difficult to deal with serial correlation in the nested logit model.[9] This is also not

[9] Goldberg (1995), pg. 908

desirable if we conceive of forward looking behavior and strategy-making of software firms. The last and the only feasible candidate is the set of characteristics of movies released on DVD. As we show when describing the data, there has been a certain tendency in releasing video on DVD. For instance, action/adventure movies and blockbuster movies are more likely to be released on DVD probably because DVD videos have special features such as 'behind the scenes', interactive games, and bonus cuts, which are valued more by fans of these types of movies. If indeed there exists a tendency, then we can use, for example, the ratio of action movies among all movies released in DVD format as an instrument. The decision to make a certain movie is usually made with a focus on theatrical revenue and long before video release and, hence should not depend on DVD release probability.

As it is typical in the discrete choice demand literature, good instruments for prices are functions of the characteristics of other products. The sums of the characteristics of all DVD player models and the number of models are chosen as instruments for the within-nest share.

5.2 Software Supply Equation

Let the profits from releasing contents on DVD take the following form:

$$\pi_{mt}^{SW} = (P_{mt} - c_{mt}) Y_t \rho^{(t_m - t_{m0})} d_{mt}(X_m, \varepsilon_{mt}) - F_t$$

where d_{mt} is a demand shifter of a movie m on DVD, which depends on the movie characteristics X_m and a shock ε_{mt}. Each movie is assumed to be characterized with this single index $X_m\beta$, where β is a vector of parameters. $\rho^{(t_m - t_{m0})}$ captures the fact that the popularity of a movie decreases over time after its theatrical or VHS release. t_{m0} denotes the date of theatrical or VHS release of movie m, and ρ falls in the range $[0, 1]$. Y_t is the potential market size

Let's also assume that d_{mt} is distributed with the truncated pdf $f_t(d_t)$, with $\int_{\tilde{d}_t}^{\bar{d}_t} f_t(d_t) = 1$. The break-even movie is characterized by d_{mt}^*, such that

$$d_{mt}^* = \frac{F_t}{(P_{mt} - c_{mt}) Y_t \rho^{(t_m - t_{m0})}}$$

Now the binary choice variable representing the DVD release decision can be expressed as

$$I_{jt} = 1(d_{mt} > d^*_{mt})$$

In order to show the determinants of the likelihood of DVD release of individual titles, we assume that d_{mt} has a lognormal distribution with shape parameter $0 < \sigma \leq 1$, that is, $d_{mt} = \exp \alpha^{-1}(X_m\beta + \varepsilon_{mt})$. Both $X_m\beta$ and ε_{mt} are normally distributed.[10] We also assume that the fixed cost - marginal revenue ratio is constant during a year and the same across movies, $\frac{F_{mt}}{P_{mt}-c_{mt}} \equiv \omega_{mt} = \omega_t$ for all m at period t.[11] Then, the probability of a movie being released on DVD is

$$\begin{aligned} P(I_{mt} &= 1) = P\left(\varepsilon_{mt} > -X_m\beta - \alpha \ln(Y_t) - \alpha \ln \rho(t_m - t_{m0}) + \alpha \ln(\omega_t)\right) \\ &= \Phi\left(X_m\beta + \alpha \ln(Y_t) + \ln \rho(t_m - t_{m0}) - \alpha \ln(\omega_t)\right) \end{aligned} \tag{2}$$

Table A6 shows the results from a probit regression that uses year dummies to capture the common variation across years, which includes the term ω_t . As expected, the installed base has a positive effect and the market age $(t_m - t_{m0})$ of a movie has a negative effect on the probability of DVD release.[12] Looking at some of the coefficients on X_m, we can verify that there is in fact some relationship between movie characteristics and the likelihood of a DVD release. First, high budget movies are more likely to be released on DVD. The budget of a movie instead of box office score is used as a proxy for being a blockbuster movie since the box office score occasionally includes revenues from video sales and hence is clearly correlated with the error term. Second, we observe that some genres are favored by studios. Action/Adventure movies have a higher possibility of being released on DVD, but many of the coefficients are not significant. Horror and thriller movies seem to have lower demand as DVD titles compared to action/adventure movies, as evident from

[10] The assumption of a lognormal distribution is in this case appropriate, since the distribution of movies gets eventually skewed to the left, due to left-over movies from earlier DVD releases.

[11] Practically we observe none of the three variables in the ratio. Although we can justify the assumption by arguing that in a monopolistically competitive market the price-cost margin remains constant and that there is not much variation in the fixed cost, this is obviously a very strong assumption

[12] The negative coefficient on age could be due to selection, because there are some movies which are never released on DVD. Our maximum duration of time is 86 months. We constructed another variable by limiting age to 24 months and ran the same regression, but there was no significant change in the results (α became .494 and $\alpha \ln \rho$ became -.081 and both were still significant). We thank Martino De Stefano for pointing this out.

the negative and significant coefficients.

