Discussion Paper No. 11-084

Market Structure and Market Performance in E-Commerce

Franz Hackl, Michael E. Kummer,
Rudolf Winter-Ebmer, and Christine Zulehner

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 11-084

Market Structure and Market Performance in E-Commerce

Franz Hackl , Michael E. Kummer,
Rudolf Winter-Ebmer, and Christine Zulehner

Non-technical Summary

Analyzing the link between market structure and market performance is of central importance in the field of industrial organization. The aim is to understand the role of market structure, i.e., the number of firms in a market, their sizes and the products they offer, in determining the extent of market competition and market performance.

In particular, antitrust and regulatory authorities are interested in knowing how many firms it takes to sustain competition in a market. For example, the expected relation between the number of firms in the market and market outcomes such as prices or qualities is at the core of merger assessments. These questions are of central importance to society and to the consumers, because a minimum level of competition ensures both sufficient provision of the good and reasonable prices for the consumer.

In this paper, we investigate the interaction between market structure and market performance in e-commerce for consumer electronics. We use data from Austria's largest online site for price comparisons combined with retail-data on whole sale prices provided by a major hardware producer. We observe firms' prices as well as their input prices, and all their moves in the entry and the pricing game. With this information we can analyze how sellers' markups over the producer's wholesale price react to the number of firms that compete in the market. We also look at the impact of market structure on market performance over the product life cycle, as other studies focusing on market structure find that entry has, especially at the beginning of the life cycle, a significant impact on prices.

An important contribution of this paper stems from a novel way of dealing with the phenomenon that, it is extremely easy for e-commerce-shops, to add and remove items from their product portfolio. This makes the analysis very difficult, because usually the number of shops will be related to how attractive an item is to sell, and will thus depend on the variables we wish to explain. This situation is an example of the "endogeneity problem", which can pose a threat to the validity of empirical estimates. We are dealing with this issue by using the information on how many shops typically listed earlier cameras at a particular stage of the life-cycle (in the past). This variable will capture overarching factors in the listing decision (e.g. distribution patterns), that cannot so easily be changed and is hence immune to such issues as whether an item is currently en vogue or not.

We find a very short lifecycle of usually less than a year and a highly significant and strong effect of the number of firms on markups. Ten additional competitors in the market reduce the markup of the cheapest firm by more than 1.5 percentage points on average. We also find that this effect is strongest in the first month, but also in the end of the lifecycle. Interestingly, markups were found to be lowest in months 2-4, but at the same time, the number of firms seems to be less relevant to pushing down prices in that stage of the lifecycle. For the consumer this means that by waiting three more weeks she will get the same price reduction she would get by going to a market with one additional firm.

Das Wichtigste in Kürze

Die Analyse des Zusammenhangs zwischen der Struktur und der Funktionsfähigkeit eines Marktes ist eine der zentralen Fragen auf dem Gebiet der Industrieökonomie. Dabei gilt es zu verstehen, wie die Marktstruktur (also die Anzahl der Firmen im Markt, deren Marktanteil sowie die Anzahl, Qualität und Charakteristika ihrer Produkte) das Ausmaß des Wettbewerbs im Markt beeinflusst.

Vor allem Wettbewerbs- und Regulierungsbehörden wollen wissen wie viele Firmen nötig sind um den Wettbewerb in einem Markt aufrecht zu erhalten. Um nur ein Beispiel zu nennen, bei der Bewertung einer beabsichtigten Firmenfusion ist der zu erwartende Zusammenhang zwischen der Anzahl der Firmen im Markt und den Preisen oder der durchschnittlich angebotenen Qualität die zentrale Entscheidungsgrundlage. Diese Fragen sind folglich von hoher Relevanz für die Gesellschaft und die Endverbraucher, da ein Mindestniveau an Wettbewerb sowohl die Bereitstellung der Güter als auch transparente Preise sicherstellt.

In dieser Arbeit untersuchen wir den Zusammenhang von Struktur und Funktionsfähigkeit von Märkten im Bereich des E-Commerce. Hierfür verwenden wir Daten der größten österreichischen Online-Preisvergleichsseite und Vertriebsdaten eines der größten Hersteller von Haushaltselektronik. Wir beobachten die Großhandelspreise, die Preise der Firmen und deren Inputpreise, sowie alle Aktionen der Einzelhändler (Markteintrittsentscheidungen und Preissetzung). Mit dieser Information sind wir in der Lage zu untersuchen, wie der Aufschlag auf den Großhandelspreis auf die Anzahl der Firmen im Markt reagiert. Darüber hinaus analysieren wir, wie sich der Zusammenhang von Marktstruktur und Funktionsfähigkeit über den Lebenszyklus der Produkte verändert, da die Marktstruktur, wie in früheren Studien gezeigt wurde, vor allem am Beginn des Lebenszyklus große Auswirkungen auf die Preise haben kann.

Ein wesentlicher Beitrag dieser Arbeit liegt darin, explizit zu berücksichtigen, dass E-Commerce Händler sehr einfach Produkte in ihr Sortiment aufnehmen und bald danach wieder aus dem Sortiment auslisten können. Dies erschwert eine objektive Analyse, da davon auszugehen ist, dass die Anzahl der Firmen von vielen stark variablen Faktoren abhängt. So können Händler beispielsweise darauf reagieren, wie stark ein Produkt in den Medien präsent ist oder wie seine Profitabilität variiert. Wir können dieses Problem lösen, indem wir auf die Information aus früheren Lebenszyklen zurückgreifen, um so zu erfahren, wie viele Händler ein Produkt in einer bestimmten Phase des Lebenszyklus „üblicherweise“ in ihr Sortiment aufnehmen. Diese Variable ist abhängig von langfristigen, zum Teil nicht beobachtbaren, Faktoren (wie zum Beispiel der Vertriebskette), die auf die Sortimentswahl einen Einfluss haben, aber nicht rasch geändert werden können nur weil ein Produkt gerade in Mode ist.

Wir gelangen zum Ergebnis, dass der Lebenszyklus der untersuchten Produkte sehr kurz ist und dass ein starker negativer Zusammenhang zwischen der Anzahl der Firmen und der Preisaufschläge im Markt besteht. Zehn Firmen mehr im Markt senken den Preisaufschlag des Bestbieters um mehr als 1,5 Prozentpunkte. Außerdem finden wir, dass der Effekt im ersten Monat besonders stark ist, in den Monaten zwei bis vier etwas abflacht, um dann ab dem sechsten Monat wieder zuzunehmen. Gerade dieses letzte Ergebnis ist gewissermaßen überraschend. Für den Konsumenten bedeutet dies, dass er durch das Abwarten von 3 Wochen dieselbe Preisreduktion erwarten kann wie wenn er einen Markt sucht, auf dem ein Händler mehr anbietet.

Market Structure and Market Performance in E-Commerce *

Franz Hackl
University of Linz

Michael E. Kummer
ZEW and CDSE, Mannheim

Rudolf Winter-Ebmer
University of Linz & IHS, Vienna

Christine Zulehner
University of Linz & WIFO, Vienna

December 22, 2011

Abstract

We investigate the effect of market structure on market performance in the market for consumer electronics. We exploit product life cycle information to build an instrumental variable for the number of firms in a market, a variable which hitherto had to be treated as exogenous in comparable studies on seller-behavior in e-commerce.

We combine data from Austria's largest online site for price comparisons with retail data on wholesale prices provided by a major hardware producer for consumer electronics. We observe input prices of firms, and all their moves in the entry and the pricing game. Using this information for 70 digital cameras, we generate instrumental variables based on the shops' entry decisions in the past. We find that instrumenting is particularly important for estimating the effect of competition on the markup of the price leader.

JEL-Classification: L11, L13, L81, D43

Keywords: Retailing, Product Life Cycle, Market Structure, Market Performance, Markup, Price Dispersion

*Correspondence: Michael Kummer: Universitaet Mannheim, Department of Economics and ZEW, L7, 1, 68131 Mannheim, Germany, Email: Kummer@zew.de. This publication is the result of a project sponsored within the scope of the SEEK research programme which was carried out in cooperation with Johannes Kepler Universitaet (JKU) Linz, Austria and the Zentrum fuer Europaeische Wirtschaftsforschung GmbH (ZEW), Mannheim, Germany

1 Introduction

Analyzing the link between market structure and market performance is of central importance in the field of industrial organization. In particular, antitrust and regulatory authorities are interested to know how many firms it takes to sustain competition in a market. For example, the expected relation between the number of firms in the market and market outcomes such as prices or qualities is at the core of merger assessments.

We investigate the interaction between market structure and market performance in e-commerce using data from Austria's largest online site for price comparisons combined with retail data on wholesale prices provided by a major hardware producer for consumer electronics. We observe firms' prices as well as their input prices, and all their moves in the entry and the pricing game. To measure the rate at which oligopoly margins decline toward zero, we analyze how quickly the break-even price-cost margins fall as the number of market participants increases from one to two firms, two to three firms, and so on. We also look at the impact of market structure on market performance over the product life cycle, as other studies focusing on market structure find that entry has, especially at the beginning of the life cycle, a significant impact on prices.[1]

The use of data from online markets has been pioneered by the studies of Brynjolfsson and Smith (2000) and Baye et al. (2004) who used online data to study the relationship of prices and competition on online markets. They also analyzed the distribution of prices and found that price dispersion increases with the number of competitors. Since then, a large variety of issues have been studied with online data. Baye et al. (2009) have examined the Kelkoo price comparison site and noted that there is a big discontinuity in clicks at the top listed products, a result which can be explained with clearinghouse models. Both, Ellison and Ellison (2005) and Ellison and Ellison (2009) have examined the competition of Internet retailers and have identified different internet-specific firm strategies which are applied in online markets to cope with the increased price sensitivity of online markets.

