NET Institute*

www.NETinst.org

Working Paper #08-23

October 2008

Market Entry in E-Commerce

Maximilian Kasy
University of California, Berkeley

Michael Kummer
Johannes Keppler University

MARKET ENTRY IN E-COMMERCE -

Working Paper

Maximilian Kasy*and Michael Kummer†‡

October 14, 2008

Abstract

We analyze the behavior of start-ups in e-commerce, namely on Austria's leading price-comparison-site, a multi-product environment with almost complete information. We use weekly panel data on price-quotes of digicams, Audio/HiFi-equipment and hardware. We furthermore use advanced estimation methods, which, having only recently been introduced to IO, aim at using a minimum of modeling assumptions. Thus, being able to trace the behavior of roughly 350 start-up companies and 600 incumbents, we investigate whether start-ups have a different composition of product-portfolios, charge lower prices and offer fewer goods.

1

1 Introduction

We are interested in market entry in e-commerce and the competitive behavior of start-ups. We analyze the pricing strategies of entrants in their respective markets (competitive behavior, portfolio composition, etc.), once they have made their product choice. Our aim is to derive stylized facts concerning the strategic behavior of start-up firms in e-tailing. One aspect we are interested in is the fact, that start-ups, by definition lack customer-evaluations, the tool provided as signal for quality on Geizhals.at. In particular, we investigate whether entrants invest in low prices until they manage to acquire the signaling tools that consumers typically rely on (e.g.: a solid number of customer reviews).

We furthermore use the multiproduct nature of the data to analyze the size of start-up's product portfolio and we ask, to what extent entrants compete in more competitive markets than incumbents.

In this paper we are interested in the following main questions concerning the pricing-behavior of startups:

Q1. do the start-up-companies charge a lower price in order to generate demand and to acquire the needed customer-evaluations?

Q2. if so, do they stop charging lower prices as soon as they have acquired a certain number of evaluations?

We furthermore want to analyze the portfolio-strategies of e-tailers, which might vary between start-ups and incumbents:

Q3. do start-ups list significantly fewer products than incumbents?

Q4. do start-ups list items in markets where competition is more intense?

Our analysis of the possible differences in the portfolio and pricing strategies is inspired by a set of widely accepted "stylized facts" about entry which have long been established in the literature (cf. for example Dunne, Roberts and Samuelson [DRS89] or Mata and Portugal [MP94]). They are summarized in

Geroski [Ger95]. We revise them in the context of a multi-product e-tailing market which is characterized by an environment of almost perfect information.

Moreover, asking for the portfolio compositions of entrants is inspired by two recent theories on the patterns of firm-size-distribution. The first stems from Cabral and Mata [CM03] and is based on financing constraints of start-ups, while the second from Klepper and Thompson [KT06] explains the faster growth of small firms using the number of submarkets in which firms are active. Both of them recognize the achievements of Jovanovic's [Jov82] pioneering work, which is based on imperfect information. However, they propose additional mechanisms to explain phenomena which the earlier theory cannot account for. Thus, gaining insights on the portfolio composition of start-ups sheds light on the questions that result from this research.

To achieve these aims we describe both, the e-tailers' choice of their product-portfolio and pricing behavior and how they change over time. We do this relying on non-parametrical methods of estimation, i.e. by nonparametrically estimating the Local Average Partial Effect of experience on price (Q1 and Q2) and portfolio size (Q3), holding firm and good constant and controlling for any aggregate developments. To this end we have constructed a non-parametric kernel estimator which we deem optimally suited for the purposes of this analysis. In order to back up our estimations we run a set of more standard parametric fixed effects panel regressions which are based on more restrictive assumptions. We furthermore construct a variable which measures the intensity of competition in order to analyze whether startups compete in markets with fewer customers per competitor or not. (Q4)

Section 2 reviews related literature and provides an account of the behavioral relations this paper seeks to shed light on. In Section 3 we provide a brief description of our data while the parameters of interest, the strategy for identification and estimation, as well as a definition of local Average Partial Effects are provided in Section 4. In Section 5 we present our results and Section 6 concludes providing a few thoughts on possible avenues for further research on the topic.

2 Theoretical Outlook

2.1 Behavioral Relations of Interest

While the approach to identification is described in 4.1, this chapter aims at describing the phenomenon we wish to analyze.

2.1.1 Q1 and Q2: Do the start-up-companies charge a lower price until they have generated enough demand to acquire customer-evaluations?

While questions Q3 and Q4 involve a panel of more than 300 entrants that we observe for up to 62 weeks, in answering Q1 and Q2 we dispose of a panel indexed by firm x good identifiers as well as time. More precisely, we estimate the local Average Partial Effect of firm-age on the pricing of a startup[1], using data on 361 startups that together, list 4442 goods over the course of 62 periods. In total we use four different definitions of age, to all of which we will refer to as "age" in the following:

- A1: the effect of physical age on prices. A1 is measured in weeks and is difficult to observe for customers

- A2: the effect of the number of reviews on prices. A2 is measured in single reviews accumulated since the firm began to be active. It can be observed by every participant in the market.

- A3: the effect of the number of clicks on prices. A3 serves as a measure for success in proposing interesting offers to the customers. It is measured in 100s of clicks accumulated since the start of operation and cannot be observed by customers.

- A4: beyond the measure of clicks we also dispose of a "Last-Click-Through"

---

[1] a formal discussion of the relationship of interest is provided in section 4.1, and will thus be skipped here.

variable[2], which bears a stronger relationship with demand. (cf. Dulleck et al. [DHW08]). We use this variable to analyze the effect of "Last-Clicks-Throughs." A4 is used as an indicator of actual demand and it is measured in single clicks accumulated since operations were taken up.

Of those indicators, only the second one is known to every participant in the market. Recent investigations, for example by Reinstein and Snyder [RS05] or Duan, Gu and Whinston [DGW05], have suggested that A2, having the function of a signal, should bear a direct relationship with the demand a shop can generate and thus to the pricing policy.

2.1.2 Q3: Do Start-ups have a Smaller Product-Portfolio than Incumbents?

In order to answer this question we generated a variable which contains the number of products a firm was offering each week. Calling the resulting variable, which captures portfolio size S we proceed to analyze the size of a shop's product portfolio as a function of age, again conditioning on the specific firm i in period t

$$S = S(f, t, a) \tag{1}$$

As before the relationship of interest $\frac{\partial}{\partial a} S(f, t, a)$ is estimated by local APEs, as outlined in section 4.1.

2.1.3 Q4: Do start-ups list items with either significantly smaller total demand or with higher levels of competition?

Finally, to answer this question we look at the average number of Last-Click-Throughs (demand) per competitor across the products that the start-ups are listing in a particular week. Denoting this indicator with Cm we analyze:

$$Cm = Cm(f, t, a) \tag{2}$$

---

[2]for a description of the concept and an outline how the variable has been generated, please refer to the Appendix

We, again, estimate $\frac{\partial}{\partial a} Cm(f, t, a)$ for each firm at every "age" and averaging over all firms with the same "age" we obtain the local Average Partial Effect.

