What is in a price? Evidence on quality signaling for experience goods

Rroshi, Daniela; Weichselbaumer, Michael

DOI:
10.57938/0c79dc48-9599-4e10-9d52-abb8092e3689

Published: 01/03/2021

Document Version:
Publisher's PDF, also known as Version of record

Department of Economics
Working Paper No. 311

What is in a price? Evidence on quality signaling for experience goods

Daniela Rroshi
Michael Weichselbaumer

March 2021

WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

EFMD EQUIS ACCREDITED AACSB ACCREDITED ASSOCIATION OF MBAS ACCREDITED

What is in a price? Evidence on quality signaling for experience goods*

Daniela Rroshi† Michael Weichselbaumer†

January 2021

Abstract

We study the effect of quality disclosure on prices using quality tests released by the main consumer protection agency in Germany. Both durable and non-durable consumer products are covered, representing about 5 percent of weighted expenditures for all products in the German CPI. Cross-section results of the price–quality relation before quality disclosure show that higher prices are positively correlated with higher quality for durable goods, and negatively correlated for non-durable goods; both results are in line with theoretical models of price signaling. In the dynamic analysis, we employ a RD-type approach around publication of the quality evaluation for identification. Results show a positive effect of quality disclosure on prices for high quality durable products and a negative effect for low quality products suggesting that the information improves matching. Opposite results hold for non-durable products. Survival estimates show that products of low quality leave the market earlier.

JEL Classifications: L15, D18, D22

Key Words: Asymmetric Information, Consumer Protection, Quality Disclosure.

*We thank Mario Lackner, Charles Louis-Sidois, Mariya Teteryatnikova, and seminar participants at the Vienna University of Economics and Business for helpful comments and suggestions.

†Vienna University of Economics and Business, Department of Economics, Vienna, Austria.

1 Introduction

Consumers frequently face quality uncertainty when deciding which one of competing products to buy. Household appliances or entertainment electronics, and products like sunscreen, child safety seats or bicycles are groups with many competing products on the market where it is practically impossible to make systematic comparisons for consumers. In many countries, consumer protection agencies exist to provide an independent source of information, for example through quality testing. We use such data to answer two main questions. Do prices signal quality? Do quality tests from the consumer protection agency have an impact on the market outcomes?

Asymmetric information in markets for experience goods results in inefficient market outcomes, characterized by a domination of low quality producers. One way for firms to mitigate the consequences of asymmetric information is to voluntarily provide information about product quality to consumers. Grossman (1981) and Milgrom (1981) show that in particular high-quality firms have incentives to do so (the so-called unravelling argument). However, both the theoretical and empirical literature agree that in many instances firms lack the incentives to voluntarily disclose information, usually because disclosure is costly (Jovanovic (1982) and Grossman and Hart (1980)) or because it intensifies price competition (Board, 2009). In these cases consumer policy can improve welfare either through direct regulation like minimum quality standards and mandatory disclosure or by subsidizing the acquisition of information (Tirole, 1988).[1]

Consumer policy considers information provision approaches to be more efficient in reducing information asymmetry compared to regulation, because these do not put any restrictions on consumer choice (Beales et al. (1981) and Vickers (2004)) as compared to other forms of regulation. The theoretical and the empirical evidence is ambiguous about the effect of increasing consumer information on market outcomes and about the way in which this information should be disclosed. In many countries consumer protection agencies conduct product tests that put the performance of products along many dimensions under scrutiny. The results of such tests are published on a regular basis in consumer magazines that usually preclude product advertisements to preserve independence, e.g. *test* in Germany, *Konsument* in Austria, *Consumer Reports* in the US, *Which?* in the UK, *UFC-Que Choisir* in France. Product quality tests are a core activity of consumer protection agencies, also in terms of revenue, implying incentives to select

[1] Apart from quality disclosure there are other quality-assurance mechanisms that provide consumers with information such as brands, experience, warranties and licensing (Dranove and Jin, 2010).

goods according to the property of having high asymmetric information problems for consumers. Those goods are an ideal set of products to study the effect of an information shock on prices.

Consumer test magazines potentially have a high impact, through at least two channels. First, through direct readership and referencing in other media. The German magazine has around 70 million visits on their website and the average number of magazine issues sold monthly in 2019 was 368,000. Around 20,000 articles in printed newspapers, some 2,500 TV reports and 4,500 radio reports refer to the results (Stiftung Warentest, 2019). Second, online price comparison websites (e.g., in Germany, idealo.de and geizhals.de) provide the test results or link to them together with information on various product characteristics. Indirect spread of information arising through word-of-mouth is an additional way how test results can influence consumer information.[2] Therefore, as these test results reach large numbers of consumers, we expect them to be able to have effects on prices.