The prime advantage of e-commerce is the easy availability of large amounts of data on retail-prices at very little cost for the researcher. Moreover, it is generally possible to observe all the prices and price changes of the firms and to reconstruct the sequence in which they react to each other. More than that, it is possible to obtain data for many different markets, be it books or consumer electronics. However, the researcher faces the disadvantage that he does not always observe the entire market, but only a segment, and he usually cannot tell whether a posted price has also induced a transaction or not.

Our point of departure is the study by Haynes and Thompson (2008b), which exploits data on digital cameras. Their study, which is related to a similar study by Barron et al. (2004), provides useful insights into the evolution of prices and price dispersion as a function of the market structure. Underlining the potential of research on e-commerce data, the study uses data for about 400 models of digital cameras in the US. It also takes first steps towards taking the life cycle of consumer electronics into account, even though mostly by regarding life cycle effects as a nuisance parameter. Moreover, the authors point to the problem of potentially endogenous explanatory variables and emphasize the need of adequately incorporating seller heterogeneities. They cannot do so, because they only observe the aggregate data that's

[1] Examples are Berry (1992), Campbell and Hopenhayn (2005), Carlton (1983), Davis (2006), Dunne, Roberts and Samuelson (1988), Geroski (1989), Mazzeo (2002), Seim (2006), and Toivanen and Waterson (2000, 2005).

publicly available. The present study proposes to expand on their analysis, by developing an instrument for the number of sellers in a market and to focus on the determinants of the product life cycle. Haynes and Thompson (2008a) use less detailed data to take a first step towards explaining entry and exit behavior in a shopbot. To do so, they estimate an error-correction model and show that the entry and exit into a market is correlated with a measure of lagged price-cost-margin and the number of competitors. Also, in the marketing literature Moe and Yang (2009) recently analyzed the product life cycle in e-tailing. However, like the literature in IO, their data did not allow them to take the endogeneity of entry and exit into account.

Our paper sheds light onto the question how market structure affects the functioning of a market and the price level. Most importantly we are able to take a first step to circumventing the usual difficulty instrumenting the number of competitors in a market. Clearly, even if under potentially reversed signs, these questions are also of great interest to consumers and manufacturers who wish to maximize their benefits. We find a highly significant and strong effect of the number of firms on markups. Ten additional competitors in the market reduce median markups by 0.22 percentage points and the minimum markup by 0.57 percentage points. However, accounting for the potential endogeneity of markups and the number of firms in the market, we see a substantially higher negative effect: ten additional retailers reduce the markup of the median firm by 0.85 percentage points and the markup of the cheapest firm by 1.72 percentage points.

The remainder of the paper is organized as follows. We present the theoretical predictions in Section 2 and describe the data as well as the empirical strategy in 3. We discuss our estimation results in Section 4 and conclude with Section 5.

2 Theoretical Predictions and Relationship of Interest

The point of departure for the present study is a series of two papers by Barron et al. (2004) and Haynes and Thompson (2008b). Both confront the predictions of competing model with data that relate the market structure to price level and price dispersion but both have to take the number of competitors in a market as exogenously given, which they themselves point out is possibly not warranted. In what follows we shall briefly summarize their discussions of the differing predictions of the competing models to be tested.

Grossly speaking, we distinguish three groups of models which allow for price dispersion and hence a violation of the law of one price: Firstly, search theoretic models (Varian (1980), Rosenthal (1980)), which successfully allow price dispersion by introducing heterogeneity in the search costs of consumers. Secondly, models of monopolistic competition (e.g.: Perloff and Salop (1985)) can account for price dispersion, when extended by introducing asymmetries across firms, such as heterogeneous producer cost or heterogeneous producer demand (cf. Barron et al. (2004)). Thirdly, Carlson and McAfee (1983) present a search theoretic model which accommodates two sources of heterogeneities by assuming a non-degenerate distribution of producers' marginal cost and heterogeneous visiting cost of the consumers. Combining these two types of heterogeneity results in somewhat different predictions about the behavior of price and price-dispersion, given an increase in the number of competitors. Finally, also a simple structure-conduct-performance model (Bain (1951)) can be tested in this context (although with somewhat vaguer predictions), as was pointed out by Haynes and Thompson (2008b).

While all these models differ significantly in their setup, they all have something to say about the impact of market structure on prices and price dispersion. Hence evidence about this relationship is important to test them and to tell which of them is most suitable to think about a market at hand. For the present purposes it suffices to skip a detailed discussion and merely provide a very brief overview over the different predictions of the models.[2]

The first group of search theoretic models (building on different consumer types, that are equipped with different search costs (e.g. Varian (1980))), predicts that an increased number of sellers results in a larger price dispersion and, somewhat against intuition, a higher average price. The second group of models with differentiated sellers and either production cost or buyer cost asymmetries would expect that a larger number of sellers is associated with a lower average price and smaller price dispersion. Thirdly, the model by Carlson and McAfee (1983) predicts that average prices would go down while price dispersion is expected to rise. According to a structure-conduct-performance model where the incumbents face the threat of entry, prices should decrease or stay equal when more firms enter the market, depending on the strength of the entry-threat. The model is somewhat silent about price dispersion.

3 Data and Empirical Strategy

Price search engine: In our analysis we use data from the largest Austrian price comparison site www.geizhals.at[3]. This platform is Austria's unchallenged market leader for comparing prices quoted by retailers of consumer electronics. For the study in this paper we use *daily* data on 70 items[4] from a major hardware manufacturer[5] which launched during the period from January 2007 till December 2008[6]. We define a camera's birth by it's appearance on geizhals.at. The cameras were on offer at up to 212 sellers from Austria and Germany.

Available Data: For time t (measured in days) we observe for each product i and retailer j the $price_{ijt}$, the $shipping\ cost_{ijt}$ posted at the website[7] and the $availability_{ijt}$ of the product[8]. Additionally, we observe the customers' referral request ($clicks_{ijt}$) from the geizhals.at website to the retailers' e-commerce website as proxy for the consumers' demand. Customers have the possibility to evaluate the *(service)quality* of the firms on a 5-point scale the average of which is listed together with the price information on geizhals.at. *Wholesale prices* for each product i at time t were obtained from the Austrian representative of the international manufacturer. We do not claim, that these wholesale prices correspond perfectly to the retailers' marginal

[2] For a more detailed discussion of the models the interested reader is referred to the two papers by Barron et al. (2004) and Haynes and Thompson (2008b) our work builds on or to the original papers. Their discussion is clever, concise and insightful, but repeating it here would not add any further insights.

[3] Based on this data set Dulleck et al. (2011) analyze the search and purchasing behavior of buyers in which the reliability of the retailer gets more important the closer it comes to actual buying decisions.

[4] digital single lens reflex cameras (DSLR), single lens reflex cameras (SLR) and laser printers

[5] The hardware manufacturer is a multinational corporation specialized in manufacturing of electronic equipment in several areas. The manufacturer asked to keep his name anonymous. If somebody wants to check the validity of our results we can offer more detailed information on the manufacturer.

[6] For our instrumentation strategy we will use also the product life cycle of cameras entering the market starting from May 2006.

[7] Shipping cost is the only variable which has to be parsed from a text field. We use the information on cash in advance for shipping to Germany, which is the type of shipping cost most widely quoted by the shops. Missing shipping cost are imputed with the mean shipping cost by the other retailers.

[8] If the product is available immediately or at short notice the dummy is 1, if the product is available only within 2-4 days it is 0.

cost. Even though the manufacturer's distribution policy indicates that the retailers should be served by the local representative it may happen, that single retailers procure commodities for instance from the Asian market. Moreover, the local representative might offer special promotions including lower wholesale prices in exceptional cases (e.g. if a retailer commits to promote the manufacturers good in a special way). Finally, it has to be mentioned that besides the wholesale price the retailers in e-commerce might have additional cost each time they order. Despite the fact that we cannot be entirely sure that each and every e-commerce retailer is always ordering at the wholesale prices we were provided with, they are a very good proxy for the actual marginal cost the retailers' are confronted with[9]. $Price_{ijt}$ and $wholesale\ price_{it}$ were used to calculate the firms' $markup_{ijt}$ (using the Lerner index) and the markets' $price\ dispersion_{it}$.

Organization of data: We reorganized the data in a way so that the product life cycles of all digicams start at the same day 1. Hence, we have shifted the product life cycles of the digicams so that we can analyze the impact of market structure on markup and price dispersion in a cross section of 70 product life cycles. This reorganization of data is also important to guarantee that observations are iid. Especially the independence assumption is crucial as listing decisions of e-commerce traders are strategic variables: If we studied product cycles in real time the decision to list digicam X might be related to the listing decision of the follower model Y. By shifting the product life cycles to identical starting points the iid assumption concerning our data structure is valid. We define the end of a product life cycle if the amount of referral requests diminish to less than 500 remaining *clicks*. Finally, we collapse the data in order to create a panel with products as units of observation and thus obtain a daily unbalanced panel with information on the *products' age*, the *number of firms*, *average markups*, *markup of the price leader*, different measures for *price-dispersion*, and the *number of clicks*.