2.2 Review of the Related Literature on Market Entry and E-Commerce

We are interested in empirically investigating the behavior of start-ups in e-commerce and their investment into a credible signal for quality and reliability. There has been a wealth of publications revolving around e-commerce. Smith and Brynjolfsson [SB01] and Ellison and Ellison [EE04] study pricing behavior of e-tailers. Baye, Morgan and Scholten [BMS04] have published important results on price dispersion and the frequency of changes in prices. Furthermore, Emre, Hortacsu and Syverson [EHS05] find that the change in search cost and increase in price elasticity can encourage new entry of low cost-firms. Those studies have usually focused on single markets, used less frequent data and we are not aware of a study that is able to observe the shops' entire product portfolio over time. One of the key-strengths of this analysis is the high quality of the data we can use to trace the behavior of the start-ups on every market they enter.

Earlier related research on start-ups have been conducted by Robinson Fornell and Sullivan [RFS92)], who analyzed the characteristics of market-pioneers and found that more competitive firms have a tendency to pioneer new markets . Mata and Portugal and Guimares [MPG95] have analyzed the longevity of plants in Portugal through the 1980s. They find that the size of entry and, most of all, the post-entry growth are decisive determinants of entrants' longevity, which they take as evidence for the pioneering theory of Jovanovic [Jov82] that start-ups cannot easily predict how efficient they will be. Finally, Audretsch [Aud91] and Argawal [Arg98] explore the post-entry performance (and exit-behavior) of small start-ups under different technological regimes. Argawal concludes that a market environment characterized by high technological activity reduces entrants' chance of survival. Yet, apart from the fact that we are not knowledge-able of a study of the entry behavior of start-ups in e-commerce, our focus on

the actions and strategies of start-ups is somewhat different from the literature cited above.

The idea of the impact of customer reviews on purchase decisions has been found by several other researchers, such as Dulleck et al. [DHW08], Reinstein and Snyder [RS05], Duan, Gu and Whinston [DGW05] or Smith and Brynjolfsson [SB01]. However, asking whether start-ups are aware of this effect and thus actively consider them in their pricing strategy appears to be an additional contribution to the literature.

Moreover, and maybe most importantly, we investigate the composition of start-ups' product portfolios. Geroski [Ger95] suggests that entrants start on a small scale and take relatively long to grow bigger. He furthermore suggests that entry into a particular product market occurs in waves and correlates with exit. If that is true it appears likely to us that the portfolio of products that start-ups list is composed differently, because start-ups generally enter markets at a different point in the product life cycle. Ellison, Fudenberg and Moebius [EF03] have provided a study which is related to portfolio specialization in e-commerce.

3   The Data

For our estimations we use a data set which comes from the Geizhals database, which is currently being collected at JKU-Linz. www.geizhals.at is Austria's best known site for price comparisons, which is also widely used in Germany and which lists mainly (but not exclusively) on-line-shops based in those two countries. As of today, there are roughly 1300 shops advertising their products on the home page. Our data has been collected three categories of products, namely: Hardware, Video/Photo/TV and Audio/HiFi.[3]

The shops pay a fixed fee for being able to make any amount of offers and are then charged on the basis of how many users click on the link form geizhals.at

---

[3]In theory there are nine categories of products (Hardware, Software, Games, Video/Photo/TV, Phone etc., Audio/HiFi, DVD, Consumer Electronics and Sports) and our raw data surveys roughly 180000 goods, which can be retrieved in up to 10 intra-day-steps. However, computational limits suggest to restrict our analysis to the selection proposed above.

to the shop. We are therefore able to observe and count the clicks of the users, when they leave geizhals.at and go to the page of an e-tailer. Obviously, the clicks are not only a valuable indicator of purchases[4] but also of the advertising value of the offers. Thus, we can observe how start-ups react to the traffic on their page and adapt the composition of their product-portfolio.

Furthermore, we are able to observe every offer (=price) of each shop for a certain item, the number of competitors in an individual product-market, and a proxy for quality of both goods and the services of a shop [5].

To our knowledge the geizhals-database is the most detailed and extensive of its kind and thus, highly suitable for the analysis of the competitive behavior of start-ups in e-commerce markets. We will now move on to give a brief survey of the data set we used. Further background information on the database, which is still being at JKU-Linz is provided in the Appendix.

3.1   Structure of the Data

As indicated above, we observe 1000 companies over a period of 62 weeks (cf. table 1). The data is drawn from the categories Hardware, Video/Photo/TV and Audio/HiFi and we observe price-quotes for 21658 items in total. The data contains a firm-indicator, a good-indicator and the price of every quote. Moreover we observe the clicks and the Last Click Throughs that occured throughout the week t and which have been measured for every offer ij. Additionally we have data on the shops, which contains information on the number of reviews and the average grades it obtained on the reviews in week t. We furthermore observe the total number of clicks and last-clicks on the shop's page and the date[6] when it was active for the first (birthdate) and the last time. For the products that are being listed, we observe an identifier and an identifier of the category it belongs to. We furthermore observe how many shops were listing the item in week t and a set of benchmark-prices (lowest (tenth) price in the market).

---

[4]see Smith and Brynjolfsson [SB01] or Bai [Ba02] on the conversion rates from last clicks to actual purchases (about 50%)

[5]consumer evaluations of goods, consumer evaluations of shops

[6]dates are measured in unix-seconds

Variable	Mean	Std. Dev.	Min	Max
age (A1)	86.95	48.04	0.02	164.42
Reviews (A2)	121.07	184.76	0.00	1941.00
past clicks (A3)	899.97	1956.25	1.00	20107.00
past Last Clicks (A4)	3216.15	6894.97	0.00	60827.00
incumbent	0.84	0.37	0.00	1.00
items listed	569.97	437.78	1.00	2385.00
relative price	0.03	0.41	-8.42	7.52
absolute price	227.62	487.60	0.05	12179.00
Basic characteristics:	firms	goods	observations	periods
	965	20919	4988929	62

Table 1: all firms - characteristics

Finally, we again, observe standardized quality-reviews and total clicks for the products.

We could access the two-year-historic-records of the two years before we begin to observe price-offers, which allows us to better distinguish start-up companies and incumbents. From those records, we obtained the number of reviews and average grade of incumbents upon entering our period of observation.

3.1.1 Descriptive Statistics

For our estimations we use a sample of entrants (cf. table 3) and the sample of incumbents is used to control for overall movements in the market price. In the sample of entrants we use only firms that entered in the 62 weeks, we observe. Incumbents (cf. table 2) are identified by the fact that they were active on Geizhals.at before our period of observation begins. (i.e. before April 2006). Together we use data from 965 firms which together quote 4988929 prices. Of these 361 are entrants who list 484 items/week on average (compared to the average of 586 listings by the incumbent firms). Not surprisingly, incumbents are not only older, but also the mean for the other three definitions of experience (A2 to A4) is significantly higher for the incumbent group. However, average relative and absolute price is lower for the entrants and also the range of relative prices is slightly wider for this group.