Although the probit regression is sufficient to verify the existence of network effects in the software side, and gives some justification for using movie characteristics as instruments for software variety, a reduced form aggregate supply equation 3 is used in the estimation below. It is because we want to estimate both the hardware demand and the software supply simultaneously, and because we want to derive title variety equation from equation 2, that we need a more complicated structure.[13] The aggregate supply of software is given by

$$\ln N_t = X_t^S \beta_2 + \eta_2 \ln Y_t + e_t \tag{3}$$

where X_t^S is a vector of the aggregate characteristics of titles, such as the number of action movies or the sum of the budgets of movies at time period t. Y_t is the installed base, which is given by the accumulated sales of DVD players.

We basically estimate two simultaneous equations: the hardware demand equation and the reduced form software supply equation.

6 Results

6.1 Aggregate Level Estimation

For a preliminary check, we estimate two aggregate level equations for the DVD market. The system of equations is

$$\begin{aligned} \ln HW_t &= a_1 + \gamma \ln P_t + \eta_1 \ln N_t + \xi_t \\ \ln SW_t &= a_2 + \eta_2 \ln IB_t + \zeta_t \end{aligned} \tag{4}$$

where HW_t and SW_t are the monthly sales of players and the total number of title releases, respectively. N_t denotes title availability, given by the total number of different DVD titles. IB_t is the installed base of DVD players. P_t is the quality adjusted average price of DVD players,

[13] The number of titles produced by a studio s is $N_{st} = K_t \int_{d_t^*}^{\bar{d}} f_t(d) dd$, where K_t is the number of movies which are candidates for DVDs, $f_t(d)$ is the pdf of the demand shifter. We need to make assumptions on the distribution of recently available movies. But the real difficulty lies in dealing with the changing distribution of d over time, caused by the left-over candidate movies. We are tackling this problem in a dynamic model in another work-in-progress.

obtained from a hedonic regression of prices on the characteristics of players.[14]

Our first concern is on the stationarity of our aggregate variables. Equation (4) is potentially a spurious regression equation, if the dependent and independent variables have similar upward trends. In that case we are very likely to find a significant coefficient for $\ln N_t$, even if there is no actual relationship but just a common trend (Davidson and MacKinnon 1993). In this respect, all variables in the model are tested for stationarity (4). We cannot reject the non-stationarity of some level variables, but all logged variables are stationary.

Second concern is on the possible serial correlation of the independent variables and the error term. Testing for autocorrelation reveals that the disturbances in the hardware demand equation (4) are serially correlated[15]. To deal with this problem we include lagged dependent variables in the regression equation. An additional advantage of having lagged dependent variables is that they help us capture "direct effects" or "knowledge spillover" by controlling for previous periods' volume of sales.

The last issue that requires consideration in the aggregate level estimation is the possible endogeneity of N_t and IB_t. As mentioned earlier, we use aggregate characteristics of movies as instruments for $\ln N_t$. These include the cumulative number of VHS movies in each genre, MPAA rating, and the studio of production. We use the natural logarithm of price as an instrument for the installed base in the software supply equation. The Durbin-Watson statistics were below the lower bound in all the regressions. After adding an AR(1) term, the Durbin-Watson statistics are all above the upper bound. Quarter dummies are also added to control for seasonal effects.

The regression results are presented in Tables 1 and 2 below. For the hardware demand equation we use both the cumulative number of DVD titles and that of DVD movies released in theaters since 1996 as measures of software availability. Both measures appear to have a positive effect on player sales. We also report Newey-West t-statistics as a robustness check to control for unknown serial correlation, as in Gandal, Kende, and Rob (2000). All the coefficients remain significant. After instrumenting, the effect of title variety is still significant, although the magnitude falls somewhat. Our results indicate that when the number of DVD movies increases by 10 percent, monthly DVD

[14] We do not have prices in the software supply equation since, as described before, we assume that the price-cost margin remains constant across titles and time.

[15] Following Gandal, Kende, and Rob(2000), we calculate the Durbin-Watson statistic to check for possible serial correlation and use Newey-West robust standard errors to deal with it.

players sales increase by about 1.5 percent in the aggregate level. Price coefficients are all negative and significant as expected.