Descriptives: Table 1 contains summary statistics of the collapsed two-dimensional panel-data. Each observation in the descriptives refer to a single product i at a given day t in the product life cycle. We will use the markup (=Lerner Index) and the price dispersion as endogenous variables. Whereas the median markup amounts to 17.8% on average, the mean markup for price leaders is only 4.6 %. These numbers are of comparable size as in Ellison and Snyder (2011), who report an average markup of 4% for memory modules on Pricewatch.com. We use different measures for the price dispersion: the coefficient of variation and the standard deviation of the distribution of prices, as well as the absolute price gap between the price leader and the second cheapest price. The absolute price gap varies between 0 and 515.9 Euro with a mean of 11 Euro. On average a product life cycle amounts to 166.7 days with a mean of 104.3 firms which are offering the digicams. Visual inspection of the data (see Figure 2) shows that the estimated markup declines with age, and, more importantly, as the number of firms increases (each observation again corresponds to the data of a single product i at a given day t). However, this pattern is by no means very abrupt, as one might expect in perfectly transparent e-commerce markets. We rather observe a well positive average markup, also with 70 and more firms in the market. In the top left panel, the $median\ markup_{it}$ is scattered against the $number\ of\ firms_{it}$ in the corresponding market and the top right panel shows the average. The number of firms ranges from 0 to slightly more than 200 and the median

[9] According to the Austrian distributor the Austrian and German list of wholesale prices are almost identical. Note the manufacturer's incentive to keep cross-border sales between distributors and retailers as low as possible if the manufacturer were to pursue substantial price discrimination between countries.

markup ranges from 0% to 35.2%. It must be noted that we also observe negative markups especially for the minimum price firms - the average markup of price leaders is 4.8% with a standard deviation of 7.9%. In our dataset we observe for 26.92 % of all best price offers negative markups. This is in line with Ellison and Snyder (2011) who report also a substantial amount of price offers with negative markups for Pricewatch.com. Negative markups might have several possible causes: They might simply point to sell-outs after overstocking, it might be a hint to cases where retailers are not procuring via the official retail channels but exploit price differentials with, for instance, Asian markets. Finally, loss leader strategies might be responsible for negative markups where a digicam is offered at a price below marginal cost in order to attract new customers or to make profits with complementary goods. In the middle row the median mark-up is plotted against the age of the product. Again the markets' median markups fall on average with the duration of the product life cycle. If the product is available the retailer has the incentive to raise the price to the rivals's level. In the lower row, the median markup is plotted against the age of the camera (in months). We typically observe a camera between 7-8 and 15 months. While the line for the averages looks very smooth, the scatter plots on the left of the graphs reveal however, that there is large heterogeneity. Apparently there are three types of digicams: Some appear to be listed by fewer shops (20 and 60, respectively) and then to be taken off the market sooner, whereas another group of cameras seems to be listed by roughly 150 shops on average and then to be taken off the market only after 14 months. The apparent segregation of markets is striking. As expected we observe rather fast market entry within the first two months - after that the amount of firms stagnates. Summarizing the descriptive results it can be stated that markup declines very slowly, given that firms have to compete in prices in this market. Secondly, the life cycle of digital products is short enough to not only allow observing their entire lifecycle, but also observe many thereof, which is the feature our instrumentation strategy shall build on.

Empirical Strategy: In order to estimate the impact of market structure on markups and price dispersion, we estimate the following fixed-effects regression as our baseline model:

$$markup = \alpha_j + \alpha_1 * age + \alpha_2 * age^2 + \beta_1 * numfirms + \beta_2 * (numfirms)^2 + \epsilon_{jt}$$

Dependent variables *depvars* are the minimum markup, price dispersion (measured as the coefficient of variation) and the markup of the median-price and we regress each of them separately on the number of firms in the market that day. Moreover, we include a quadratic age-trend and thus measure life-cycle effects as a byproduct.[10] However, before we can do so we have to account for the fact, that it is very easy to list and unlist an item. Hence the number of sellers can react extremely fast on the market characteristics like markups or price dispersion.

Sources of Endogeneity: In all markets - but in particular in an e-tailing shopbot market - it is important to treat market structure as endogenous: due to simple and low-cost market entry and exit, e-tailers can easily adapt to changing circumstances by listing a particular product.

Generally, we are worried about two sources of potential endogeneitiy: i.) heterogeneity of the cameras in the market (quality, design-features etc.) and ii.) the shops' ability to quickly react to rising or falling markups of to be earned with a specific model by simply moving in

[10]In all the estimations we included month dummies to account for seasonality effects.

and out of the respective markets. Since our design is based on fixed effects estimation, we are less concerned about the heterogeneity of products. However, we still have to worry about the second source of endogeneity, which is due to the simultaneous determination of the markups and the number of firms.

If, for example, some unobserved factor temporarily drives up markups for some item, shops which did not sell the item before, might move into this market. Thus, we would expect to observe more shops in markets where higher markups can be reaped and vice verca. This, in turn, might result in the econometrician estimating a less negative relationship of interest than actually appropriate. This is easily illustrated by looking at the system in which markups and the number of firms are simultaneously determined, and where we are interested in the first of the following two relationships.[11]

$$markup_{jt} = \delta_j + \alpha_1 * numfirm_{jt} + \beta_1 Z_1 + \varepsilon_{1,jt} \tag{1}$$

$$numfirm_{jt} = \gamma_j + \alpha_2 * markup_{jt} + \beta_2 Z_2 + \varepsilon_{2,jt} \tag{2}$$

Substitution of the second into the first equation and rearranging, gives the reduced form for the number of firms:

$$numfirm_{jt} = \pi_j + \pi_{21} Z_1 + \pi_{22} Z_2 + u_{jt} \tag{3}$$

with $\phi := \frac{1}{1-\alpha_2\alpha_1}$, $\pi_j = \phi(\gamma_j + \alpha_2\delta_j)$, $\pi_{21} = \phi\alpha_2\beta_1$, $\pi_{22} = \phi\beta_2$ and finally $u_{jt} = \phi(\alpha_2\varepsilon_1 + \varepsilon_2)$.

In equation (1) the issue is whether $numfirm_{jt}$ and $\varepsilon_{1,jt}$ might be correlated due to simultaneity (z_1 and ε_1 are by assumption uncorrelated). Yet, from equation (3), it is easy to see, that Cov($numfirm_{jt}$,$\varepsilon_{1,jt}$) is typically not 0, since u_{jt} is a linear function of $\varepsilon_{1,jt}$. Unless $\alpha_2 = 0$, this tells us that estimation of equation (1) by OLS will result in biased and inconsistent estimates of the α_1 and β_1. More precisely we have:

$$Cov(numfirm_{jt}, \varepsilon_{1,jt}) = \frac{\alpha_2}{1-\alpha_2\alpha_1} * Var(\varepsilon_1) \tag{4}$$

and hence regressing *markup* on *numfirm* will return the biased estimate:

$$\widehat{\alpha_{1,OLS}} = \alpha_1 + \frac{Cov(numfirm_{jt}, \varepsilon_{1,jt})}{Var(numfirm_{jt})} = \alpha_1 + \frac{\alpha_2}{1-\alpha_2\alpha_1} * \frac{Var(\varepsilon_1)}{Var(numfirm_{jt})}$$

Closer examination of this result reveals, that the bias $\frac{\alpha_2}{1-\alpha_2\alpha_1} * \frac{Var(\varepsilon_1)}{Var(numfirm_{jt})}$ is positive, whenever $\alpha_2 > 0$ (more firms enter, when markups are high) and $\alpha_1 < 0$ (markups decrease when the number of firms increases). Therefore, if these conditions are satisfied, we would expect an upward bias of the OLS-coefficient of *numfirm*.

[11] Note that in the illustration we only look at a linear regressor to be instrumented and neglect the quadratic term. This corresponds to the estimations we show in Columns (1) and (2) of tables 3 and 4.

Instrumentation strategy: In order to cope with this endogeneity problem we follow an IV-approach and instrument the number of firms. For that purpose we exploit the fact that in the full Geizhals.at data, we observe the complete lifecycle of many products of the same producer and that they were launched in different points in time. For markets with brand names a part of the listing decisions can be explained by common patterns, such as established supply-relationship shops might have with a producer or a wholesale importer, or variations in the availability. These patterns remain the same over time and they are not influenced by contemporary fluctuations on a specific camera. We exploit this fact in order to devise an instrument for the number of firms, based on the shops' behavior in the markets of previously introduced products. In other words, we use the timing of listing decisions of e-tailers for brand products of our manufacturer in the past as an instrument for current listing decisions. Clearly, such past decisions - in particular if they come from different markets (e.g. digital cameras versus computer products) will be relevant predictors of the listing decisions today. On the other hand, the instrument is based on the exclusion restriction, that the shops, when taking their listing decisions in the past, did *not* take into consideration the current, then future, digital cameras.