Variable	Mean	Std. Dev	. Min	Max
age (A1)	100.01	40.56	1.24	164.42
Reviews (A2)	138.69	194.94	0.00	1941.00
past clicks (A3)	996.53	2101.25	1.00	20107.00
past Last Clicks (A4)	3612.21	7405.80	0.00	60827.00
items listed	586.36	449.97	1.00	2385.00
relative price	0.03	0.41	-7.91	7.47
absolute price	229.21	501.86	0.42	12179.00
Basic characteristics:	firms	goods	observations	periods
	604	18431	4190820	62

Table 2: incumbent characteristics

Variable	Mean	Std. Dev	Min	Max
age (A1)	18.37	13.72	0.02	60.62
Reviews (A2)	28.51	60.41	0.00	369.00
past clicks (A3)	392.96	656.93	1.00	5897.00
past Last Clicks (A4)	1136.48	2008.05	0.00	19847.00
items listed	483.89	354.99	1.00	1566.00
relative price	0.02	0.42	-8.42	7.52
absolute price	219.31	404.45	0.05	15499.00
Basic characteristics:	firms	goods	observations	periods
	361	13281	798109	62

Table 3: entrant characteristics

Table 4: Basic Descriptive Statistics of the rawdata from JKU's Geizhals-database

No. of periods (t)	*Total No. of items (j)*		*Total No. of listings (i at j)*		*No. of prices quoted (=obs.)*	
62	21658		512944		5546319	
	Firms (i)		*Start-ups*		*Incumbents*	
number	1000		395		605	
evaluations	33352		3298		29528	
exits	300		125		175	
	mean	*sd*	*mean*	*sd*	*mean*	*sd*
total No. of price quotes each week	82700.76	32984.75	14984.47	7542.31	67716.29	29166.37
total No. of goods listed in the market each week	6442.77	2634.88	3363.11	1440.87	6082.76	25545.84

note: the figures are taken from the rawdata and include obsevations which were not included in the estimations (flaws, single-period-offers, etc.)

4 The parameters of interest - Definition, Identification, Estimation and Inference

4.1 Identification of the effect of firm "age" on pricing

4.1.1 Setup and notation

For matter of clarity we will focus the following discussion on question 1 and 2 regarding the effect of age on pricing. However exactly the same arguments hold regarding questions 3 and 4, with the obvious replacement of dependent variables and conditioning only on firms as opposed to firms and goods. Our object of interest will be log prices, denoted as P, for different goods, firms and points in time. All random variables will be denoted as uppercase, all realizations as the corresponding lowercase (the distinction will be important in the interpretation of conditional expectations below).

The following notation will be used throughout:

- A: firm "age" since market entry, as measured either by calender time or sales or number of ratings received in different specifications

- T: actual calender time

- G: a good identifier

- F: a firm identifier

- D: a dummy for whether a good, as indexed by A, T, G, F is offered on Geizhals

The question of interest is the effect of firm age on pricing. Two sources of unobservability make identification difficult:

- Unobservability due to certain goods not being offered by different firms at different points in time

- Unobervability due to counterfactual firm age and calender time combinations

We will essentially define away the first problem by looking only at the local effect of "age" on pricing for the distribution of goods offered by firms of a certain age in the sample population. It is the second problem that constitutes the more fundamental difficulty, namely disentangling the effect of inflation - or rather deflation in the case of consumer electronics - from the effect of firm aging. We will have to find ways of controlling for the change in prices we would have observed in the absence of firm aging.

But first let us define more precisely our object of interest. We will take prices to be a function of A, T, G, F, i.e. $P = P(a, t, g, f)$. This includes actual, observed prices as well as counterfactual prices for "impossible" combinations of age and time for a given firm.[7] The effect of age on price, holding calender time

---

[7]For the notion of counterfactual outcomes compare for instance Holland [Hol86]

fixed(!), is then given by

$$P_a(a, t, g, f) := \frac{\partial}{\partial a} P(a, t, g, f)$$

We will now define our object of interest, the local average partial effect of age on prices, as a suitable average of these[8]:

$$g(a) = E[P_a(A, T, G, F) | A = a, D = 1]$$

In words we take the average of the effect of age on log prices over the distribution of goods offered by firms of age a in the sample universe. As mentioned before, this definition assumes away the problem of goods not being offered by some firms at some time points, by simply averaging over the appropriate distribution conditional on $D = 1$.

For simpler notation below, let

$$P_t(a, t, g, f) := \frac{\partial}{\partial t} P(a, t, g, f)$$

be the "pure inflation" effect of time on price and

$$P_e(a, t, g, f) := \frac{\partial}{\partial \epsilon} P(a + \epsilon, t + \epsilon, g, f)$$

be the "empirical" time effect. It is only this latter that is identified from the data, conditional on $P(a, t, g, f)$ being observable (in a time neighborhood). Note that

$$P_e = P_a + P_t$$

that is if we can identify the pure inflation effect for a given (a, t, g, f), then we can also identify the pure age effect since P_e is basically observable: Define

$$h(a) = E[P_t(A, T, G, F) | A = a, D = 1]$$

---

[8] For a general discussion of average structural functions and average partial effects compare [Woo05]

13

and

$$j(a) = E[P_e(A, T, G, F)|A = a, D = 1]$$

where again the latter is identified and can be estimated from the data, for instance using kernel based local linear regression.[9] Then $g(a) = j(a) - h(a)$, and we have to identify "inflation", properly weighted by the basket of goods under consideration, $h(a)$.

4.1.2 Identification via "mature" firms

Our first identification approach will use the price evolution of mature firms to estimate the counterfactual evolution of prices without an age effect - essentially assuming that for old firms there is no more systematic effect of age on pricing and that inflation is on average the same for young and old firms - and then construct the age effect for young firms as the difference between their compounded age and time effect - the actual evolution of their prices over time - and this estimate of the pure time effect.

Formally, define

$$h_1(t, g) = E\left[P_e(A, T, G, F)|T = t, G = g, A > \bar{A}, D = 1\right]$$

and

$$\hat{h}(a) = E[h_1(T, G)|A = a, D = 1]$$

where $\bar{A}$ is a chosen constant beyond which we assume there is no more systematic effect of aging on pricing. $\hat{h}$ is again identified and estimable using local linear regression. If $\hat{h} = h$, Our problem of identifying g is solved! Sufficient conditions for this to hold are given by

Assumption 1.

$$E\left[P_a(A, T, G, F)|T = t, G = g, A > \bar{A}, D = 1\right] = 0$$

---

[9]Compare for example [LR06]

14

(i.e. there is no age effect for old firms)

$$E[E\left[P_t(A,T,G,F)|T,A>\bar{A},D=1\right]|A=a,D=1]=E\left[P_t(A,T,G,F)|A=a,D=1\right]$$

4.2 Estimation via average local linear regression

For the implementation of this program we decided to go for a fairly straight-forward application of local linear regression. In particular, a weighted least squares regression of log prices, corrected for inflation according to the above argument by subtracting log average prices for old firms appropriately specified, on our various definitions of age is fitted in a neighbourhood with respect to age, holding firm and good constant, where the weights correspond to (the square root of) the Epanechnikov kernel appropriately rescaled, i.e

$$\hat{\beta}_t = \frac{1}{\sum_{f,g} Obs(f,g,t)} \sum_{f,g} \left[argmin_b \sum_{s:|s-t|<h,obs(f,g,s)} K\left(\frac{s-t}{h}\right)(P(f,g,t)-(1,s-t)b)^2 \right]$$

where $obs(f,g,t)$ is an indicator for whether the respective price is observed, and $Obs(f,g,t)$ is an indicator for whether at least two such prices were observed in an h neighborhood of t.

This is done for every firm and good and every age on a predefined grid. the result is then averaged, for a given age, over all firms and goods such that there were sufficient observations in an age neighbourhood for calculating the slope. This average is taken as an estimate of the average effect of age on pricing at the given age. Furthermore, this local average partial effect is, for presentational purposes, integrated to deliver something that can with some caveats be interpreted as average relative log price at different ages, ceteris paribus.