Table 1. Aggregate Hardware Demand Estimation

	OLS all titles	OLS - movies	Newey-West - movies	IV - movies
LCDMOVIES		0.151 (2.26)*	0.151 (2.32)*	0.146 (2.18)*
LDVDTITLES	0.331 (3.88)**			
LQDVDPRICE	-0.199 (2.27)*	-0.408 (4.29)**	-0.408 (5.00)**	-0.409 (4.32)**
LDSALES_1	0.441 (4.35)**	0.490 (5.04)**	0.490 (6.33)**	0.494 (4.98)**
Q1	-0.777 (5.03)**	-0.739 (4.63)**	-0.739 (4.56)**	-0.741 (4.65)**
Q2	-0.569 (4.11)**	-0.584 (4.08)**	-0.584 (4.12)**	-0.583 (4.09)**
Q3	-0.636 (4.53)**	-0.651 (4.54)**	-0.651 (4.31)**	-0.650 (4.52)**
Constant	5.648 (4.82)**	7.776 (5.51)**	7.776 (6.53)**	7.758 (5.47)**
Observations	77	78	78	78
DW statistic	2.40	2.24	2.24	2.24
R-squared	0.95	0.95		0.95

Robust t statistics in the parentheses; * significant at 5%; ** significant at 1%
Dependent variable is log of monthly aggregate DVD player sales (LDSALES).
Instruments for the title variety are cumulative number of movies for each genre, MPAA rating, and studio.

Table 2. Software Supply Equation Estimation Results

	OLS	IV	OLS w/ extra	IV w/ extra
LCDSALES	0.176 (4.07)**	0.125 (3.25)**	0.200 (4.35)**	0.147 (2.81)**
LDMOVIES_1	0.391 (2.30)*	0.485 (2.91)**	0.297 (1.62)	0.384 (2.14)*
Constant	-0.928 (1.89)	-0.422 (0.37)	-1.426 (1.32)	-0.737 (0.64)
DW static	2.42	2.67	2.03	2.13
Observations	76	76	71	71
R-squared	0.66	0.65	0.74	0.74

Robust t statistics in parentheses; * significant at 5%; ** significant at 1%
Dependent variable is the log of monthly release of DVD movies. The last two columns include the number of new VHS movies sorted by genre and rating
but the coefficients are not reported.
Instrument for log of cumulative player sales is the log of price.

Table 2 shows the estimation results from the software supply equation. The first two columns correspond to the OLS and the IV regressions with only the installed base and the lag of the

dependent variable as the explanatory variables. The natural logarithm of the price of DVD players is used as an instrument for sales of players in the IV regressions. The coefficients on the installed base show yet another direction of the network effect. As more people own DVD players, more titles are provided. Instrumenting lowers the magnitude of the coefficients since it removes the upward bias in the estimates. We also ran the regressions with some extra movie characteristics, such as the number of new release action/adventure movies and R-rated movies in VHS format given the month, as additional independent variables. The additional variables add to the explanatory power of the estimation but do not change the parameter estimates substantially.

6.2 Product Level Estimation : Nested Logit

Next, we estimate the product level nested logit model, given by equation 1. The results are presented below in Table 3. For the nested logit estimation we need to make an assumption on the market size. The number of U.S. households is chosen as the size of the market for DVD players and VCR's. The market size in the first period, M_0, is 100 million. Consumers are assumed to exit the market once they buy a VCR or a DVD player. Thus, the market size at period T is $M_0 - \sum_{t=1}^{T} [Sales_t(VCR) + Sales_t(DVD)]$.

Table 3. Nested Logit Regression of Hardware Demand

	BASE	IV (HWchar)	IV (HW/SW char)
HPRICE	-0.038 (12.31)**	-0.076 (9.86)**	-0.091 (9.92)**
LCMOVIES	0.970 (31.41)**	0.787 (26.52)**	0.681 (18.26)**
LWITHIN	0.955 (334.00)**	0.841 (105.51)**	0.805 (83.47)**
SEASON	1.201 (81.40)**	1.159 (63.83)**	1.142 (56.84)**
COMP	0.081 (5.31)**	0.094 (4.87)**	0.121 (5.50)**
OPT	0.000 (0.03)	-0.042 (2.26)*	-0.053 (2.54)*
COAX	0.002 (0.16)	-0.018 (1.14)	-0.029 (1.63)
DD	-0.052 (4.15)**	-0.079 (5.16)**	-0.092 (5.31)**
DTS	0.065 (4.22)**	0.045 (2.01)*	0.049 (1.97)*
CDR	0.123 (9.19)**	0.161 (10.14)**	0.192 (10.19)**
MP3	0.136 (10.58)**	0.153 (8.00)**	0.161 (7.49)**
DVDR	0.118 (11.11)**	0.165 (10.68)**	0.187 (10.30)**
VHS	0.892 (8.61)**	1.493 (12.13)**	1.693 (13.67)**
PROG	0.076 (6.56)**	0.083 (4.62)**	0.094 (4.52)**
REC	0.312 (8.68)**	0.454 (7.63)**	0.518 (7.44)**
PORT	0.188 (7.56)**	0.247 (5.31)**	0.300 (5.57)**
MULTI	0.088 (8.33)**	0.079 (5.90)**	0.091 (5.82)**
Constant	-12.133 (65.16)**	-11.607 (66.54)**	-11.148 (52.46)**
Observations	7388	6979	6979
R-squared	0.98	0.97	0.96

Robust t statistics in parentheses; * significant at 5%; ** significant at 1%
The first stage R^2 is 0.88 for the first IV regression, and 0.98 for the second.
Dependent variable is the difference between the log of market share of product j and the outside option.