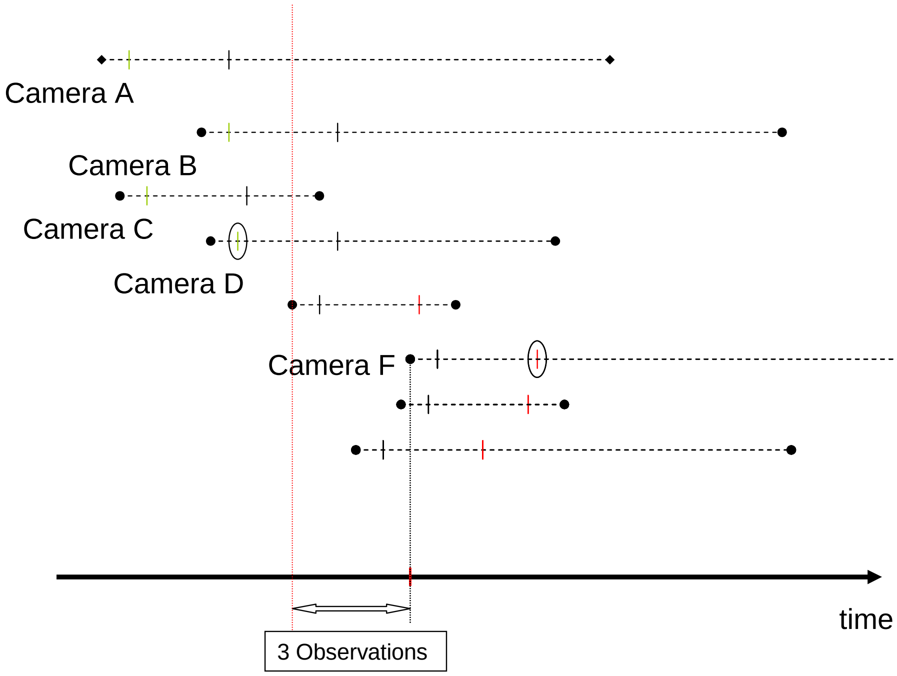
The *instrumentation strategy* is illustrated with an example in Figure 1, which represents an e-tailer, which we shall call 0815 for the sake of illustration, and its listing decisions over time: let us consider whether shop 0815 will list a product F, on the 10th day after introduction. This decision is represented by the encircled red line on item F. We predict the probability of this event by the shop 0815's general probability of listing a similar item that has been on the market for 10 days. Considering only the items that saw light *before* product F was introduced we first fix the number of such items we wish to consider (to three, here). Then we can calculate how many of those items, were listed by shop 0815 on the tenth day after they appeared and taking the share gives us a readily available estimate of shop 0815's probability to list product F on its 10th day of existence.

This method can easily be extended to the first, second, third, twentieth, etc. day. Thus, simply calculating the share of products listed on a given day in earlier life cycles, we obtain an estimate of shop 0815's probability to list an item (F) on it's first, second, tenth, etc. day of existence. Thus, repeating the exercise in order to predict what shop 0815 will do with Camera D on the third day (highlighted by the green encircled dash), we look at the "average decision" for products A, B, and C on the third day. Note, that we ignore F and its predecessors when instrumenting D, because they were introduced *after* D. Also note, that for instrumenting F we ignored A - D, because we had decided that those cameras lay too far in the past of F.[12] In a last step, we aggregate these probabilities across shops to obtain the predictor of the number of shops that will offer item F on a given day. [13]

[12] N.B.: When calculating the instrument, we fixed the number of earlier cameras to 3. This number can be varied by the researcher, depending on how fine grained the instrument needs to be and the number of available products. As fixing the number of cameras to three resulted in a reasonably strong instrument, we favored this specification over one, where we have fewer data available in the estimation. However, we also tried an alternative specification with 5 cameras, which resulted in very similar estimates. Moreover, rather than fixing the number of cameras, the researcher might prefer to fix a time-period (e.g. 6 months before F was introduced) and include all cameras that saw light during that period. The latter way of framing the instrumentation strategy does not ensure however, that the same number of products is used for calculating the shares needed for the instrument. Therefore to obtain valid standard errors additional bootstrapping - methods will be needed.

[13] Note that the method we suggest can be used even if only market-level data are available, simply by taking the average of the number of firms that listed the earlier product on the given day.

Figure 1: Instrument uses firm's listing behavior in earlier lifecycles



NOTES: If we want to predict how many shops will list a product on any given day q after introduction we use a shop's general probability of listing one of the three items that entered the market before product j, q days after they were introduced. Examples: To predict listing behavior for Camera D on day 3 (encircled green dash), we would use information on Cameras A, B, and C on their respective third days of existence (highlighted green dashes). However, we would not use the information from the Cameras that saw light after D (illustrated by a black dash on day 3). Now, to predict how many shops listed camera F on day 10 (encircled red dash), we would use the information only from the cameras that saw light later than (excluding) Camera D, provided they entered the market before F (highlighted red dashes). Cameras A, B, C (black dash on day 10) and also models younger than camera F would be ignored for the computations relevant to Camera F.

First-stage regressions: As we use the time patterns of previous listing decisions in completely different markets our instrument should not have a direct causal implication on today's markups and price dispersion. Table 2 presents the first stage regressions and show that the instrument is strong enough to explain the markets actual entry decisions depicted by the number of firms at each point in time of the product life cycles. Columns (1) and (2) compare the contribution of the instrument to explaining the number of firms, and columns (3) and (4) show the contribution to its quadratic term. It is easy to see that the instrumental variables of interest are significantly different from 0 with a probability of more than 99.99%. Moreover they improve the predictive value of the model. The R^2 in the baseline regression without the instrumented number of firms (not shown in table) amounts to 0.434. Adding our instrument for the number of firms raises the R^2 by 0.0111 in column (1) from 0.434 to 0.446. For the other columns even higher marginal R^2 are computed. We also computed the F-statistics to test the null-hypothesis that the excluded instruments are irrelevant in the first-stage and they exceed the critical value of 10 substantially. With F-values well above 300 we can prove that our instruments are strong enough to explain the variation in the number of firms over the lifecycle. For the following analysis we use columns (2) and (4) of Table 2 to calculate the predicted number of firms for the second stage regressions.

4 Results

4.1 Market Structure and Market Performance

Tables 3 and 4 show our basic results for the impact of market structure on markups. These baseline specifications are parsimoniuous, as they consider only the number of firms on the market - either linearly or in quadratic terms - and the product life cycle. Moreover, to account for seasonal effects, we add an indicator variable for offers that were quoted in December[14]. All product-specific influences are covered by a product fixed-effect. Columns 1 and 3 show OLS estimations, whereas in Columns 2 and 4, our instrumental variables approach is used.

Our results indicate a highly significant and relatively strong effect of the number of firms on markups. Not accounting for the endogeneity of the number of firms and using OLS, we would estimate the effect of ten additional competitors in the market to reduce median markups by 0.22 and minimum markup by 0.57 percentage points. The reaction of the cheapest firm is significantly higher as compared to the reaction of the median firm, which might be explained by the high frequency with which prices are changed in online markets, where the cheapest price is a focus of considerable attention of both consumers and firms.

If we instrument for the number of firms, we see a substantially larger negative effect: 10 additional retailers reduce the markup of the cheapest firm by 1.72 percentage points and the markup of the median firm by 0.85 percentage points. These figures are large in economic terms considering the standard deviation of the number of firms in our sample - 57 firms. As was discussed above, OLS is likely to underestimate the true effect of an additional firm on the markup, as it does not account for the fact that attractive items also attract more firms. Again, the reaction of the cheapest firm is considerably higher as compared to the median firm.

In Columns 3 and 4 we use a quadratic specification of the number of firms: it turns out

[14]The results for this variable are omitted in the tables to follow.

that there is a solid negative - but decreasingly negative - influence of the number of retailers on markup, both for the cheapest as well as the median markup. Numerically, for the cheapest price, the negative influence of the number of firms ceases at 560, for the median markup with 340 firms. As the maximum number of firms in our sample is 203, we can safely assume that for most part of our sample, this negative relationship is a valid description. Also note that, while the median markup remains positive throughout, the minimum markup falls below 0 for very large numbers of sellers.

Looking at the impact of the product cycle on markups, the picture is not entirely clear: in all 2SLS regressions markups grow in the beginning and go down after the first few months until the end of the product life cycle. For the minimum markup the turning point is between five and six months (Columns 2 and 4 of Tables 3 and 4) which is around the mean duration of a product life cycle of 5.5 months. For the median markup the turning point is with eight to nine months slightly above the mean duration of the product life cycle.

To investigate the impact of the number of sellers on price dispersion we concentrate on the coefficient of variation (Table 5). While the OLS regressions show a somewhat negative relation between the number of firms and price dispersion, in the 2SLS results in Columns 2 and 4, we see a strong positive relationship. In the linear case, increasing the number of firms by 10 increases the coefficient of variation by 0.03. The situation is quite similar in the quadratic case (Column 4): up to a level of 120, increasing the number of firms always leads to higher price dispersion and also after that, though declining, it remains large beyond the level of 200 firms.

The combined results on markups and price dispersion are compatible with the model (Carlson and McAfee, 1983), i.e. a search theoretic model which accommodates two sources of heterogeneities by assuming a non-degenerate distribution of producers' marginal cost and heterogeneous visiting cost of the consumers. The other search theoretic models are not in line with our findings of a decreasing median markup. Models of monopolistic competition, on the other hand, predict a decreasing price dispersion, which contradicts with our findings about the coefficient of variation.

In Table 6 we present the effect of the market structure on shipping cost. The patterns are largely the same as in Tables 3 and 4. While OLS predicts a positive relationship of shipping cost and the number of firms, Columns 2 and 4 reveal a robust negative relationship with an insignificantly small quadratic term. Ten more firms actually decrease the average shipping cost in that market by 7 cent, again a large number bearing in mind the standard deviation of 57 firms.

4.2 Lifecycle Effects

In this section we investigate, whether the profit-squeezing effect of a higher number of firms is the same in different phases of the product life cycle. To do this, we extend the model to check for different effects of market structure on markups over the product cycle.

In Table 7 we estimate the baseline model and add crossterms, interacting the number of firms with age (both linearly and quadratically). For ease of interpretation of the coefficients, in Figures 3 and 4 we also plotted how markups are predicted to depend on the number of firms, separately for different stages of the product life cycle.