For a development of the general econometric theory of such local average partial effects and so called partial means more generally, in particular asymptotic normality and rates of convergence, the interested reader is referred to Newey [New94].

4.3 Confidence bands based on bootstrap

Finally, to be able to do inference of the thus estimated relations we need to get a handle on the sampling variance of our estimator. As is practice in much of the applied literature, we will use a bootstrap procedure to construct confidence bands.

Crucially, however, we need to account for dependencies between observations of prices for different goods, firms, and times. In the context of parametric panel estimation, this is usually discussed under the name of clustering, here it will imply a procedure of hierarchical bootstrapping. In order to get a handle on the nature of such dependencies, a clear notion has to be formed on the (fictitious) population of interest as well as the data generating process producing the sample from this population. We have decided to consider the process as sampling first from a population of firms, then within each firm sampling from a set of possible goods. This choice corresponds to accounting for dependencies of the price setting of a firm across goods. It does not consider, however, dependencies of goods across firms. However, to the extent that there is a multiplicative effect (i.e. additive in terms of logs) which is constant across young and old firms for a given good and point in time, this is accounted for by our procedure of considering prices net of mean log prices for old firms. Nor does this hierarchical procedure take into account dependencies of the prices of different goods and firms at a given point in calendar time. This, however, we consider as a matter of definition: we look at pricing in this particular period in history.

This said, our bootstrap procedure is implemented as drawing repeatedly from the sample of firms under consideration, then for each such firm goods are repeatedly drawn from the set of goods offered by that firm until we have as many goods (with possible repetitions) as observed for that firm. The drawing of firms and goods is repeated until we have a sample of the original size. For this "new" sample, the local linear regression estimator is calculated again. The whole is repeated a large number of times - large meaning in practice 500. From this set of estimated functions two types of confidence bands are constructed for both

the estimated local average partial effect and for its cumulative sum, namely a pointwise and a uniform confidence band. The former is constructed by taking the pointwise $\alpha/2$ and $1 - \alpha/2$ (α usually being .05) quantiles of the estimating functions. The second is then constructed by "blowing up" the first, until the appropriate uniform coverage probability is achieved. This procedure ensures fairly tight uniform confidence bands.

4.4   More conventional estimators

For matter of comparison we will also present a couple of more standard estimators to check our results and to help the reader put our methods into context. In particular, we will run a series of fixed effects regressions of normalized prizes (or one of the other dependent variables) on a polynomial in age and a set of firm/good dummies:

$$P = \alpha_{F,G} + \sum_{i=1}^{k} \beta_i A^i + \epsilon_{F,G,T}, \; \epsilon_{F,G,T} \perp A$$

The estimated relation between age and prize should look roughly like the cumulated local average partial effects from our nonparametric approach. An even more flexible specification is given if we allow the coefficients of the age polynomial to vary in an unrestricted manner with firms and goods, i.e.

$$P = \alpha_{F,G} + \sum_{i=1}^{k} \beta_{F,G,i} A^i + \epsilon_{F,G,T}, \; \epsilon_{F,G,T} \perp A$$

$$\beta_i = E[\beta_{F,G,i}]$$

We would expect the these average coefficients would asymptotically (as sample size and order of the polynomial go to infinity) deliver exactly the same age/price schedule as out average LLR estimator, if we abstract from nonrandom missingness of observations.

5 Results

We now present the results of our non-parametric description of the relationships between our four indicators of experience of start-ups and their behavior. This chapter is divided in four subsections. We first provide the local Average Partial Effects, which correspond to the pricing behavior of start-ups, before moving on to their portfolio composition (subsection 2). The third subsection repeats the estimations of the pricing strategies for the subgroup of successful start-ups, and subsection 4 finally is dedicated to the more standard parametric estimations. All of the following graphs have the same structure: Our estimates of local average partial effects together with pointwise and uniform 95 percent bootstrap confidence bands are displayed on the right, the derived - but more intuitive - cumulated partial effects are displayed on the left, normalized to be 0 at the right endpoint (at "high age"), again with pointwise and uniform confidence bands.

5.1 Pricing Policy

A general pattern is revealed by the analysis of the relationship between experience and pricing. After controlling for any form of firm/good level heterogeneity and aggregate price changes start-ups increase prices as they grow older or. In other words, set their prices at a lower level than incumbents upon entry. Furthermore, the comparison of figure 1 and 3 suggests that the turnover of a company has a stronger influence on its pricing policy than mere aging. As far as the accumulated reviews are concerned (cf. figure 2), the pattern is consistent with the other two dimensions of experience and our estimations suggest the prevalence of lower pricing in the first phase of a firms life.

Finally, accumulated LCTs (Last Click Throughs) bear a comparatively similar relationship to the pricing strategy as accumulated clicks. This can be seen in figure 4, which shows how prices increase over time, reflected by generally slightly positive local APEs.

Figure 1: Effect of Age (in weeks; A1) on Prices
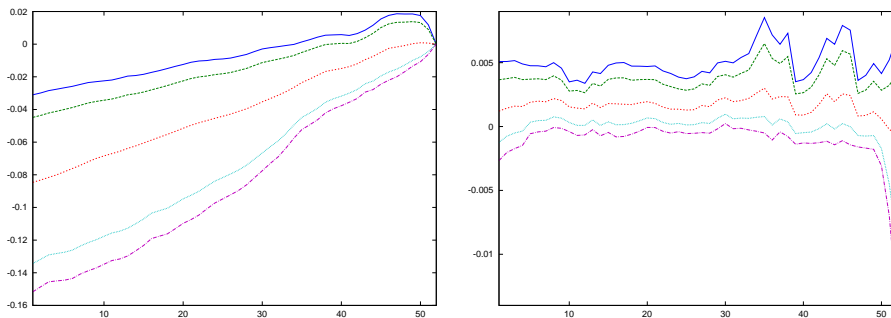
The figure shows the estimated APEs of age on prices. Left we show the effect of age on the level of prices, the right graph shows the marginal effects. Prices start out on a low level and continuously increase until the firm reaches its first year.

Figure 2: Effect of Accumulated Reviews (A2) on Prices
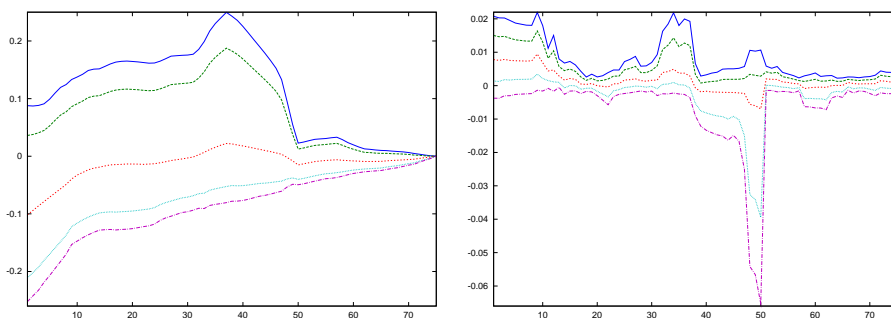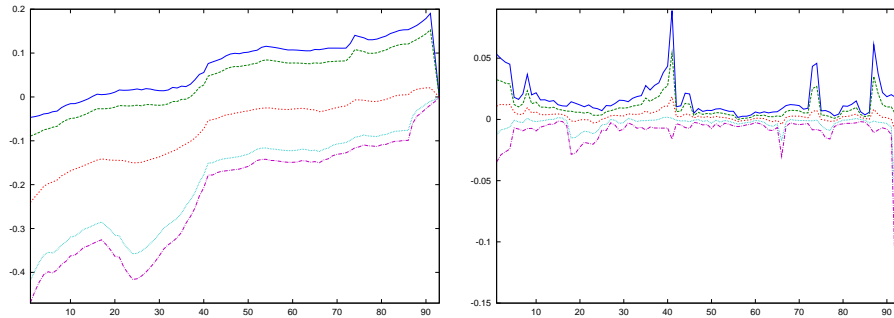

The figure shows the estimated APEs of the number of reviews on prices.