In order to minimize the measurement error and to reduce the magnitude of unobserved heterogeneity, some outliers are dropped from the sample. First, we sort out the observations with price higher than \$1,345. A price of \$1,345 is the outer fence of the severe outlier, which is characterized by 75 *percentile* + 3(75 *percentile* − 25 *percentile*). We also exclude the observations with price less than \$10.

The first column reports the results from the base OLS estimation of equation 1. Most of the coefficients have the expected signs. Greater variety of titles has a positive effect on consumer utility and, hence on hardware demand. The OLS regression results suggest that a 1% increase in the number of movies released on DVD raises the market share of a DVD player model relative to no-purchase option by 0.98%. As expected, the higher the price, the lower is utility. The positive and highly significant seasonal dummy variable suggests a strong Christmas effect. The large coefficient on the within-nest share tells us that nesting products into categories of VCR and DVD is indeed meaningful. Results on digital audio outputs (optical and coaxial) are negative

but not significantly different from zero. In the latter two columns we report the results from the IV regressions. Initially, we use only hardware characteristics as instruments, in line with other empirical papers on network effects.[16] Berry, Levinsohn, and Pakes (1995) find that product characteristics, the sum of characteristics of the same brand products, and the sum of other brands' characteristics serve as valid instruments. We find that OLS estimates for price, software variety, and the within-group share are upward biased. After instrumenting, the coefficients become smaller in magnitude but the signs do not change. When we include software side instrumental variables such as the cumulative number of movies in each genre and rating, we have even smaller but still significant coefficients. The effect of software availability on hardware demand is 0.68 and is still significantly different from zero.

In the above regressions we chose the logarithm of the number of DVD movie titles as the functional form of $f(N_t)$. We also ran the same regressions with alternative variables and different functional specifications, and the results are provided in table A7 in the appendix. In the second column of the table we present the results from a regression with a quadratic specification for DVD movie titles. The variable HCMOVIES is the cumulative number of post-1996 movie titles. HCMOVIES2 is the squared HCMOVIES term. We still find a positive effect of variety on consumer utility. Additionally, we observe that the marginal effect of an increase in titles diminishes. In the last column of table A7, log of all video (DVD and VHS) titles, instead of only movies, is used. Due to the scale of the variable the coefficient is smaller but still positive.

6.3 Simulations

As stated earlier, an integrated body of hardware firms may be able to internalize the network externality in a way a single firm is not able to. One producer of hardware among many will not find it profitable to lower the price of one its products in order to exploit the network effect. But for a monopolist that produces all the models in the industry, the potential gain of a price reduction might be much larger. In this section we simulate the effect of a one period change in the price of one product, holding all future prices constant. We compute the change in revenues for the chosen product, the firm producing the product, and the whole industry. More specifically, we calculate the change in revenues that results from the reduction in the price of one DVD player model of a

[16] As far as we know only Gandal, Kende, and Rob (2000) used software side variables (fixed cost) as instruments.

single firm.

When a firm lowers the price of its product, there are two effects at play. First, some consumers who already wanted to buy a DVD player switch from other models to the one with the reduced price. We call this the 'business stealing' effect. Second, some other consumers who were not going to buy any DVD player now find it optimal to buy the cheaper model. These purchases put the network effect into action. An increase in the installed base of DVD players induces studios to release more movies on DVD. As software variety expands, DVD players become more popular and sales of all models increase. We call this second effect the 'market expansion' effect. A single hardware firm does not have as much incentive to lower the price of its product because it is not be able to fully internalize the market expansion effect.

Our counterfactual experiment runs as follows: We perturb the equilibrium set of prices by lowering the price of one DVD player for just one period, holding all other prices constant. We simulate the change in software availability and the resulting increase in the demand for DVD players using our results from sections 6.1 and 6.2. Then, we compare the change in revenues of the single firm with that of the whole industry. According to our simulations, a 10% reduction in the price of one product raises the sum of current and future revenues of the price cutting firm by $253 million. In the first month, the revenues of the firm decrease by 2.3 million.That is because the own price elasticity of demand is less than one (0.6), and there is intra-brand business stealing. Over the time period of our dataset, total industry revenues increase by $798 million, a significantly larger increase than that for a single firm. On the other hand, consumer surplus increases as consumers enjoy increased software availability and as more consumers buy hardware. We find that the consumer surplus of the representative consumer rises by $233 over the period.[17] Revenues from the sales of VCR's decrease by $4 million, as some consumers switch from buying VCR's to buying DVD players.