In this article we evaluate whether information release about product quality has an effect on markets for consumer goods, to provide evidence on the usefulness of such activities. Our analysis focuses on the effect of quality information disclosure on prices and the exit of products of different quality from the market. In the first part of the empirical analysis we test the signaling role of prices by investigating the cross-sectional relationship between prices and quality. These models predict a reduction of price distortions if more information is available to consumers. Theory predicts that the direction of the implied price change depends crucially on the durability of the product. Therefore, in the empirical analysis we distinguish between durable and non-durable products. Our identification strategy relies on the exogenous change of consumers' information about quality through the publication of the test results. The data cover many different products, tested at different points in time, allowing us to investigate the broad effect of this intervention. We combine the test quality data with daily price data, which allows us to identify the immediate effect of the test results on prices.

The cross-sectional results confirm the theoretical predictions of the price signaling literature. We observe higher prices for high quality durable products, and lower prices for high quality non-durable products before the publication of the test results. In the dynamic analysis, we find that the immediate price effect of quality disclosure on price dynamics is an increase of about 1 percent for high quality durable products and about 2-5 percent decrease for high and medium quality non-durable products, which is a pattern not in line with the theoretical

[2]For the effect of school-quality information on school choice, Koning and van der Wiel (2013) show that the indirect diffusion of quality results is equally important as direct exposure to the information in the newspaper that originally published this information.

models discussed. Looking at the survival of products after the quality grades are published, we observe that no goods that are of the highest quality leave the market within 500 days, and that market exit increases to about 10 percent for low quality levels. The lowest quality products, though, leave the market at a smaller rate than all but the very highest rated products. The results show a behavior of products of insufficient quality — in terms of price levels, as well as for price dynamics and market exit — of imitating the behavior of good quality products.

This study contributes, first, by providing empirical evidence on the role of prices as quality signals using a unique dataset on product quality that covers a wide range of experience goods. We find that prices vary proportional to quality. Moreover, combining the product quality information with daily prices allows us to test the effect of quality information on prices and compare the results to implications of the theoretical models of price dynamics in experience goods markets.

The second contribution relates to the empirical literature on the effects of mandatory or third party quality disclosure on firms' behavior. Empirical evidence on the effect of disclosure on quality provision by firms is available mainly in the education, food/beverages and health markets (Dranove and Jin (2010) provides a review). We provide empirical evidence on the effect of disclosure on firms' pricing behavior for a wide range of experience goods markets using a common empirical background.[3]

Finally, our paper also contributes to the economics literature on consumer protection (Armstrong et al. (2009), Armstrong (2011) and Vickers (2004)) by evaluating one of the main instruments of the consumer protection agencies. Our results show that providing information is important for consumers and helps them to make better choices. The price changes suggest that when consumers get to know the quality of their products they substitute low quality products with better quality products.

In the next section, we discuss the implications from the theoretical literature on information disclosure and compare our approach to the empirical literature. In Section 3, we explain our identification strategy. Section 4 describes the data we use. Section 5 presents the results. Robustness checks that address modeling and data issues follow in Section 6. The article ends

[3]Several studies in finance evaluate regulatory and mandatory disclosure measures aiming to protect investors and borrowers in these markets. For instance, Greenstone et al. (2006) investigate the effect of mandated disclosure of financial information on stock returns and firm performance. Brown and Jansen (2020) look at the effect of usury and wage garnishment laws on borrowers outcomes in the auto loan industry. Another stream of literature investigate how stock prices react to new information, see for a more detailed discussion the work of Hollenbacher and Yerger (2001) and Dellavigna and Pollet (2009).

with a discussion and conclusion in Section 7.

2 Price as Quality Signal

Prices are the most fundamental instrument for firms to signal quality to consumers in experience goods markets. Firms use either prices alone to signal quality or in combination with other instruments. There is an extensive theoretical literature on warranties, umbrella branding and reputation as signaling instruments. Firms can also voluntarily disclose quality information through product certification or advertising. This discussion of the theoretical literature is on quality signaling through prices. In the second part of this Section we review the related empirical literature.

2.1 Models

The theoretical models on price signaling convey different predictions regarding the sign of the correlation between price and quality. The predictions depend on, first, the nature of the good — whether it is a durable or non-durable product — and second, on the type of producers — whether it is a high quality versus low quality producer.