Figure 3: Effect of Accumulated Clicks (100s of clicks; A3) on Prices
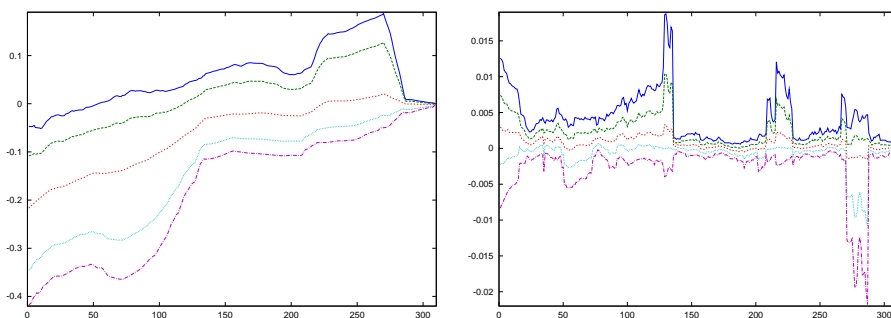
The figure shows the estimated APEs of accumulated 100s of clicks on prices. Prices start out on a low level and continuously increase until the firm reaches its first year. However, the effect is insignificant most of the time.

Figure 4: Effect of Accumulated Last Click Throughs (i.e. demand, measured in 10 clicks; A4) on prices

The figure shows the estimated APEs of accumulated 10s of LCTs on prices. LCTs largely mirror the pattern of clicks.

## 5.2 Strategic Composition of the Product Portfolio

In the following we present the description of how start-ups in the e-tailing business compose their portfolio of products and how these choices are modified as the young firms gain experience.

5.2.1 Experience and the Size of the Portfolio

The evolution of the relative size of the start-ups' portfolio as they gain age in weeks is shown in figure 5. The results are inconclusive and seem to point towards no significant change of portfolio as a firm matures.

Furthermore, the relationship between portfolio size and accumulated reviews is shown in figure 6, whereas figure 7 shows the portfolio size together with accumulated clicks. While the firms that are young in reviews show the same heterogeneity in terms of the initial size of the portfolio, we observe a significant decrease of relative portfolio size until the first reviews are accumulated. Following this decrease, the companies appear to follow the movements in the market. Finally, the evidence does not suggest a very clear pattern for the evolution of the portfolio with respect to past demand of young firms.

Figure 5: Age and Portfolio Size
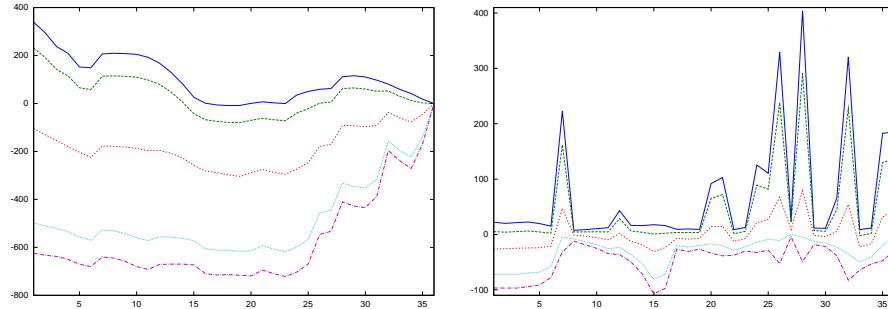
The graph shows the relationship between the age of the start-up companies and the size of their portfolio (relative to the established-firms' portfolio). The dependent variable is the number of items listed.

Figure 6: Reviews and Portfolio Size
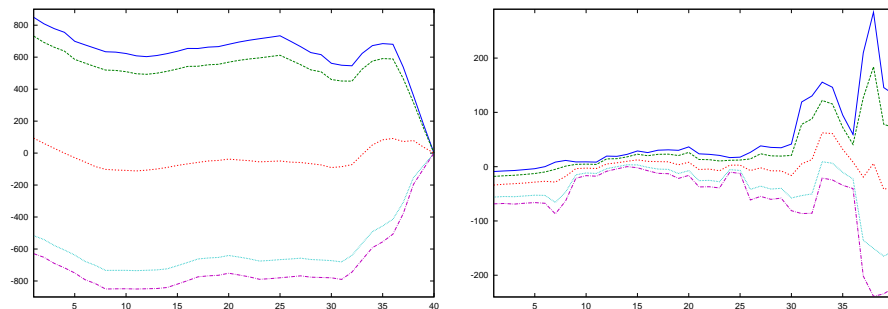
The graph shows the relationship between the accumulated reviews of the start-up companies and the relative size of their portfolio. The dependent variable is the number of items listed. The graph suggests a reduction in portfolio-size during the first phase of review accumulation.

Figure 7: Past Clicks and Portfolio Size
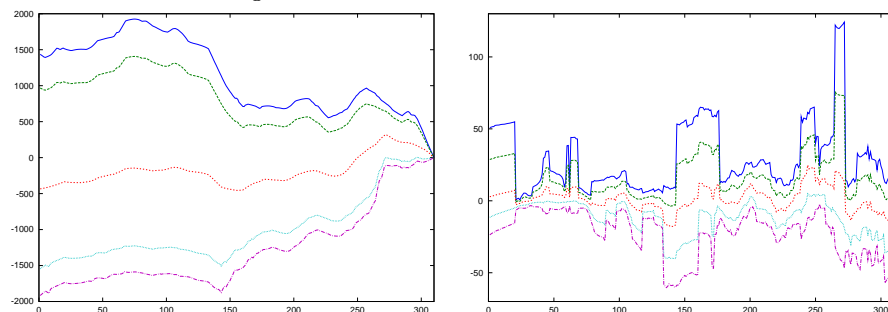
The graphs shows the relationship of clicks accumulated in the past and number of items listed in the portfolio (relative to the established-firms' portfolios).