The large increase in revenues is the result of the fast expansion of software variety. As we see in the figure 3, the number of DVD titles rapidly catch up with the number of VHS titles when a lower price induces more consumers to buy DVD players. The number of DVD titles is bounded upward by the number of VHS titles following our earlier assumption that all movies are released

[17] We chose the largest hardware firm Sony and its most popular product for our experiment. The results depend on the market share and the price elasticity of the product among other things.

on VHS.

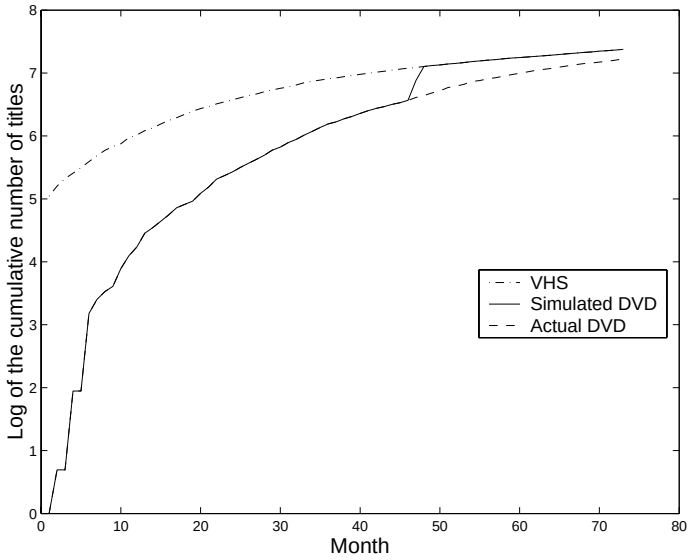

Figure 3: Simulation of the number of DVD titles

This simple experiment shows that there are significant gains to consolidation. A horizontally integrated hardware side can exploit the network effect to increase its revenues.[18] The increase in industry revenues is much greater than the increase in the revenues of a single firm. This confirms our belief that hardware firms can do better by making joint pricing decisions in order to exploit the network effects. Economists are usually skeptical about any form of integration since these lead to increased market concentration. Under network effects, however, consumers benefit from the increased availability of software. Without more information on costs of production, or setting up a structural dynamic model of hardware supply, it is not possible to provide more clear answers to this issue. We need to solve for and compute the set of new equilibrium prices to reach a conclusion on the welfare impacts of integration.

7 Conclusion

Indirect network effects arise when the benefit of a product is increasing in the use of a complementary set of goods. For instance, the utility one gets from owning a DVD player (hardware) increases

[18] We delibreately avoid mentioning profits since calculating the change in profits would require us to have marginal cost information.

in the number of titles that are available on DVD (software). At the same time demand for software depends on the number of users of the hardware. In this paper we estimate the complementarity between hardware and software in the U.S. DVD market using data on hardware sales and title availability.

First, we estimate an aggregate level equation and find that the elasticity of hardware demand with respect to software availability is about 0.15%. Next, we estimate a nested-logit model at the level of the individual product. The identification of the network effect comes from the differences in software availability between two different systems: VHS and DVD. The endogenous software availability variable is instrumented by various movie characteristics. Our results indicate that a 1% increase in software availability raises demand for hardware by 0.68%. Compared to other estimates in the literature this number is quite high. The reason for this difference is that we have complete information on software availability and thus are able to measure the network effect accurately. Finally, we estimate a software supply equation with the number of DVD titles as the dependent variable. The coefficient on the hardware installed base variable indicates the other direction of the network effect. As more DVD players are sold, more titles are released on DVD.

If left to itself, a market of complementary goods can end up in an inferior equilibrium where the hardware adoption rate and software availability are less than optimal. Our results show that network effects are strong in the DVD market, leaving room for a negative externality of considerable size. Without more information on software provision costs, we cannot calculate the optimal size of the market. One thing we do is verify the value of integration to hardware firms. We simulate the effect of a price change while holding all future prices constant. We compute the change in revenues for the product, for the firm, and for the industry. Comparing these measurements gives us an idea of how much of the network effect is being missed by an oligopoly structure. According to our simulations, the revenues of the single firm increase by $126 million, whereas total industry revenues increase by a much larger $424 million. Consumer surplus increases as consumers enjoy increased software availability. These results indicate that the horizontal integration of hardware firms does not only increase total revenues but may also raise social welfare. One word of caution on the interpretation of the results is necessary. At the new equilibrium with an increase in software availability, hardware firms will charge higher prices. Thus, the net change in consumer surplus cannot be found without computing the set of new equilibrium prices.