Central to the signaling literature is the mechanism that in an asymmetric information environment high quality producers will set the price such as to distinguish themselves from the low quality producers. Early theoretical literature shows that in monopoly markets with incomplete information and exogenous quality, high-quality producers will signal quality by setting a price above the full information price, whereas the low quality producers choose the full information price. Generally, this will lead to upward price distortions. In the model by Farrell (1981) and Wolinsky (1983) signaling works because of the presence of a share of informed consumers; in the models of Milgrom and Roberts (1986) and Bagwell and Riordan (1991) signaling is viable because the production of high quality costs more than low quality. Daughety and Reinganum (2008) analyze price signaling in the context of imperfect competition with horizontal product differentiation. They show that in contrast to a monopoly setting, the equilibrium price with imperfect competition is higher than in the full-information equilibrium for both high quality and low quality firms due to the strategic interaction between firms. The model predicts that, when more information becomes available, prices decline to a much greater extent compared to the monopoly situation, because both low quality and high quality products will reduce signaling distortions. Janssen and Roy (2010) show that this result holds and prices signal quality even when there is no horizontal product differentiation and there is tough price competition. An essential assumption for the predictions in Bagwell and Riordan (1991), Daughety and Reinganum (2008) and Janssen and Roy (2010) is that high quality products are associated with higher production costs than low quality products. These studies argue that in practice this assumption is reasonable and relevant for most durable products.

Based on these models, we expect the following patterns for *durable* goods. As long as consumers have little information about quality in experience goods markets, high quality is signaled by high price, therefore we expect to see a monotonic positive relationship between quality and price. Regarding price dynamics, as information about the quality diffuses, the price distortions will be reduced and prices of both high quality and low quality products will fall. This will materialize in particular when an exogenous shock occurs that increases the amount of information consumers receive. However, we expect that the price reductions will be more pronounced for high quality products because they experience the largest distortions.[4]

In markets for *non-durable* goods, repeat purchases and therefore reputation considerations play an important role. High quality can be signaled through a low price (Tirole (1988) and Shapiro (1983)). This can be observed especially in the case of introductory offers when a new product is brought into the market and the producer wants to induce the customers to try the product by means of a low price. This reputation mechanism is more important for non-durable products because consumers can learn product quality by experience within a relatively shorter period of time. In these cases firms will have incentives to maintain high quality levels in order to avoid reputation damages if consumers would observe a low quality in a given period. Moreover, because a high quality product is more likely to induce repeated purchases, a low introductory price is more valuable to high quality producers.

The expected pattern for non-durable products will therefore be as follows. We expect a negative relationship between price and quality for non-durable products. Prices of high quality non-durable products will increase over time as more information becomes available.

2.2 Empirics

Early empirical studies offer mainly descriptive evidence and find a weak relationship between price and quality (Riesz (1979), Gerstner (1985) and Curry and Riesz (1988)). Caves and Greene

[4]There are also models where quality is endogenous and firms set a price premium for high quality (models of quality-guaranteeing price). We focus on changes in firms behavior in a short period surrounding the event of the publication. During this short period, quality can be assumed to be fixed and we abstract from quality changes and reputation effects.

(1996) analyze the use of prices as signals of quality measuring the amount of information consumers hold based on survey data. The analysis relies on cross-section data for some 200 products evaluated in the Consumer Reports. They find that price–quality correlation depends on the amount and type of information consumers hold. The results suggests that prices serve only as signals for products of frequent but unimportant purchase or for products where brand is important for the consumer choice.

A growing literature investigates the effect of online reputation mechanisms on sellers' behavior by exploiting the fact that in an online environment the researcher knows the information sets of both buyers and sellers. However, these studies rely on buyer reviews as an indicator of quality or reputation. Cabral and Hortacsu (2010) show that a negative feedback received by a seller on eBay is associated with a decrease in price and increase in sales. A related study that also looks at the eBay feedback mechanism is provided by Jin and Kato (2006). They conducted a field experiment and let professionals evaluate baseball cards in order to obtain a quality measure. They find evidence of fraud. Some sellers claim high quality and less-experienced buyers pay high prices without receiving the high quality claimed. Reputation does not solve the problem; while it helps consumers to identify the truthful producers, more reputable sellers do not offer better quality. Lewis (2011) looks at the effect of voluntary disclosure of product characteristics on prices. He finds that the degree of information that sellers reveal about the product in terms of the number of photos and text descriptions of products strongly influences prices. These studies exploit changes in the reputation system of eBay for identification but do not explicitly consider an exogenous increase in the amount of information consumers hold. Focusing on the used car market, Hollenbacher and Yerger (2001) look at the effect of Consumer Reports' reliability evaluations on resale prices, finding that used vehicles that belong to the compact/sub-compact category experience a price decrease following a negative evaluation relative to other products.

Quality information disclosure can also affect sellers' market exit behavior and quality adjustment. Cabral and Hortacsu (2010) argue that in an anonymous market one should expect an effect of product ratings on exit, because rational sellers change their behavior just before leaving or are more likely to leave after having received the negative feedback. This hypothesis is confirmed. They find that sellers with lower reputation are more likely to exit. Klein et al. (2016), though, do not find evidence that an increase in market transparency increases sellers' exit. For a change in market transparency, they interpret as a natural experiment the change of eBay's settings to display seller information by default. Their results suggest a reduction in moral hazard and sellers that improve their performance in a more transparent market environment.