21

5.2.2   Experience and the Intensity of Competition

We now discuss the estimation of the local APEs of experience on our variable of intensity of competition. Somewhat surprisingly, the results mirror the evolution of portfolio-size, which hints at a possible relationship between those two variables. This could indicate that the young firms manage to add products with lower intensity of competition to their portfolio, while removing products with high intensity of competition.
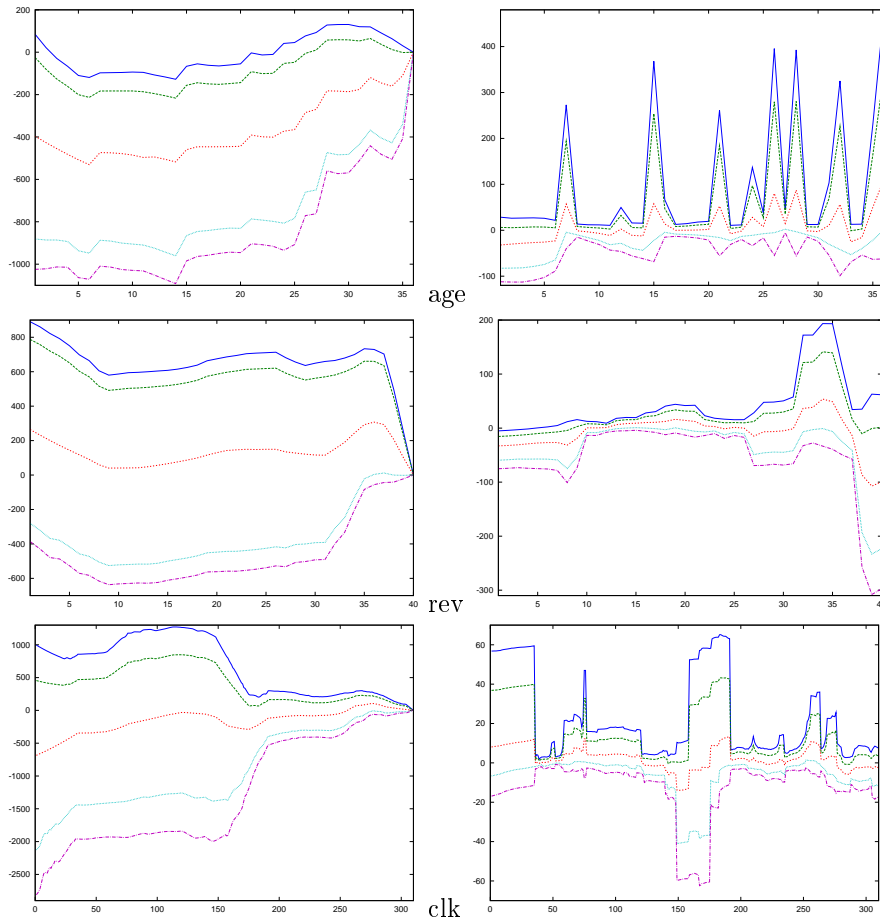
5.3   Successful Firms Only

Finally, inspired by Cabral and Mata's [CM03] analysis of surviving entrants, we were interested in the question whether the behavior of successful firms on Geizhals.at follows a different pattern. We thus reduced the sample to those start-ups which managed to acquire more than 15000 clicks or more than 20 reviews. We re-estimated the local APEs following only the life of the firms in the reduced sample and the results are shown in figure 9. Unfortunately, as a result of reduced sample size, our standard errors blow up and the "point estimate" of the age-pricing schedule seems slightly flatter than before. This might well be a result of sampling error, however, and over all we don't see any distinctive behavior for this sample of successful firms.

5.4   Standardmethodology

The following tables (cf. tables 5 to 8) contain (firm x good) fixed effects regressions of prices on our various definitions of age. Age has been normalized so that the interval covered in the corresponding graphs of our nonparametric estimates is [0,1], and log normalized prices have been multiplied by 100. As a consequence the sum of the coefficients excluding the constant is the estimated percent change of prices over the respective age intervals under consideration. The results are broadly consistent with our nonparametric estimates.

Figure 8: Experience and the Intensity of Competition

The dependent variable is average Last Clicks per Competitor. It is shown together with age (on top), accumulated reviews (middle) and accumulated past clicks (at the bottom).

Figure 9: Reduced Sample of Successful Firms

age

reviews
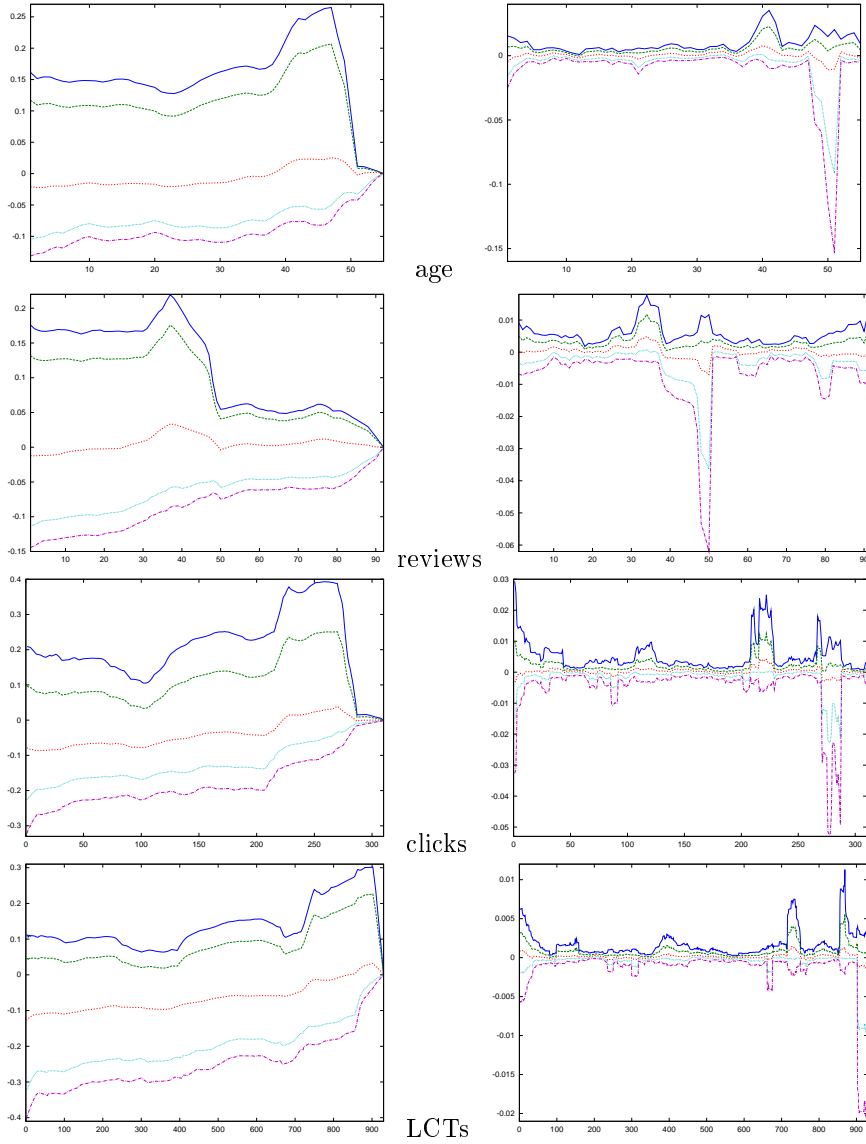
clicks

LCTs

The figure shows the repetition of the estimation above for the four measures
of experience (age, reviews, clicks and LCTs), when including only firms which
reached a minimum level of clicks and reviews.