In these plots, each line represents a product of certain age and we plotted the curve for products right after their introduction, and after 1, 2, 3, 6 and 9 months in the market, respectively. To make the picture clearer, we concentrate for each phase of the life cycle on the typical situation concerning the number of firms.[15] Interestingly, our plots show a very consistent pattern, which appears to consist of three phases. In Figure 3 we see the pattern for minimum markups. Apart from the months in the middle, we see a clear pattern: markups decline with more firms. This pattern, however, is less pronounced during months 2 - 4 where we observe a movement to the right, with more firms entering while markups remain stable. The estimated curves for those months resembles a J shape, indicating a small negative effect. Only towards the end of the cycle (month 6 and later) competitive pressures become more pronounced and the relationship becomes clearly negative again.

Comparing the 6 curves, we can look at how markups develop over time: First markups fall drastically, but then they stabilize and even slightly recover. Figure 3 shows the minimum markup remaining which decreases below 0 within 2 months. However, after that it remains close to 0 with more or less fierce competition, depending on the lifecycle phase. For the case of median markups in Figure 4, we can see a fairly similar pattern: More firms in the market means lower median markup, particularly in the beginning and towards the end of the lifecycle. Again, there is an initial reduction of markups over the time of the life cycle, but this trend turns around after three months and then the median markups even slightly rise again.

4.3 Robustness

Several robustness checks were performed and we shall present them in what follows. First, we test the robustness of the basic results by using varying definitions of price dispersion and by using other definitions of markups, i.e. by including shipping costs into sales prices. Moreover, at the end we take account of the fact that some of the price offers attract less attention of potential buyers; we use click-weighted markups to control for this.

Our first robustness check in Table 8 concerns our definition of price dispersion. We experiment with different definitions: apart from the coefficient of variation we use the standard deviation of prices and a coefficient of variation calculated in such a way, that the prices are weighted with the number of clicks they received. All these variations show a similar pattern: increasing number of firms is first increasing, then slightly reducing price dispersion; in all cases, the turning point is above the average number of firms in the sample. Interestingly, applying the click-weighting increases the turning point even further, to 175 firms.

The next robustness check in Table 9 concerns the measurement of prices. Consumers typically pay the product price plus shipping costs. It is well-known that firms can follow specific price-setting strategies to set visible prices - the product price - very low and non-visible prices, like shipping costs, etc. relatively high (see Ellison and Ellison (2009)). In such a case, the total price including shipping costs should be used to calculate the markup of the firm. Unfortunately, we do not know "actual" shipping costs of the firms, therefore, we calculate an artificial markup: product price plus announced shipping costs minus wholesale price. As there are different shipping costs possible, we concentrate on those, which are mostly observed in the data, which are shipping costs to Germany when paying cash in advance. If firms can vary their announced shipping costs, they should also react to the market structure; i.e. the number of firms in the market. In Table 8 we show that, in fact, our qualitative results

[15] We plot only in the region between the 33rd and 67th percentile of the distribution concerning firm sizes to avoid extrapolation of the polynomials.

are fairly similar, just more pronounced: both minimum and median markup decline with the number of firms and price dispersion is increasing.

We investigate further whether our results are influenced by the fact we treat all product offers symmetrically in our regressions. In particular, in questions of price dispersion researchers mistrust typically price offers which are way too high (cf. Baye et al. (2004)). This suggests to weigh price offers with the number of clicks they are receiving in order to give the low ranked - and maybe less reliable - price offers less weight.[16] When we do this in Table 10, we see our main results unchanged or reinforced.

Finally, we wanted to see whether our results are due to changes in the composition of the shops offering an item over the life cycle. Particularly the presence of larger shops or of shops who sell not only on geizhals.at, but also dispose of a brick and mortar outlet might affect the outcomes in our estimation. Therefore in Tables 11 - 13 we include the composition of shops in the regression. In these tables, Column 1 shows the quadratic 2SLS estimation from Tables 3, 4 and 6. We then add the share of firms that have the item stocked, the share of firms with low reputation, the share of low-price firms, the share of large firms and the share of shops with a brick and mortar facility. All of these shares are scaled on a range from 0 to 100, i.e. if in Table 11 the share of large firms increases by ten percent, this is associated with a drop in median markups by 0.62 percentage points.

First of all it should be noted, that the pattern that we observed in the baseline-specification, is unaffected by taking into account several measures of shop composition. There is some relatively small variation in the size of the coefficients and for shipping costs some effects are no longer significant when controlling for the composition of shops. Yet, grosso modo the patterns remain largely unaffected. Secondly also the estimated coefficients for the shop composition's impact on the markups are of some interest. In all the tables, markups and shipping cost fall as the share of larger firms (Column 5) and the share of firms with low reputation (Column 3) increases. The decrease of markups is more pronounced for the size of the firms. The pattern is somewhat ambiguous for the share of low price firms (Column 4). While a higher share of low price firms slightly decreases median markups and shipping cost, the estimation for the minimum markup appears to indicate otherwise. Even though accompanied by a greater coefficient for the effect of the number of firms the coefficient for the share of low-price firms is positive, which is somewhat counterintuitive. Finally, the share of firms which have the item on stock (Column 2) and the share of firms which also have a brick and mortar facility (Column 6) is related to an increase in both markups and shipping cost.

5 Conclusions

In this paper we estimate the effect of market structure on market performance in e-commerce. As endogeneity of market structure and market performance might be an issue, we use the behavior of sellers in high frequency product life cycles to develop an instrumental variable for the number of firms in a market. To analyze the effect of market structure, we use data on 70 different digital cameras and find that an increase in the number of sellers in a market by 10 reduces the mark up of the price-leader by 1.72 percentage points and that of the median firm by 0.85 percentage points. While we also find negative correlations between market structure and performance using OLS regressions, we show, that these coefficients are likely to be biased

[16] Note that the third column of this table is the same as in Table 8 as it was relevant to both questions.

upwards, due to simultaneity. To correct for this problem we propose an instrumental variable strategy, which is based on regularities in listing behavior over the product life cycle. Indeed, our results, which allow for a causal interpretation, are stronger. Moreover, we find a positive effect of the number of firms on the coefficient of variation of prices.

When we differentiate market structure effects over the full life cycle of a product, we find a negative impact especially in the beginning and the late phases of the life cycle. Moreover, we find somewhat diminished effects in the stage that corresponds to an age of 2-4 months. Our results refer to e-tailing in the presence of a price-search engine with very narrowly defined products. In such a situation, consumers can very easily collect information about prices and reliability of the sellers. Still, it takes a large number of sellers and a relatively long time until markups of firms dissipate.

The markup of the price-leader diminishes as well over the life cycle of the product. If we evaluate our results at sample means we can compare the competitive effect of more firms to the effect of time: having one more firm in the market reduces the mark up of the price leader by the same amount as three additional weeks in the product life cycle. In other words: By waiting three more weeks a consumer will get the same price reduction she would get if she went to a market with one additional firm.

References

Bain, J.S., "Relation of profit rate to industry concentration: American manufacturing, 1936-1940," *The Quarterly Journal of Economics*, 1951, *65* (3), 293–324.

Barron, J.M., B.A. Taylor, and J.R. Umbeck, "Number of sellers, average prices, and price dispersion," *International Journal of Industrial Organization*, 2004, *22* (8-9), 1041–1066.

Baye, M.R., J. Morgan, and P. Scholten, "Price dispersion in the small and in the large: Evidence from an internet price comparison site," *The Journal of Industrial Economics*, 2004, *52* (4), 463–496.

_ , J.R.J. Gatti, P. Kattuman, and J. Morgan, "Clicks, discontinuities, and firm demand online," *Journal of Economics & Management Strategy*, 2009, *18* (4), 935–975.

Berry, Steven T., "Estimation of a model of entry in the airline industry," *Econometrica*, 1992, *60* (4), 889–905.

Brynjolfsson, E. and M.D. Smith, "Frictionless commerce? A comparison of Internet and conventional retailers," *Management Science*, 2000, *46* (4), 563–585.

Campbell, J. R. and H. A. Hopenhayn, "Market Size Matters," *Journal of Industrial Economics*, 2005, *53*, 1–25.

Carlson, J.A. and R.P. McAfee, "Discrete equilibrium price dispersion," *The Journal of Political Economy*, 1983, *91* (3), 480–493.

Carlton, D. W., "The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables," *Review of Economics and Statistics*, 1983, *63*, 440–449.

Davis, Peter, "Spatial Competition in Retail Markets: Movie Theaters," *Rand Journal of Economics*, 2006, *forthcoming.*

Dulleck, Uwe, Franz Hackl, Bernhard Weiss, and Rudolf Winter-Ebmer, "Buying Online: An Analysis of Shopbot Visitors," *German Economic Review*, 2011, *12* (4), 395–408.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson, "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries," *Rand Journal of Economics*, 1988, *19* (4), 495–515.

Ellison, G. and S.F. Ellison, "Lessons about Markets from the Internet," *Journal of Economic Perspectives*, 2005, *19* (2), 139–158.

_ and _ , "Search, obfuscation, and price elasticities on the internet," *Econometrica*, 2009, *77* (2), 427–452.

Ellison, S.F. and Ch.M. Snyder, "An empirical study of pricing strategies in an online market with high frequency price information," *MIT, Department of Economics, Working Paper 11-13*, 2011.

Geroski, Paul A., "The Effect of Entry on Profit Margins in the Short and Long Run," *Annales d'Economie et de Statistique*, 1989, *15-16*, 333–353.

Haynes, M. and S. Thompson, "Entry and Exit Behavior at a Shopbot: E-sellers as Kirznerian Entrepreneurs," 2008.