Overall the empirical evidence on quality information provision after disclosure is mixed, also when considering evidence from settings other than an online environment. Jin and Leslie (2003) investigate the effect of mandatory disclosure of quality in the restaurant market. They find that the display of grade cards causes restaurants to make hygiene quality improvements and reduces the incidence of food-borne diseases. Looking at the health sector, Dranove et al. (2003) in their review show that in general there is no improvement in quality of the service offered by doctors. In health markets, rather than quality improvements, obligatory information provision has negative consequences on consumer welfare because providers engage in selective behavior and refuse to treat severely ill patients. For instance, Werner and Asch (2005) show that the incidence of cardiac surgery for minority patients declined after the publication of quality report cards.

Some studies investigate the impact of other disclosure mechanisms on consumer demand. Mathios (2000) exploits the change in regulation which made voluntary nutritive information disclosure mandatory for salad. They find that prior to regulation, the low-fat salads (high quality) were more likely to voluntarily disclose information and there is a significant effect of nutritive elements disclosure on consumer choice. Freedman et al. (2012) investigate the effect of product recalls of children's toys that violate product safety and find industry and brand effects. Koning and van der Wiel (2013) find a positive effect of school-quality information published in a newspaper on school choice in the Netherlands.

An alternative to systematic quality evaluation by a consumer protection agency is product ratings either through consumers or through experts. In Anderson and Magruder (2012), the effect of eBay ratings on restaurant customer flows is largest for those restaurants on which there is a lack of information from other sources. Studies on the impact of expert reviews on consumer demand are Reinstein and Snyder (2005) for movies, and Friberg and Grönqvist (2012) and Hilger et al. (2011) for wine. Ackerberg (2003) and Ippolito and Mathios (1990) show that the information provided through advertising affects consumer demand. Overall, these studies conclude that consumers react to quality information and substitute away from low quality products towards high quality products, provided that the information is easy to understand and contains new information.

Relying on consumer or expert ratings as an indicator of quality gives less clear and more subjective quality measures. Consumer ratings may be manipulated. Expert reviews may be appropriate only for specific settings such as for wine, restaurants, theatres, or books; they may be influenced by personal taste; they can be hard or impossible to reproduce. In contrast to these studies or to studies that investigate voluntary disclosure of information by firms, we rely on an objective and exogenous measure of quality provided by an independent agency as the source of information. Product test procedures are documented in detail together with the test results. The German consumer protection agency is financed primarily through the sale of their publications and receives some government subsidies. Financing through advertisements is prohibited.

The behavioral economics literature emphasize that whether consumers react to quality ratings or not depends on the amount of attention they pay to the information provided (Dellavigna and Pollet, 2009). Pope (2009) shows that the consumer reaction depends on the reported measures of quality and recommends that these should be easier to understand, given the limited cognitive ability of consumers. The information provided by the test magazine from which we collect our data is easy to comprehend. It provides a lot of detail on testing procedures and tested characteristics, but also a single number aggregating all results for a product that resembles school grades (a sample can be seen in Figure A.1).

The empirical studies on dynamic effects of quality disclosure focus on single markets, usually exploiting either the nature of information in online platforms or a change in the regulatory environment. Our study considers a wide range of experience goods markets and the sample allows us to distinguish between durable products and non-durable products in a unified setting.

3 Setting and Research Design

Our identification strategy exploits the exogenous increase in the amount of information consumers hold and the quasi-random timing of the quality disclosure to identify the effect of quality on prices. Identification relies on price comparison of the products within the interval of $[-105, +105]$ (+/- 15 weeks) days around quality evaluation. The 105 days window is a somewhat arbitrary compromise of having a window that is large enough to have precision in the estimation and having enough products that have a long enough time series of data available. We will address this in a robustness check below. For identification to be valid, two important conditions have to be met. First, test results are not communicated before the event. Second, quality is stable within the window.

The timing of quality disclosure is coupled with the publication date of the test magazine. Concretely, the publication date is defined as the day when the test results are published online and sent out to subscribers and newsstands. This happens 2 to 12 days before the first of the month of the issue in question, with the precise date available for every monthly issue. The product categories that are tested are not communicated in advance to consumers or producers. Products are bought anonymously to avoid that producers know that their products are being tested.[5] Producers also cannot influence the choice of the topics or ask the consumer protection agency to test their products. Free sample products from producers are never asked for, nor used or accepted. Testing is sometimes outsourced to independent testing institutes, which remain anonymous to the public and producers and conform to the same testing standards of anonymous purchase. Producers are consulted only after the products have been tested to ask them if, according to their opinion, all relevant product components are incorporated in the test, but no (preliminary) test results of their products are communicated before publication of the magazine (Stiftung Warentest (2014)).

There is a large range of products and market segments tested and no information on the products tested is published before the magazine appears. Therefore, we consider the choice of products in a particular issue of the magazine as quasi-random. The arrival of the quality information on the publication date is treated as an information discontinuity about quality.