A1	6.1802	8.0199	7.7853	-1.8298
	(1.8456)	(4.4046)	(3.1586)	(6.526)
A1sq		-2.1512		25.9474
		(4.4914)		(18.8851)
A1cb			-2.3461	-20.728
			(3.215)4	(13.5648)
Constant	0.1347	-0.1066	-0.1453	0.5711
	(0.6343)	(0.8405)	(0.8042)	(0.6971)

Table 5: StdA1

Table 5 shows the parametric estimates for the effect of age (A1) on prices. (Standard Errors are provided in brackets)

A2	3.234	5.3179	4.2976	7.6296
	(2.0257)	(5.0974)	(3.9814)	(7.8326)
A2sq		-2.5491		-9.6746
		(4.7971)		(27.2541)
A2cb			-1.6287	5.3796
			(3.9156)	(21.7127)
Constant	2.6359	2.4931	2.5447	2.3951
	(0.2945)	(0.474)	(0.4478)	(0.3932)

Table 6: StdA2

Table 6 shows the parametric estimates for the effect of accumulated reviews (A2) on prices. (Standard Errors are provided in brackets)

A3	6.2219	13.7515	10.6446	12.9308
	(1.821)	(5.6608)	(3.4938)	(17.7163)
A3sq		-7.9751		-5.7898
		(5.7109)		(42.9175)
A3cb			-5.4508	-1.5313
			(3.4176)	(27.598)
Constant	2.9602	2.0041	2.2378	2.0631
	(0.5031)	(0.8354)	(0.7099)	(1.4495)

Table 7: StdA3

Table 7 shows the parametric estimates for the effect of accumulated clicks (A3) on prices. (Standard Errors are provided in brackets)

A4	3.7206	10.5152	7.5803	12.8161
	(1.6962)	(4.7073)	(3.2638)	(9.7778)
A4sq		-6.4213		-11.8855
		(3.6407)		(20.3634)
A4cb			-3.9580	3.5697
			(2.2771)	(12.6267)
Constant	5.7483	4.4565	4.7803	4.2303
	(0.769)	(1.2208)	(1.1001)	(1.4586)

Table 8: StdA4

Table 8 shows the parametric estimates for the effect of accumulated LCTs (A4) on prices. (Standard Errors are provided in brackets)

6   Conclusions and Further Research

In this paper we analyze empirically the competitive behavior of startup firms in e-commerce. To that end we use comprehensive data from Austria's biggest search-engine for price comparisons, tracing 361 start-up firms for up to 62 weeks. We furthermore use data from roughly 600 incumbent firms to disentangle the effect of price movements which affect the entire market from the effect of aging on prices. Using non-parametrical methods of estimation (i.e. local generalized least squares regressions), we investigate the evolution of pricing strategies, portfolio size and portfolio composition throughout the first months of the entering firms.

Being interested in the role of customer-reviews and whether start-ups pursue special strategies to acquire reviews more quickly, we used four definitions of age. The first is based on age in weeks and the second on accumulated reviews, which are typically filed by customers who bought from firm i in earlier periods. Definitions three and four are based on accumulated clicks and last-clicks respectively and are thus also a measure of past turnover.

We find that start-ups typically start out on a lower price level and increase prices as they grow older. Moreover our evidence suggests that age measured in accumulated past clicks is more relevant to start-ups' behavior than their mere aging. Our analysis of the evolution of portfolio-size of start-ups, does not reveal a clear pattern of portfolio-change as firms mature. An initial reduction

26

of items is particularly strong if age measured as accumulated reviews and does not play an important role when age is measured in clicks. Finally, results from the analysis of changes in the intensity of competition mirror the evolution of portfolio-size, which hints at a possible relationship between those two variables.

For further research we suggest looking at the quality entrants provide and analyze possible relationships between the intensity of competition in the markets and the average quality grades shops receive. It could also be useful to decompose our measure for intensity of competition and to look at number of competitors and market volume separately. Portfolio composition in an multiproduct-retail environment is certainly an interesting topic as such. However, we also deem it important to control for different portfolio-composition when analyzing the possibly different price and quality levels of start-ups and incumbents. This is an issue which we consider worthwhile tackling in further investigations. There is another issue which it might be interesting to move on to at a later stage: Inspired by the works of Golsbee and Chevalier [GC03] or Ellison and Ellison [EE04], it could be worthwhile to compare the price elasticities that incumbents and and start-ups are facing.

7   Appendix

7.1   Background Information on Technical Aspects of the Data and the Estimation Procedure

Geizhals.at is a price search engine which collects the price offers via standardized protocols from a predefined group of sellers and presents them electronically via its web-platform. This is a different search service than a shop-bot, which usually does an arbitrary price search for products on the whole web and offers the results of this web search on line. Typically the quality and reliability of price offers in price search engines are higher than in shop-bots.

Currently the data at JKU-Linz cover a time period of more than 12 months. On a subset of products information can be tracked over a long period in order

to analyze the whole product cycle from its introduction in the market to the phasing out of the product. Data are directly transferred from Geizhals.at to JKU's server on a day-by-day basis. The single files in its entirety represent a relational database system comprising the following information:

As of today, the data covers price offers for about 180000 products from a total of 1397 sellers. Furthermore the database contains data of customers' quality evaluations of sellers (78,369) and 185,310 product evaluations.[10] Prices can be updated from the sellers about 10 times a day. Furthermore the data comprise detailed information on about 100000 daily customer clicks to retail shops together with the respective views and filter options of the customers.[11]

The raw data at JKU's server doesn't lend itself to direct manipulation, but it has to be transferred to a relational database using MySQL and from there it can be imported into econometrical software packages. While this implies some programming efforts, it also allows to extract additional or data with higher density, whenever this is needed.

As for the calculations that have been performed to complete this paper, our estimations were run on JKUs multiprocessoral computing facility and we wrote the estimator in the open source software-package Octave. Eventually, we used 32 processors to perform the estimations of start-ups' pricing behavior. However, adapting the estimator to multiprocessoral computation is not trivial and was only made possible thanks to the support of Johann Messner at JKU and thanks to recent programming efforts by Creel [Cre05] and Fernandez et al. [FAR06] which make parallel usage of processors in Octave possible. Being thus equipped, it was possible to reduce estimation time to approximately 7 hours per estimation (800,000) observations. Further information concerning the estimator or the code is available from the authors upon request.

---

[10]The reviews use a standardised grading scheme (1-5 according to the Austrian grading system in schools) which lends this data to econometric analysis. The numbers correspond to June, 27th. 2007 and have been retrieved using MySQL-requests.

[11]And a wealth of further information which include the evaluation of products and shops, location of shops, delivery cost and many more. The raw data at JKU's server currently comprises roughly 400GB.

7.2 Background Information on the Last Click Through Variable

While looking at clicks to the web site of an online shop is the ideal way to analyze the capacity of an offer to arouse interest, it is useful to contrast this advertorial function with actual purchases. It is clear, that the actual act of purchasing a product is unknown, given that it happens at the e-tailer's own web site. However, recent investigations have shown, that Last-Click-Throughs (LCT)[12] can be used as a reasonable proxy for the purchasing decision (cf. for example Smith and Brynjolfssens' [SB01] or Bai's [Ba02] contributions). Using this well established concept we use a measure for buying decisions, which has been derived by Dulleck et al. [DHW08], from which we obtain another demand-oriented measure of the experience of a shop (A4). In the following, we provide a few details on how this measure has been derived.[13]

Analyzing the click behavior of a customer over time we have to define a 'searching period' which is finished with an actual purchase. The approach which has been chosen for identification of search periods relies on hierarchical clustering. The clicks with respect to their minimal temporal distance were sequentially added to get a dendrogramm from which, after fixing the hierarchical level, we obtained the relevant search intervals. While choosing a low level results in many search spells, choosing a high level rendered fewer intervals. Thus Grubbs' Test for Outlier Detection was applied to identify the optimal search intervals.[14] Furthermore some additional minimal requirements were introduced to account for outliers, which are typically observed in the procedure.[15]

Finally, customers might not only search for one specific product, they might look at substitutes during their search as well. However, the data from

---

[12]If a customer is searching for a product, she might meander around different web sites, comparing characteristics of the shops, but she will finally settle for the preferred shop and buy there online. This "last click" is called Last Click Through.