An interesting question that remains to be answered is the outcome of vertical integration of DVD player manufacturers and movie studios. We have some formal and plenty of anecdotal evidence on mergers and cooperative behavior between the two sides. Our findings suggest that making coordinated hardware pricing and title release decisions may benefit vertically integrated firms. On the other hand, hardware prices will most likely rise, making it difficult for us economists to predict the welfare aspects of this interaction. We plan to extend the present model to take into account this strategic coordination aspect of the industry.

Appendix A. Additional Tables

Table A1. Summary of DVD players data in annual averages

year	number of models	monthly sales	price
1997	97.7	15443	483.28
1998	120.8	58651	501.67
1999	150.6	254771	481.27
2000	180.8	466895	432.56
2001	260.1	830658	358.31
2002	365.3	1227821	282.27
2003(Aug)	444.3	990335	218.33
Average	219.2	620301	345.59

Table A2. Annual summary of DVD titles: number of titles and share of genre

year	titles	act/adv*	sci-fi	drama	animation	family	sus/thr*
1997	445	0.164	0.038	0.229	0.045	0.052	0.004
1998	1445	0.134	0.042	0.207	0.037	0.014	0.013
1999	2827	0.079	0.040	0.184	0.065	0.016	0.037
2000	4009	0.095	0.040	0.159	0.074	0.021	0.036
2001	6214	0.108	0.034	0.137	0.098	0.021	0.045
2002	8009	0.090	0.025	0.130	0.117	0.030	0.044
2003	5335	0.085	0.019	0.167	0.115	0.036	0.067

* act/adv – action/adventure; sus/thr – suspense/thriller

Table A3. Annual summary of DVD titles: share of MPAA rating

year	G	NR	PG	PG-13	R
1997	0.031	0.443	0.117	0.076	0.301
1998	0.011	0.562	0.080	0.077	0.224
1999	0.025	0.600	0.059	0.061	0.218
2000	0.021	0.636	0.052	0.048	0.204
2001	0.020	0.642	0.051	0.042	0.166
2002	0.023	0.692	0.041	0.040	0.123
2003	0.029	0.653	0.033	0.039	0.127

Table A4. Description of Aggregate Variables

Variable	Description	Observation	Mean	SD	Min	Max
LDSALES	log of monthly DVD player sales	81	12.16	1.71	8.20	15.29
LCDSALES	log of cumulative DVD player sales	81	14.76	2.28	9.31	17.56
LQDVDPRICE	log of quality-adjusted DVD player price	81	4.56	1.27	2.28	6.44
LDMOVIES	log of monthly new release DVD movies	76	2.82	0.68	0	3.83
LCDMOVIES	log of cumulative DVD movies	78	5.76	1.65	0	7.30
LDVDTITLES	log of all DVD titles available	77	8.02	1.75	3.18	9.91

Table A5. Description of Product Attributes

Variable	Description	Obs	Mean	Std. Dev.
COMP	Component video output. For higher quality picture.	8814	0.742	0.438
OPT	Optical digital audio output. For higher quality sound.	8814	0.848	0.359
COAX	Coaxial digital audio output. For higher quality sound.	8814	0.785	0.411
DD	Built-in Dolby Digital audio decoder. For higher quality sound.	8742	0.300	0.458
DTS	Built-in Digital Theater Systems (DTS) decoder. For higher quality sound.	8742	0.119	0.324
CDR	CD-R or/and CD-RW playable.	8742	0.480	0.500
MP3	MP3 playable.	8742	0.290	0.454
DVDR	DVD-R, a DVD recording format, playable.	8778	0.148	0.355
VHS	VHS playable. Some DVD players ("combo") play VHS.	81	1	0
PROG	Progressive scan. For higher quality picture.	8778	0.209	0.406
REC	Recorder.	8814	0.013	0.112
PORT	Portable.	8814	0.078	0.269
MULTI	Multidisc. Value is 1, if a player has a multiple disc exchanger.	8814	0.213	0.409

Note: All variables take the value of 1 if a player has that feature and 0 otherwise.
Only VCRs have the VHS playback feature when DVD-VCR combos are dropped from the sample.

Table A6. Probit regression of DVD movie release

log of Installed Base	0.511 (14.23)**
Age	-0.048 (21.56)**
log of budget	0.053 (2.36)*
Animation	0.317 (1.23)
Comedy	-0.056 (0.64)
Comedy/Drama	-0.329 (1.24)
Drama	-0.147 (1.62)
Family	-0.037 (0.21)
Foreign	-0.455 (1.72)
Horror	-0.473 (4.05)**
Musical	-0.171 (0.52)
Mystery/Suspense	0.134 (1.11)
SciFi/Fantasy	-0.133 (1.16)
Thriller/Suspense	-1.649 (5.54)**
NR	0.175 (0.58)
PG	0.023 (0.10)
PG13	0.128 (0.51)
R	0.238 (0.95)
Constant	-9.775 (10.17)**
Observations	5560

Absolute values of z statistics in parentheses. * significant at 5%; ** significant at 1%
Genre Action, MPAA rating G, and year dummy 1997 are omitted.
Studio dummies are included but the coefficients are not reported.