_ **and** _ , "Price, price dispersion and number of sellers at a low entry cost shopbot," *International journal of industrial organization*, 2008, *26* (2), 459–472.

Mazzeo, M.J., "Product choice and oligopoly market structure," *RAND Journal of Economics*, 2002, *33* (2), 221–242.

Moe, W.W. and S. Yang, "Inertial Disruption: The Impact of a New Competitive Entrant on Online Consumer Search," *Journal of Marketing*, 2009, *73* (1), 109–121.

Perloff, J.M. and S.C. Salop, "Equilibrium with product differentiation," *The Review of Economic Studies*, 1985, *52* (1), 107.

Rosenthal, R.W., "A Model in which an Increase in the Number of Sellers Leads to a Higher Price," *Econometrica: Journal of the Econometric Society*, 1980, pp. 1575–1579.

Seim, K., "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics*, 2006, *37* (3), 619–640.

Toivanen, Otto and Michael Waterson, "Empirical Research on Discrete Choice Game Theory Models of Entry: An Illustration," *European Economic Review*, 2000, *44*, 985–992.

_ **and** _ , "Market Structure and Entry: Where's the Beef?," *Rand Journal of Economics*, 2005, *36*, 680–699.

Varian, H.R., "A model of sales," *The American Economic Review*, 1980, pp. 651–659.

Table 1: Summary statistics of the collapsed two-dimensional panel-data with the info on the level of goods and time.

Variable	Mean	Std. Dev.	Min.	Max.	N
average price in EUR	948.8	1398.7	99.1	7864.3	15893
median price in EUR	938.4	1386.8	98	7990	15893
minimum price in EUR	853.1	1293.5	78	7084.6	15893
number of sellers	104.3	57.5	1	203	15893
age in days	166.7	111.5	1	450	15893
wholesale price in EUR	764.2	1113.3	79	5801.4	15893
indicator: clicks for product i exist at t	0.9	0.3	0	1	15893
aggregate clicks at product i	27.3	40	0	646	15893
average clicks per shop offering product i	0.3	0.7	0	20.5	15893
markup of price-leader in %	4.8	7.9	-28.2	35.2	15893
median markup for product i in %	17.8	3	0	35.2	15893
markup of price leader incl. shipping cost*) in EUR	7.8	6.8	-22.4	36.1	15511
median markup incl. shipping cost*) in EUR	19.3	3.7	-3.8	36.1	15511
coefficient of variation of the prices	0.1	0.2	0	5	15801
standard deviation of prices in EUR	67.1	169.4	0	6050.7	15801
coefficient of variation of prices incl. shipping cost*)	0.1	0.2	0	4.9	15186
absolute price gap between best and second price in EUR	11	26.8	0	515.9	15801
average shipping cost in EUR	7.7	1.5	0	23.6	15511
average reputation on a scale from 1.0 (best) to 5.0 (worst)	1.7	0.2	1.1	3.5	15810
average availability (1: on stock, 2: within 2-4 days)	1.5	0.1	1	2	15208
share of German shops	0.7	0.1	0	1	15893
share of shops with brick and mortar facility	0.5	0.1	0	1	15872
clickweighted markup of price-leader in %	4.9	8	-28.2	44.3	14401
clickweighted median markup for product i in %	9.1	7.4	-28.2	96.4	14401
clickweighted coefficient of variation of the prices	0.1	0.1	0	4.1	13639

NOTES: The unit of observation is product i at time t. (product-time panel). The time-variable are days since market introduction. *) Markups and the measures for price dispersion including shipping cost refer to the gross price including the cash-in-advance fee for shipping cost the customers have to pay.

Table 2: First stage regressions for instrumenting the number of firms

VARIABLES	(1) # of firms/10	(2) # of firms/10	(3) (# of firms/10)^2	(4) (# of firms/10)^2
instrument for # of firms/10	0.20***	0.89***	0.92***	14.62***
	(0.011)	(0.027)	(0.224)	(0.522)
instrument for (# of firms/10)^2		-0.05***		-0.98***
		(0.002)		(0.034)
age (months)	2.00***	1.83***	39.41***	36.03***
	(0.026)	(0.026)	(0.499)	(0.500)
age^2	-0.12***	-0.11***	-2.44***	-2.21***
	(0.002)	(0.002)	(0.034)	(0.034)
constant	3.31***	2.22***	25.77***	3.83**
	(0.071)	(0.080)	(1.396)	(1.558)
observations	15,893	15,893	15,893	15,893
marginal R^2	0.0111	0.0373	0.0007	0.0314
products included	70	70	70	70
F test (all $u_i = 0$)	404.7	407.2	365.0	372.9
Log Likelihood	-37738	-37354	-84968	-84560
DoF (model)	73	74	73	74
rss	107452	102384	4.100e+07	3.890e+07

NOTES: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; The table shows the first stage regressions for instrumenting the number of firms by the instrumental variable that is based on listing behavior of shops over the life-cycles in earlier product markets. Columns (1) and (2) show the first stage regressions for instrumenting the number of firms. Columns (3) and (4) show the corresponding estimation for instrumenting the *square* of the number of firms. Note that the coefficients are much larger in these columns, since the variance of the predicted variable is much larger after taking the square. The R^2 of the baseline regression without the instrument (not included in the table) amounts to 0.434. The F-Statistics amount to 317,86 for testing Column (1) against the baseline model and to 833,76 for testing Column (4) against Column (3).

Table 3: Minimum markup

VARIABLES	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS
# of firms/10	-0.5729***	-1.7157***	-0.9611***	-1.7876***
	(0.012)	(0.109)	(0.030)	(0.132)
(# of firms/10)^2			0.0219***	0.0160**
			(0.002)	(0.008)
age (months)	-1.2275***	1.3248***	-1.2477***	0.8375***
	(0.044)	(0.247)	(0.044)	(0.138)
age^2	0.0323***	-0.1193***	0.0355***	-0.0889***
	(0.003)	(0.015)	(0.003)	(0.009)
constant	16.1911***	20.6920***	17.0929***	20.5172***
	(0.108)	(0.446)	(0.126)	(0.352)
observations	15,893	15,893	15,893	15,893
R^2	0.495		0.501	
products included	70	70	70	70

NOTES: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Dependent Variable: Minimum markup. The table shows the results form the fixed-effects panel regressions. Columns 1 and 3 show the estimates from the OLS; Columns 2 and 4 show the results for 2SLS panel regressions that use the instrumental variable to account for the endogeneity of the number of firms.

Table 4: Median markup

VARIABLES	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS
# of firms/10	-0.2220***	-0.8476***	-0.6656***	-0.9078***
	(0.007)	(0.063)	(0.018)	(0.075)
$(\#\text{ of firms}/10)^2$			0.0251***	0.0134***
			(0.001)	(0.004)
age (months)	-0.1890***	1.2080***	-0.2120***	0.8002***
	(0.026)	(0.144)	(0.026)	(0.078)
age^2	0.0065***	-0.0765***	0.0102***	-0.0510***
	(0.002)	(0.009)	(0.002)	(0.005)
constant	20.8724***	23.3361***	21.9029***	23.1899***
	(0.065)	(0.259)	(0.074)	(0.199)
observations	15,893	15,893	15,893	15,893
R^2	0.145		0.182	
products included	70	70	70	70

NOTES: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Dependent Variable: Median markup. The table shows the results form the fixed-effects panel regressions. Columns 1 and 3 show the estimates from the OLS; Columns 2 and 4 show the results for 2SLS panel regressions that use the instrumental variable to account for the endogeneity of the number of firms.

Table 5: Coefficient of Variation

VARIABLES	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS
# of firms/10	-0.0034***	0.0303***	-0.0030**	0.0445***
	(0.000)	(0.004)	(0.001)	(0.005)
(# of firms/10)^2			-0.0000	-0.0021***
			(0.000)	(0.000)
age (months)	-0.0022	-0.0780***	-0.0022	-0.0238***
	(0.002)	(0.008)	(0.002)	(0.005)
age^2	0.0001	0.0046***	0.0001	0.0012***
	(0.000)	(0.001)	(0.000)	(0.000)
constant	0.1292***	-0.0034	0.1282***	-0.0032
	(0.004)	(0.015)	(0.005)	(0.014)
observations	15,801	15,801	15,801	15,801
R^2	0.008		0.008	
products included	70	70	70	70

NOTES: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Dependent Variable: Coefficient of Variation. The table shows the results form the fixed-effects panel regressions. Columns 1 and 3 show the estimates from the OLS; Columns 2 and 4 show the results for 2SLS panel regressions that use the instrumental variable to account for the endogeneity of the number of firms.

Table 6: Shipping Cost

VARIABLES	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS
# of firms/10	0.0602***	-0.0708**	0.0446***	-0.0765**
	(0.003)	(0.029)	(0.009)	(0.039)
$(\# \text{ of firms}/10)^2$			0.0008*	0.0010
			(0.000)	(0.002)
age (months)	-0.1399***	0.1248**	-0.1413***	0.0992***
	(0.012)	(0.060)	(0.012)	(0.035)
age^2	0.0025***	-0.0132***	0.0027***	-0.0115***
	(0.001)	(0.004)	(0.001)	(0.002)
constant	7.7592***	8.3759***	7.8029***	8.3693***
	(0.032)	(0.140)	(0.040)	(0.129)
observations	15,441	15,441	15,441	15,441
R^2	0.075		0.075	
products included	70	70	70	70

NOTES: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Dependent Variable: Shipping Cost. The table shows the results form the fixed-effects panel regressions. Columns 1 and 3 show the estimates from the OLS; Columns 2 and 4 show the results for 2SLS panel regressions that use the instrumental variable to account for the endogeneity of the number of firms.