[13]However, our account of the construction remains somewhat superficial and draws heavily Dulleck et al [DHW08]). The reader, who is interested in a thorough treatment of LCTs must be referred to this paper or to [SB01].

[14]The significance level chosen to distinguish the different relevant search intervals was 95%.

[15]The sequence of clicks is divided into several search spells if there is a an interspace of one week without clicks and the resulting search periods contain at least 3 clicks, in a second version a maximal interspace of one month and a minimal amount of 5 clicks is applied.

geizhals.at contains a hierarchical mapping of the products into subsubcategories, subcategories and categories, which allows to cope with this issue. The three lowest levels of the classification scheme[16] could be used to analyze consumers' search spells. Consequently a measure, which relies on seven-day search spells within the same subsub-category (third level) was selected for the present investigation.

References

[Arg98] Agarwal, Rajshree (1998): Small Firm Survival and Technological Activity , Small Business Economics, Volume 11, Number 3 / November 1998

[Aud91] Audretsch, D.B. "New Firm Survival and the Technological Regime." Review of Economics and Statistics, Vol. 73 (1991), pp. 441–450.

[Ba02]   Bai, X. (2002): Why Pay More? Why Charge Less? An Empirical Exploration of International Online Price Differentials, Heinz School of Public Policy and Management Carnegie Mellon University, Pittsburgh, PA 15213.

[BMS04] Baye, M. R., J. Morgan and P. Scholten (2004): Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site, The Journal of Industrial Economics, Vol. LII, 463-496.

[CM03]   Cabral, L. and Mata, J. "On the Evolution of the Firm Size Distribution: Facts and Theory." American Economic Review, Vol. 93 (2003), pp. 1075–1090.

---

[16]The scheme is devided into categories, subcategories, subsubcategories and products. Especially the latter three are well suited to account for the degree of substitutional relationship between the products (In total 358 subsubcategories and 40 subcategories are given. As an example the category 'Video/Foto/TV' contains the subcategory 'TV-Sets' and the subsub-category 'LCD TV sets with 30-39 inches').

[Cre05]   Creel, Michael (2005): "User-Friendly Parallel Computations with Econometric Examples", Computational Economics (2005) 26: 107128

[DGW05] Duan, Wenjing, Gu, Bin and Whinston, Andrew B., "Do Online Reviews Matter? - An Empirical Investigation of Panel Data" (January 2005). Available at SSRN: http://ssrn.com/abstract=616262

[DHW08] Dulleck, U., Hackl, F., Weiss, B. and Winter-Ebmer, R. 2008. 'Buying Online: Sequential Decision Making by Shopbot Visitors'. CEPR Discussion Paper no. 6939. London, Centre for Economic Policy Research. http://www.cepr.org/pubs/dps/DP6939.asp.

[DRS89] Dunne, Timothy, Mark J. Roberts, Larry Samuelson Source: "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries", The RAND Journal of Economics, Vol. 19, No. 4 (Winter, 1988), pp. 495-515

[EF03]   Ellison, Fudenberg and Mobius (2003): Knife-Edge or Plateau: When Do Market Models Tip?, Quarterly Journal of Economics, Vol. 118, No. 4, 1249 - 1278

[EE04]   Ellison, G. and S. F. Ellison (2004): Search, Obfuscation and Price Elasticities on the Internet, NBER Working Paper 10570.

[EHS05]   Emre, Onsel, Ali Hortacsu, and Chad Syverson (2005) , "E-commerce and the Market Structure of Retail Industries," NET Institute Working Paper #05-24.

[FK07]   Farrell, J. and P. Klemperer (2007): "Coordination and Lock-In: Competition with Switching Costs and Network Effects" in Armstron and Porter (Eds.): Handbook of Industrial Organization, Vol. 3, Ch.31. North-Holland: Amsterdam, The Netherlands, pp.1967-2074

[FAR06]   Fernández, J. et al.(2006): "SCE Toolboxes for the Development of High-Level Parallel Applications" in Alexandrov, V.N. et al. (Eds.):

ICCS 2006, Part II, LNCS 3992, Springer-Verlag Berlin Heidelberg 2006, pp. 518525

[Ger95]   Geroski, P.A.: "What do we know about entry?", International Journal of Industrial Organization 13 (1995) 421-440

[GC03]   Goolsbee, A. and J. Chevalier (2003): "Measuring Prices and Price Competition Online: Amazon versus Barnes and Noble", Quantitative Marketing and Economics, Vol. I, No 2.

[Hol86]   Paul W. Holland (1986), "Statistics and Causal Inference", Journal of the American Statistical Association, Vol. 81, No. 396. (Dec., 1986), pp. 945-960.

[Imb04]   Imbens, Guido W. (2004). "Nonparametric estimation of average treatment effects under exogeneity: a review", Review of Economics and Statistics 86 (1): 4 - 29.

[Jov82]   Jovanovic, Boyan (1982): "Selection and the Evolution of Industry", Econometrica, Vol. 50, No. 3 (May, 1982), pp. 649-670

[KT06]   Klepper, Steven and Peter Thompson: "Submarkets and the evolution of market structure", RAND Journal of Economics Vol. 37, No. 4, Winter 2006 pp. 861–886

[LR06]   Qi Li and Jeffrey Scott Racine. "Nonparametric Econometrics: Theory and Practice", Princeton University Press (November 27, 2006)

[MP94]   Mata, Jos and Pedro Portugal: "Life Duration of New Firms, The Journal of Industrial Economics", Vol. 42, No. 3 (Sep., 1994), pp. 227-245

[MPG95]   Mata, Jos, Pedro Portugal and Paulo Guimares (1995): "The survival of new plants: Start-up conditions and post-entry evolution", International Journal of Industrial Organization Volume 13, Issue 4, December 1995, Pages 459-481

[New94]   Newey, W. (1994): "Kernel estimation of partial means and a general variance estimator". Econometric Theory, 10:233-253, 1994.

[RS05]   Reinstein, D. A. and C. M. Snyder (2005): The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics, The Journal of Industrial Economics, Vol. LIII, No. 1., 27-51.

[RFS92)] William T. Robinson, Claes Fornell and Mary Sullivan (1992): Are Market Pioneers Intrinsically Stronger Than Later Entrants?, Strategic Management Journal, Vol. 13, No. 8 (Nov., 1992), pp. 609-624

[SB01]   Smith, M. D. and E. Brynjolfsson (2001): "Consumer Decision-Making at an Internet Shopbot": Brand Still Matters, The Journal of Industrial Economics, Vol. XLIX, No.4. 541-558.

[Sta01]   Stephen S. Standifird (2001): Reputation and e-commerce: eBay auctions and the asymmetrical impact of positive and negative ratings, Journal of Management 27 (2001) 279–295

[WC03]   Waldfogel, J. and L. Chen (2003): Does Information Undermine Brand? Information Intermediary Use and Preference for Branded Web Retailers, NBER Working Paper No. 9942.

[Woo05] Wooldridge, Jeffrey M. (2005). "Unobserved heterogeneity and estimation of average partial effects", Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg: 27-55 (D.W.K. Andrews and J.H. Stock, Eds.). Cambridge: Cambridge University Press.