Table A7. Other Specifications of Software Availability

	IV Logarithm	IV Quadratic	IV All titles
HPRICE	-0.091 (9.92)**	-0.100 (10.11)**	-0.023 (3.96)**
LCMOVIES	0.681 (18.26)**		
HCMOVIES		0.327 (29.28)**	
HCMOVIES2		-0.013 (28.97)**	
LALLTITLES			0.786 (43.58)**
LWITHIN	0.805 (83.47)**	0.781 (61.74)**	0.900 (135.70)**
SEASON	1.142 (56.84)**	1.125 (50.60)**	1.165 (77.37)**
COMP	0.121 (5.50)**	0.157 (7.03)**	0.044 (3.21)**
OPT	-0.053 (2.54)*	-0.047 (2.12)*	-0.040 (2.98)**
COAX	-0.029 (1.63)	-0.032 (1.66)	-0.016 (1.37)
DD	-0.092 (5.31)**	-0.100 (5.30)**	-0.042 (3.70)**
DTS	0.049 (1.97)*	0.046 (1.68)	-0.010 (0.60)
CDR	0.192 (10.19)**	0.199 (9.78)**	0.074 (6.05)**
MP3	0.161 (7.49)**	0.163 (6.72)**	0.106 (7.72)**
DVDR	0.187 (10.30)**	0.224 (10.26)**	0.098 (8.34)**
VHS	1.693 (13.67)**	1.915 (13.97)**	0.348 (3.88)**
PROG	0.094(4.52)**	0.092 (4.04)**	0.016 (1.16)
REC	0.518 (7.44)**	0.547 (6.89)**	0.172 (3.90)**
PORT	0.300 (5.57)**	0.324 (5.74)**	0.035 (1.02)
MULTI	0.091 (5.82)**	0.074 (4.31)**	0.020 (1.93)
Constant	-11.148 (52.46)**	-8.493 (118.65)**	-13.073 (95.85)**
Observations	6979	6979	6974
R-squared	0.96	0.95	0.98

Appendix B. Derivation of the linear nested logit estimation equation

Consider an individual who faces a set of J mutually exclusive alternatives and must make a choice among them. The nested logit model in this paper formulates the choice of this consumer as a two-stage process. The alternatives are divided into subgroups, such that ones with common characteristics are grouped together. Consumers first choose a group of alternatives, and then the specific product they want to consume among that group of alternatives. In the following the time subscript has been omitted for the sake of notational simplicity.

The utility consumers receive from consuming product $j \in B_f$ is

$$U_{ijf} = \delta_j + \delta_f + \zeta_{if} + \varepsilon_{ij}$$

where $\delta_j = X_j\beta + \gamma P_j + \eta f(N_f) + \xi_j$ is product level mean utility, $\delta_f = \eta f(N_f) + \xi_f$ is format level utility, and ε_{ij} is identically and independently distributed extreme value. ζ_{if} is common to all J products in group f and has a distribution function that depends on σ. Cardell (1997) shows that the distribution of ζ_{if} is such that, if ε_{ij} is an extreme value random variable, then $(\zeta_{if}+\varepsilon_{ij})$ is also an extreme value random variable, hereafter ϵ_{ij}. The parameter σ measures the within group correlation of utility levels. As σ approaches one, the decision model collapses into an ordinary logit, where nests disappear. As σ approaches zero, nests become more important, and substitution across groups decreases.

Suppose that the individual has chosen the set B_f at the upper nest. The probability of choosing product $j \in B_f$ then becomes

$$P_{B_f}(j) = \Pr(\epsilon_{i1} - \epsilon_{ij} \leq \delta_j - \delta_1 \ \ldots \ \epsilon_{iJ} - \epsilon_{ij} \leq \delta_J - \delta_1$$

Let F be the common cumulative distribution function of ϵ_{ij}, and let f be the corresponding density. Remembering that the ϵ_{ij} are i.i.d., the probability that alternative j will be chosen can be written as

$$P_{B_f}(j) = \int_{-\infty}^{\infty} f(x) \prod_{k \neq j} F(\delta_j - \delta_k + x) dx$$

Following Anderson, de Palma, and Thisse (1992) we specify the distribution of ϵ_{ij} which generates the logit market share equations. ϵ_{ij} are distributed i.i.d. double exponential (from the extreme value family of distibutions) with the distribution function

$$F(x) = \Pr(\epsilon_{ij} \leq x) = \exp -[\exp -(\frac{x}{\sigma} + \gamma)],$$

and the probability density function

$$f(x) = e^{-(\frac{x}{\sigma}+\gamma)} e^{-e^{-(\frac{x}{\sigma}+\gamma)}}$$

where γ is Euler's constant and σ is a scale parameter of the distribution. Using the above distribution and density functions, the conditional probability of choosing j among all the products in group f is given by

$$P_{B_f}(j) = \frac{\exp(\delta_j/\sigma)}{\sum_{k=1}^{J} \exp(\delta_j/\sigma)} \tag{5}$$

The proof can be found in Anderson, de Palma, and Thisse (1992).