Table 7: Interacting the number of firms and the product life cycle

VARIABLES	(1) minimum markup	(2) median markup	(3) coefficient of variation	(4) shipping cost
# of firms/10	-10.8075***	-10.2296***	0.0107	-13.0547
	(2.958)	(2.755)	(0.053)	(17.294)
$(\#$ of firms$/10)^2$	0.6655***	0.6676***	-0.0008	0.8443
	(0.203)	(0.189)	(0.004)	(1.114)
# of firms/10 x age	0.0595	-0.1210	0.0436***	-0.2331
	(0.312)	(0.291)	(0.006)	(0.501)
# of firms/10 x age^2	0.1849**	0.2283***	-0.0041**	0.3598
	(0.090)	(0.084)	(0.002)	(0.463)
$(\#$ of firms$/10)^2$ x age	-0.0654***	-0.0634***	-0.0022***	-0.0830
	(0.022)	(0.020)	(0.000)	(0.116)
$(\#$ of firms$/10)^2$ x age^2	-0.0049*	-0.0062**	0.0002***	-0.0105
	(0.003)	(0.003)	(0.000)	(0.013)
age (months)	11.1473**	14.0518***	-0.1866*	20.9313
	(5.220)	(4.861)	(0.096)	(27.526)
age^2	-1.5739**	-1.9207***	0.0160	-2.9772
	(0.700)	(0.652)	(0.013)	(3.877)
constant	23.7129***	25.9548***	0.1722***	15.9114
	(1.340)	(1.248)	(0.029)	(11.168)
observations	15,893	15,893	15,801	15,441
products included	70	70	70	70

NOTES: Standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$; The table shows the estimation results for interacting the number of firms with age. Fixed-effects 2SLS panel regressions, using the instrumental variable to account for the endogeneity of the number of firms. The dependent variables are displayed in the column heading.

Table 8: Alternative versions of price dispersion

VARIABLES	(1) coefficient of variation	(2) dispersion sd	(3) clw coefficient of variation
# of firms/10	0.0445***	41.2327***	0.0244***
	(0.005)	(4.832)	(0.004)
(# of firms/10)^2	-0.0021***	-1.9504***	-0.0007***
	(0.000)	(0.279)	(0.000)
age (months)	-0.0238***	-24.2846***	-0.0223***
	(0.005)	(4.390)	(0.003)
age^2	0.0012***	1.0273***	0.0012***
	(0.000)	(0.276)	(0.000)
constant	-0.0032	1.7691	-0.0396**
	(0.014)	(12.902)	(0.016)
observations	15,801	15,801	13,638
products included	70	70	70

NOTES: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; The table shows the results for alternative price dispersion measures as dependent variables. Column 1 shows the Coefficient of Variation, Column 2 Price Dispersion and Column 3 the Coefficient of Variation if prices are weighted by clicks. Fixed-effects 2SLS panel regressions, which use the instrumental variable to account for the endogeneity of the number of firms. Each Column has the dependent variable as in the column heading.

Table 9: Markup and price dispersion including shipping costs

VARIABLES	(1) min markup (de)	(2) med markup (de)	(3) coeff. of variation (de)
# of firms/10	-2.3518***	-1.4160***	0.0554***
	(0.143)	(0.111)	(0.006)
(# of firms/10)^2	0.0407***	0.0280***	-0.0020***
	(0.007)	(0.006)	(0.000)
age (months)	0.9331***	1.3266***	-0.0413***
	(0.150)	(0.116)	(0.005)
age^2	-0.0827***	-0.0825***	0.0023***
	(0.009)	(0.007)	(0.000)
constant	25.4791***	26.5884***	-0.0865***
	(0.474)	(0.366)	(0.024)
observations	15,511	15,511	15,186
products included	70	70	70

NOTES: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; The table shows the estimation results when computing markups and price dispersion based on gross prices (total of price + shipping cost). For computing gross prices we used the cost of shipping to Germany, that was charged for cash-in-advance payment. Fixed-effects 2SLS panel regressions, using the instrumental variable to account for the endogeneity of the number of firms. The dependent variables are displayed in the column heading. "min markup (de)" stands for minimum gross markup for shipping to Germany.

Table 10: Markup and price dispersion weighted by clicks

VARIABLES	(1) clw minimum markup	(2) clw median markup	(3) clw coefficient of variation
# of firms/10	-1.0814***	-0.1214	0.0244***
	(0.196)	(0.206)	(0.004)
(# of firms/10)^2)	-0.0166*	-0.0347***	-0.0007***
	(0.010)	(0.011)	(0.000)
age (months)	0.5966***	-0.0265	-0.0223***
	(0.151)	(0.159)	(0.003)
age^2	-0.0778***	-0.0470***	0.0012***
	(0.009)	(0.010)	(0.000)
constant	19.4796***	18.0235***	-0.0398**
	(0.652)	(0.686)	(0.016)
observations	14,401	14,401	13,639
products included	70	70	70

NOTES: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; The table shows estimation results when the observations are weighted by the clicks an offer managed to attract, before computing the markups/coefficient of variation. Fixed-effects 2SLS panel regressions, using the instrumental variable to account for the endogeneity of the number of firms. The dependent variables are displayed in the column heading, “clw” stands for “clickweighted”.

Table 11: Median-Markup and the composition of shops as shares

VARIABLES	(1) benchmark	(2) availability	(3) reputation	(4) price-level	(5) size	(6) brick-mortar
# of firms/10	-0.9078***	-0.9404***	-0.9089***	-0.8207***	-0.9046***	-0.8628***
	(0.075)	(0.088)	(0.075)	(0.099)	(0.071)	(0.080)
$(\#\text{ of firms}/10)^2$	0.0134***	0.0129***	0.0130***	0.0119**	0.0184***	0.0109**
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)
age (months)	0.8002***	0.8262***	0.8163***	0.7031***	0.5696***	0.8482***
	(0.078)	(0.084)	(0.078)	(0.087)	(0.067)	(0.078)
age^2	-0.0510***	-0.0526***	-0.0520***	-0.0467***	-0.0382***	-0.0543***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
share on stock		0.0123**				
		(0.005)				
share low rep			-0.0045***			
			(0.002)			
share low price				-0.0220***		
				(0.007)		
share larger shops					-0.0624***	
					(0.004)	
share brick mortar						0.0192***
						(0.003)
constant	23.1899***	22.9815***	23.4638***	23.6538***	24.3018***	22.0315***
	(0.199)	(0.138)	(0.234)	(0.113)	(0.188)	(0.352)
observations	15,893	15,893	15,893	15,893	15,893	15,872
products included	70	70	70	70	70	70

NOTES: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Fixed-effects 2SLS panel regressions, which use the instrumental variable to account for the endogeneity of the number of firms; The dependent variable is Median Markup. Column (1) repeats the regression in Column (4) of table 4. Columns (2) - (6) each add a measure for the composition of shops in market-day observation, in order to account for shop heterogeneities. While immediate availability (on-stock, Column (2)) and brick and mortar (Column (6)) are readily observable dummy variables, for each shop, Columns (3) - (5) are based on shops' behavior in the first three weeks of the sample.

Table 12: Minimum-Markup and the composition of shops as shares

VARIABLES	(1) benchmark	(2) availability	(3) reputation	(4) price-level	(5) size	(6) brick-mortar
# of firms/10	-1.7876***	-2.0164***	-1.7892***	-2.1079***	-1.7836***	-1.8079***
	(0.132)	(0.164)	(0.133)	(0.191)	(0.127)	(0.142)
$(\# \text{ of firms}/10)^2$	0.0160**	0.0125	0.0155**	0.0217**	0.0223***	0.0175**
	(0.008)	(0.008)	(0.008)	(0.009)	(0.007)	(0.008)
age (months)	0.8375***	1.0202***	0.8586***	1.1944***	0.5432***	0.8177***
	(0.138)	(0.158)	(0.139)	(0.167)	(0.121)	(0.139)
age^2	-0.0889***	-0.1002***	-0.0901***	-0.1047***	-0.0725***	-0.0875***
	(0.009)	(0.010)	(0.009)	(0.010)	(0.008)	(0.009)
share on stock		0.0860***				
		(0.010)				
share low rep			-0.0059**			
			(0.003)			
share low price				0.0808***		
				(0.014)		
share larger shops					-0.0796***	
					(0.007)	
share brick mortar						0.0004
						(0.006)
constant	20.5172***	19.0578***	20.8758***	18.8106***	21.9363***	20.5466***
	(0.352)	(0.258)	(0.414)	(0.216)	(0.337)	(0.621)
observations	15,893	15,893	15,893	15,893	15,893	15,872
products included	70	70	70	70	70	70