Absence of preliminary disclosure of the products to be tested lead to ignorance of producers about being tested, which is important for the stability of quality. Producers do not know if and when they are tested, and cannot tie quality adjustments to being tested. The event window is also relatively short for changing quality after the test results are published, and in particular for consumers to then also learn about possible changes in quality. Most products in the sample have detailed specifications, model names and numbers and it may be that changing the quality in many cases would result in the creation of a new product and model name/number. For these reasons, at that time when the event window starts, the producers cannot influence the quality of the product, meaning exogeneity of the quality measure.

The published test results provide a systematic evaluation of quality for a large number of products across different markets. They provide standardized measures of quality that are easy to compare across producers and easy to understand for consumers. Quality is measured on a wide range of objectively verifiable characteristics. Test results and the structure of the timing of information flows lend themselves to significantly increase the degree of information.

[5] Detailed information on the test procedure can be found here: `https://www.test.de/unternehmen/testablauf-5017344-0/`

4 Background and Data

Stiftung Warentest is the main consumer protection agency in Germany. It provides quality result tables in the monthly magazine called *test*. Its stated aim is to provide objective information about products and services that improve consumers' judgement about the products, based on systematic and reproducible testing. Stiftung Warentest is a non-profit foundation. Per its statutes, product advertisements are prohibited, to maintain independence of information from firms. Revenues are generated by selling magazines and other consumer information publications. Losses are covered by the government (2019: about 5 percent of total revenues).

For most products and characteristics, a consumer cannot judge the products even after purchase because of lack of knowledge or expertise to evaluate the technical or chemical aspects, in particular in comparison with other products. For instance, consumers are typically unable to judge whether a face cream contains dangerous ingredients that are detrimental to his health or if products have safety problems; or easily measure TV display quality under different light environments and viewing angles.

We use all 36 issues over the years 2016, 2017 and 2018 for a total of 230 tests for which the prices are available on a price comparison website.[6] Within these 230 tests, the majority of tests is for durables, and in particular many electric and electronic goods: household appliances, entertainment electronics, computers, tools, etc. Examples are notebooks, mobile phones, battery lawn mowers, speakers, dishwashers, light bulbs, and coffee makers. Non-electronic goods in the sample are for example creams for hand/feet/face, sunscreen, child safety seats, bicycles, or toothpaste. The quality of the products is evaluated along several dimensions that vary from test to test, depending on the product in question. The focus is how well the product works to fulfill the primary use of the product. For durables or electronic products, additional test dimensions can be safety, durability, ease or comfort of use, equipment, technical functionality, power consumption, noise etc. For non-durables the tests evaluate — besides performance — harmful substances, packaging, declared information etc. Sometimes these dimensions also include norms already defined by laws, but where ex ante proof is not available to consumers that producers actually comply. These detailed dimensions each receive grades for each product, from which a weighted average is calculated to obtain a total grade for each product.

[6] This is the vast majority of tests in the magazine. The main categories usually *not* on the price comparison website are food products and tests for service quality. Examples for tests of the latter are: home care intermediation service, service quality of online pharmacies, online dating agencies, or tour operators for traveling.

The 230 tests cover 2,736 tested products, which were matched to price series data from the German site *geizhals.de*, which provides price comparisons from different sellers. Prices are the daily minimum price offered per day. Products in the magazine *test* are clearly specified with product names that makes confusing similar products almost impossible. Pictures frequently available both in *test* and on *geizhals.de* further helped in identifying matches. The overall matching rate is 85.6 percent, and the products share for those ending up in the estimation sample with data available for the estimation window of +/-105 days is 61.1 percent (1,672 products). Note that this matched sample is not selected on quality (see Table 3).

The consumer protection agency, by its raison d'être, has an incentive to select those products with the highest asymmetric information problems for consumers. Therefore, the goods represented in the tests, and hence our sample, are by construction not a random sample of all products, but chosen for the presence of asymmetric information. We cannot quantify the extent of asymmetric information, but we can quantify coverage of our data in terms of representation in the CPI basket-of-goods.

We have mapped the products that are in the tests of our sample to the 645 lowest-level categories in the weighting pattern of the German Consumer Price Index (Destatis, 2019). Our data cover about 1/9 of these categories (69/645). Table 1 shows the names of these 69 categories and the number of products falling into each category from our sample. The Table is sorted in descending order by the size of the CPI weight of the CPI category. Clearly, the tested products falling into each CPI category are not all products feasible for the category, but we find this to be a useful way to illustrate the representativeness and coverage of our data. Summing up all the weights, without counting any product twice, the products cover about 5 percent of the whole German CPI baskets of goods (47.7 of 1,000).