Similarly, the probability of choosing group f in the upper nest is

$$P(B_f) = \frac{D_f^{\sigma}}{\sum_{g=1}^{G} D_g^{\sigma}} \tag{6}$$

where $D_f = \sum_{j \in B_f} \exp[(\delta_j + \delta_f)/\sigma]$ is the expected indirect utility of subset f (often called the *inclusive value* of subset f) and $\sum_{g=1}^{G} D_g^{\sigma}$ is the sum of expected indirect utilities of all subsets.

Finally, the probability of choosing alternative $j \in B_f$ can be written as the product of the two probabilities given by Equations 5 and 6:

$$P(j) = P_{B_f}(j) \cdot P(B_f) = \frac{\exp(\delta_j/\sigma)}{\sum_{k=1}^{J} \exp(\delta_j/\sigma)} \cdot \frac{D_f^{\sigma}}{\sum_{g=1}^{G} D_g^{\sigma}} \tag{7}$$

Note that this expression is the probability of the representative consumer choosing product j, which also equals s_j, the market share of product j.

We can now derive the simple analytic expression for market shares. For the outside option, $\delta_j = 0$, and hence the market share $s_0 = 1/\sum_{k=1}^{J} \exp(\delta_j/\sigma)$. Using this expression, we can define the market share of product j in terms of its deviation from the market share of the outside option in a linear form:

$$\ln s_j - \ln s_0 = \delta_j + \delta_f + (1-\sigma) \ln s_{jft|B_f}$$

The main advantage of this formulation is that it allows us to use instrumental variables techniques in the estimation. Price and software variety enter the market share equation linearly and now can be instrumented.

References

Anderson, S. P., A. de Palma, and J.-F. Thisse (1992): *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge, MA.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–90, July.

Berry, S. T. (1994): ""Estiamting Discrete-Choice Models of Product Differentiation"," *RAND Journal of Economics*, 25 (2), 242–262.

Cardell, N. S. (1997): "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 13 (2), 185–213.

Chou, C. F., and O. Shy (1990): "Network Effects without Network Externalities," *International Journal of Industrial Organization*, 8(2), 259–270, JUN.

Church, J., and N. Gandal (1992): "Network Effects, Software Provision, and Standardization," *Journal of Industrial Economics*, 40(1), 85–103, MAR.

Church, J. R., N. Gandal, and D. Krause (2002): "Indirect Network Effects and Adoption Externalities," Foerder Institute for Economic Research Working Paper 02-30.

Davidson, R., and J. G. MacKinnon (1993): *Estimation and Inference in Ecnometrics*. Oxford University Press, New York, New York.

Dranove, D., and N. Gandal (2003): "The DVD Vs.DIVX Standard War: Empirical Evidence of Network Effects and Preannouncement Effects," *Journal of Economics and Management Strategy*, 12(3), 363–386.

Farrell, J., and G. Saloner (1985): "Standardization, Compatibility, and Innovation," *RAND Journal of Economics*, 16, 70–83.

Gandal, N. (1994): "Hedonic Price Indexes for Spreadsheets and an Empirical-Test for Network Externalities," *Rand Journal of Economics*, 25(1), 160–170, SPR.

Gandal, N., M. Kende, and R. Rob (2000): "The dynamics of technological adoption in hardware/software systems: the case of compact disc players," *Rand Journal of Economics*, 31(1),

43–61, SPR.

Goldberg, P. K. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, July, 891–951.

Katz, M., and C. Shapiro (1985): "Network Externalities, Competition, and Compatibility," *American Economic Review*, 75, 424–440.

Liegey, P. R. (2001): "Developing an Hedonic Regression Model For DVD Players In the U.S. CPI," mimeo, Bureau of Labor Statistics.

Nair, H., P. Chintagunta, and J. Dube (2003): "Empirical Analysis of Indirect Network Effects in the Market for Personal Digital Assistants," mimeo, Graduate School of Business, University of Chicago.

Ohashi, H. (2003): "The Role of Network Effects in the U.S. VCR Market, 1978-1986," *Journal of Economics and Management Strategy*, 12(4), 447–494.

Rysman, M. (2003): "Competition Between Networks: A Study of the Market for Yellow Pages," forthcoming, The Review of Economic Studies.