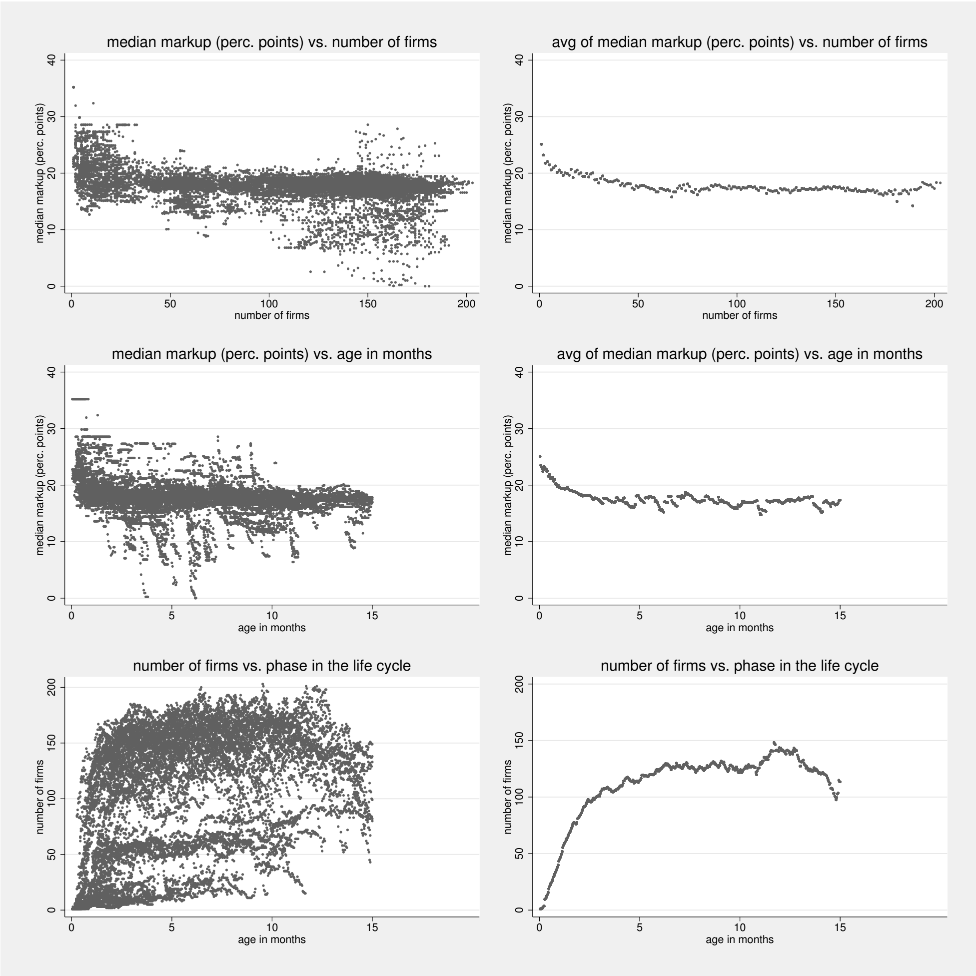
NOTES: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Fixed-effects 2SLS panel regressions, which use the instrumental variable to account for the endogeneity of the number of firms; The dependent variable is Minimum Markup. Column (1) repeats the regression in Column (4) of table 3. Columns (2) - (6) each add a measure for the composition of shops in market-day observation, in order to account for shop heterogeneities. While immediate availability (on-stock, Column (2)) and brick and mortar (Column (6)) are readily observable dummy variables, for each shop, Columns (3) - (5) are based on shops' behavior in the first three weeks of the sample.

Table 13: Shipping cost and the composition of shops as shares

VARIABLES	(1) benchmark	(2) availability	(3) reputation	(4) price-level	(5) size	(6) brick-mortar
# of firms/10	-0.0765**	-0.0763*	-0.0762*	-0.0596	-0.0986***	-0.0642
	(0.039)	(0.039)	(0.039)	(0.047)	(0.038)	(0.040)
$(\# \text{ of firms}/10)^2$	0.0010	-0.0003	0.0006	0.0005	0.0031*	-0.0008
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
age (months)	0.0992***	0.0956***	0.1133***	0.0836**	0.0321	0.1651***
	(0.035)	(0.035)	(0.035)	(0.038)	(0.032)	(0.035)
age^2	-0.0115***	-0.0115***	-0.0124***	-0.0110***	-0.0080***	-0.0157***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
share on stock		0.0129***				
		(0.002)				
share low rep			-0.0057***			
			(0.001)			
share low price				-0.0060*		
				(0.003)		
share larger shops					-0.0314***	
					(0.002)	
share brick mortar						0.0165***
						(0.001)
constant	8.3693***	7.9846***	8.7053***	8.5377***	9.0827***	7.5190***
	(0.129)	(0.080)	(0.139)	(0.071)	(0.141)	(0.182)
observations	15,441	15,441	15,441	15,441	15,441	15,423
products included	70	70	70	70	70	70

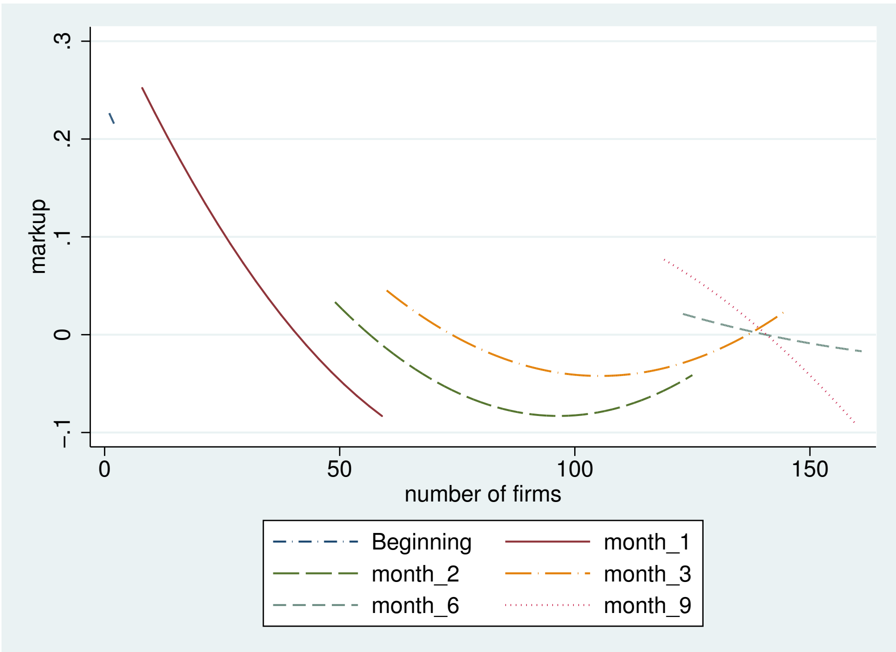
NOTES: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Fixed-effects 2SLS panel regressions, which use the instrumental variable to account for the endogeneity of the number of firms; The dependent variable is Shipping Cost. Column (1) repeats the regression in Column (4) of table 6. Columns (2) - (6) each add a measure for the composition of shops in market-day observation, in order to account for shop heterogeneities. While immediate availability (on-stock, Column (2)) and brick and mortar (Column (6)) are readily observable dummy variables, for each shop, Columns (3) - (5) are based on shops' behavior in the first three weeks of the sample.

Figure 2: Median markup plotted against the number of firms and age of product.



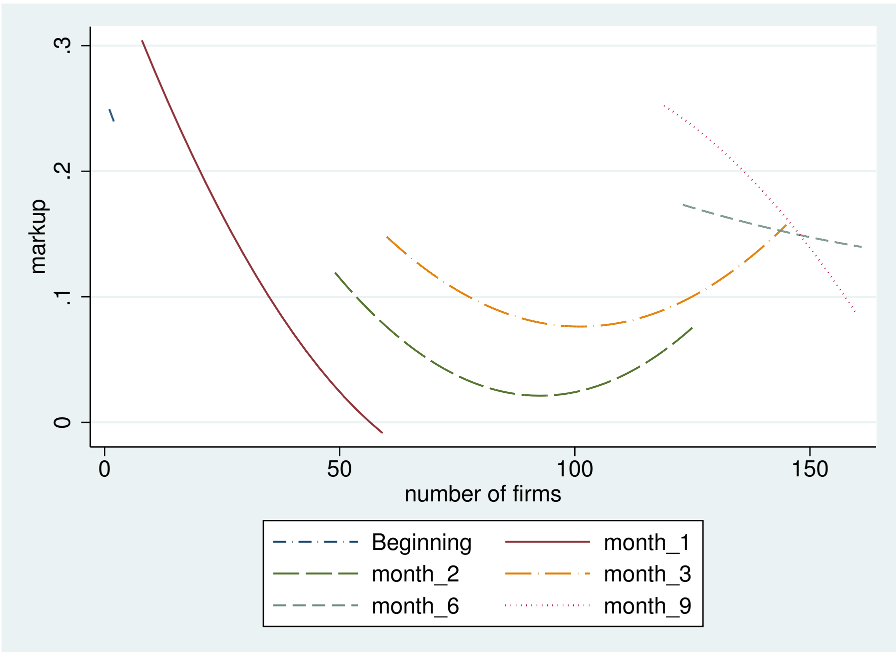
NOTES: The plot shows the empirically observed distributions of number of firms, age and median markup plotted against each other. In the top left panel, the *median markup*$_{it}$ is scattered against the *number of firms*$_{it}$ in the corresponding market and the right Column shows the corresponding averages. In the middle row the median markup is plotted against the age of the product. In the lower row, the number of firms is plotted against the age of the cameras (in months).

Figure 3: Minimum markup in different phases of the product life cycle



NOTES: The plot shows the curve estimated in table 7. Each curve shows the estimated relationship of number of firms and markup at a different point in time (right after introduction and after 1,2,3,6 and 9 months). The curves are plotted on the range from the 33rd to 67th percentile of the empirically observed distribution of the number of firms at the point in time it corresponds to.

Figure 4: Median markup in different phases of the product life cycle



NOTES: The plot shows the curve estimated in table 7. Each curve shows the estimated relationship of number of firms and markup at a different point in time (right after introduction and after 1,2,3,6 and 9 months). The curves are plotted on the range from the 33rd to 67th percentile of the empirically observed distribution of the number of firms at the point in time it corresponds to.