Table A.1 in the Appendix gives additional insight by looking at the CPI-category aggregates at the 2- and 3-digit level. There are 12 main level (2-digit) categories in the weighting pattern of the German CPI (Column (2)), of which 8 are covered. Column (3) shows the weights, which sum to 1,000 for the whole basket. Column (5) shows the 3-digit subcategories for those main categories that are present in our estimation sample. From those 31 subcategories, 16 are covered. Column (7) shows the lowest-level weights covered by our data, summed up at the 3-digit level, which again gives the value of 47.7 of 1,000. Column (8) repeats the display of the number of products in our sample that fall into that 3-digit level.

As a further point of reference for the weighted expenditures covered by our dataset, we calculated the relation between the cost of the CPI basket of goods and GDP. GDP in Germany in 2018 was 3,388.2 billion euro. Reconstructing the price of the basket of goods, we take

Table 1: Product categories of CPI weighting pattern

Product Name (1)	CPI Weight (2)	N (3)	Product Name (4)	CPI Weight (5)	N (6)
Wireless telecommunications services	9.27	17	Operating system or other PC application software	0.45	20
Non-prescription drugs	2.41	14	Children's furniture	0.40	16
Mobile phone without contract	2.14	160	Hi-fi system, speakers or the like	0.40	95
Portable computer	2.08	173	Radio, excluding car radio	0.40	18
Desktop computer	2.08	11	Coffee maker, tea urn or electric kettle	0.38	54
Blood pressure monitor or the like	1.34	29	Streaming music	0.37	9
Bra or other corsetry articles	1.26	12	Accessories for bicycles	0.36	153
E-bike or pedelec	1.05	27	Fitness equipment	0.31	25
Electric mixer or blender	1.02	45	Digital camera	0.29	99
Bicycle	0.99	20	Musical instruments, including accessories	0.29	10
Access to online services and internet	0.82	44	Cooking pot	0.28	16
Tumble dryer	0.82	47	Kitchen scales, stirring spoon, masher or the like	0.28	28
Washing machine	0.82	40	Stove with oven	0.28	13
Dishwasher	0.82	43	Microwave	0.28	16
Vacuum cleaner	0.76	72	Model railway or toy car	0.27	10
Satellite kit	0.71	44	Tricycle, scooter or other children's vehicle	0.27	15
Home cinema system	0.71	33	Construction kit, experimental kit or model kit	0.27	30
TV set	0.71	73	Toaster or barbecue grill	0.26	17
DVD player or blu-ray player	0.71	11	Memory card or USB flash drive	0.25	16
Flower pot, flower box or climbing aid	0.58	16	Suitcase, travel bag or the like	0.24	20
Petrol or electric lawnmower	0.54	53	Baby carriage or baby carrier	0.24	39
Impact drill	0.54	47	Baby or child car seat	0.24	136
Mouthwash, dental floss or the like	0.53	20	Zoom lens or the like	0.23	54
Shower gel, bath foam or the like	0.53	17	Headphones	0.22	72
Deo spray or deo roll-on	0.53	16	Printer	0.22	80
Day cream or night cream	0.53	102	Keyboard, mouse or other PC accessories	0.22	11
Toothpaste	0.53	38	Wet shaving razor, razor blades or the like	0.21	12
Hair shampoo	0.53	11	Bicycle helmet	0.20	44
Landline telephone	0.52	15	Extractor hood, ventilator or the like	0.20	31
Refrigerator	0.48	20	Wood burner or the like	0.20	10
Freezer or chest freezer	0.48	47	Iron	0.18	14
Fridge-freezer	0.48	52	Electric toothbrush	0.17	28
Batteries	0.45	20	Convenience food for infants or toddlers	0.13	15
Alarm or motion detector	0.45	42	MP3 player or the like	0.07	35
Halogen bulb, energy saving bulb or LED light	0.45	44	Total	47.72	2736

Notes: Columns (1) and (4) contain the names of all products in the Full sample corresponding to the lowest-level categories in the Weighting Pattern of the German CPI (Destatis, 2019). Column (2) and (5) contain the weight in the CPI for each category. Column (3) and (6) contain the number of tested products in each category we have in the sample of test results.

Table 2: Quality grades non-/durables and by sample

Grade	1	2	3	4	5	Total
Verbal	Very good	Good	Satisfactory	Sufficient	Insufficient	
Full sample	100	1,410	818	230	178	2,736
durables	77	1,273	773	224	156	2,503
non-durables	23	137	45	6	22	233
Dynamic sample	38	897	535	130	72	1,672
durables	37	860	522	129	66	1,614
non-durables	1	37	13	1	6	58

Notes: Number of products by quality result from the product tests. For each grade, one line reports the number of durables and one line for non-durables. The top panel shows the Full sample which covers all products of the tests we have included. The bottom panel covers the subsample of products where we have sufficient price series data for the Dynamic sample that uses +/- 105 days of price data around the event.

expenditures for housing (incl. energy costs and repair/maintenance), which were 908 euro per private household and month and in the CPI weighting scheme account for a share of 0.3247. There were 41.378 million private households in Germany in 2018. Therefore, the basket of goods covers

$$12 \cdot 908/0.3247 \cdot 41.378 = 1{,}388.5$$

billion euro, from which we derive the coverage of the basket of goods in relation to GDP as 1,388.5/3,388.2 = 0.4098. This means, the 5 percent coverage of the basket of goods of our sample corresponds to about 2 percent of GDP.

4.1 Cross-section and dynamic sample

The sample with all 2,736 products from the tests will be referred to as *Full sample* henceforth and used for the cross-section estimations. For our RD-approach we need price data for each day +/-105 days around the event. The sample of products where we also have sufficient price series data will be referred to as *Dynamic sample*. Moving from Full to Dynamic, the number of products contracts because of missing price history data on the price comparison website.

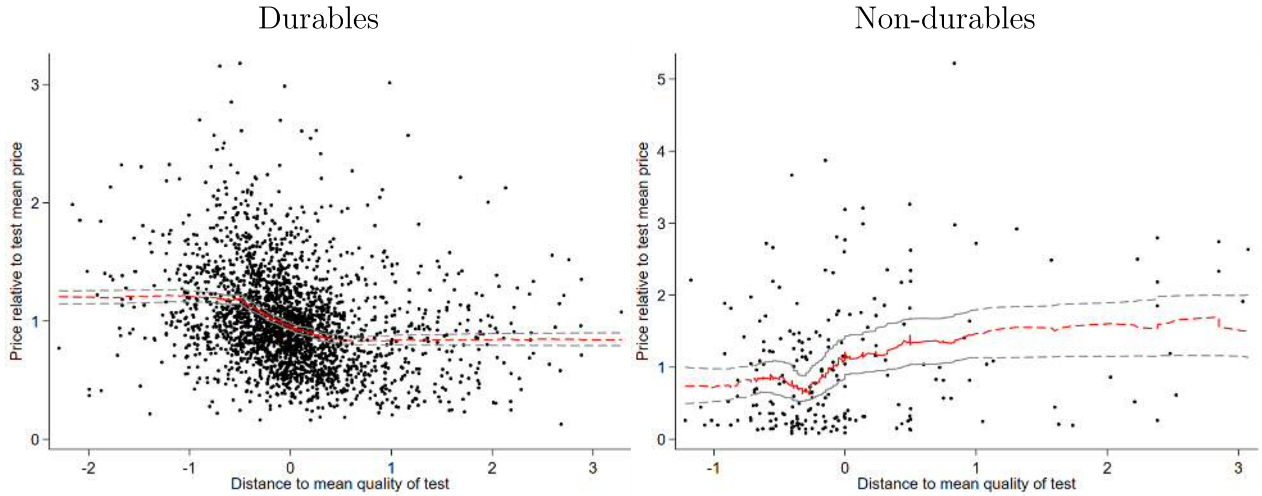
Table 2 shows the observations subdivided by type of sample, by durability, and by ordinal grade from 1 (best) to 5 (worst) achieved in the product test. Durability differentiation is highly important because the predictions for price setting and its response after quality revelation depend on it.

For the data analysis, we use the more detailed subcategorisation of products from the magazine. Product groups in the magazine refer to a type of product, like a TV set, a washing machine, or child helmet. Frequently, these product groups are subdivided when a particular characteristic makes the products weaker substitutes. For example, washing machines that are front-loaders vs. top-loaders; or, child helmets fitting a particular age group; or, TV sets of 55 inch screen diagonal vs. 65 inch. We use the lowest-level characteristic by which test tables are organized to make products within a test comparable. This is relevant for the cross-section analysis; in the dynamic analysis, the subcategories are eliminated by the product fixed-effects we include. Table 3 shows summary statistics for the Full and the Dynamic sample. The first line shows that there are 365 product groups by subcategories in the Full sample. 327 tests remain in the Dynamic sample.

Over 90 percent of the products are durables (Table 3). The average number of products per test goes down from 7.5 in the Full to 5.1 in the Dynamic sample because of missing price series data. Price summary statistics are shown, and prices relative to the mean test price; the latter will be the dependent variable in the regressions to avoid dominance of more expensive product groups in the price reaction regressions. Prices of the products vary widely by type of product, from below 1 euro for products measured in units of 100 milliliters (e.g., all sorts of creams), to up to 4,000 euro for some TV sets, E-bikes and some types of cameras. The mean price is 377 euro for Full, with a median of 200 euro; for Dynamic, it is 431 euro and 262 euro. The difference comes partly from the relative reduction of the usually cheaper non-durables, which have smaller representation on the price comparison website.

Each product receives a numeric aggregate quality value in steps of 0.1 from 0.5 to 5.5, which is derived from the test categories — but without being influenced by the price of the product. Table 3 shows a mean quality of 2.7. The quality results are mapped into five ordinal grades of quality by Stiftung Warentest: Very good (0.5-1.5), Good (1.6-2.5), Satisfactory (2.6-3.5), Sufficient (3.6-4.5) and Insufficient (4.6-5.5). For our empirical analysis, we use the five ordinal grades for differentiation of quality. The ordinal grades are easier to process and remember by consumers, and may be more relevant to them. The line *History* refers to the time in years that the product was listed on the price comparison website before the publication day of the test result. The mean is 1.3 years for durables and 3.7 years, the medians are 0.8 and 3.9 years. *Future* measures the time in years that the product was listed after the event. This variable is censored for all products still on the market when our data extraction ended. To avoid bias, we uniformly apply censoring the variable Future at 500 days (1.37 years). Quartile 1 reveals that most products remain on the market for 500 days after the event; only 6.8 percent of the

Figure 1: Price and Quality



⟵ higher quality / lower quality ⟶ ⟵ higher quality / lower quality ⟶

Notes: Price–quality plot for durable products (left) and non-durable products (right). Prices are divided by the mean price of the products in the test. Quality is centered around the mean quality of the products in a test. The middle (red) line shows the moving average of relative prices with a symmetric window containing 20 percent of the sample. The lines above and below (grey) the moving average show the basic 95 percent confidence interval from 5,000 iterations of bootstrapping the moving average values. For the border areas, where the calculation of the moving average relies on an increasingly truncated sample, the moving average and 95-percent-CI-lines are shown as dashed lines.

products (132/1,948) exit until day 500; all of the exits are durables. For the exit analysis in Section 5, we use all observations with exit data, not only those with data available for the +/-105-window. Therefore, there are 1,948 observations here instead of 1,672 in the Full sample.

Figure 1 illustrates the cross-section relationship of price and quality for durables (left) and non-durables (right). Because of the variability in prices, the Figure shows prices expressed relative to the test group mean. Quality is measured as the difference from the median quality of each product's test, which is better suited to preserve any tendency between price and quality in the cross-section.[7] Figure 1 includes a moving average which for each data point contains a symmetric window with one fifth of the dataset, together with a bootstrapped basic 95 percent confidence interval for the moving average. A lower grade means higher quality. For durables, this yields a clear negative relationship in the area where no truncation for the moving average calculation takes place (solid lines). For non-durables, the pattern is more complex, with a larger part showing a positive relationship.

[7]For example, all products in a test could be relatively good, and therefore all prices be relatively high, which would dilute a negative price-quality relationship when we divide by mean product price only. In the regressions below, we will control for test group fixed effects.

Table 3: Summary statistics

(1) Sample	(2) Variable	(3) N	(4) Mean	(5) Std. dev.	(6) Q1	(7) Median	(8) Q3
Full sample							
Tests (N=365)	Products per test		7.50	4.50	4.00	6.00	10.00
Products	Price	2,736	377.10	486.81	61.37	199.75	516.23
	durables	2,503	411.45	495.15	92.21	238.70	556.88
	non-durables	233	8.13	15.51	0.87	1.96	7.60
	Price/$\overline{\text{Price}}$	2,736	1.00	0.47	0.71	0.95	1.21
	durables	2,503	1.00	0.41	0.73	0.95	1.20
	non-durables	233	1.00	0.89	0.26	0.73	1.42
	Quality	2,736	2.68	0.90	2.10	2.40	3.00
	durables	2,503	2.70	0.89	2.10	2.50	3.00
	non-durables	233	2.46	1.02	1.80	2.10	2.90
	History (years)	1,948	1.42	1.59	0.55	0.85	1.65
	durables	1,877	1.33	1.50	0.54	0.81	1.56
	non-durables	71	3.72	2.01	2.02	3.89	5.50
	Future (years)	1,948	1.34	0.16	1.37	1.37	1.37
	durables	1,877	1.33	0.17	1.37	1.37	1.37
	non-durables	71	1.37	0.00	1.37	1.37	1.37
Dynamic sample							
Tests (N=327)	Products per test		5.11	3.60	2.00	4.00	7.00
Products	Price	1,672	431.74	489.55	111.07	262.04	595.49
	durables	1,614	447.04	491.45	122.64	276.29	616.91
	non-durables	58	6.03	8.44	1.10	2.77	8.24
	Price/$\overline{\text{Price}}$	1,672	1.02	0.43	0.72	0.96	1.23
	durables	1,614	1.01	0.41	0.72	0.95	1.22
	non-durables	58	1.29	0.76	0.60	1.20	1.82
	Quality	1,672	2.63	0.79	2.10	2.40	3.00
	durables	1,614	2.64	0.78	2.10	2.50	3.00
	non-durables	58	2.50	1.02	1.80	2.10	3.00
	History (years)	1,672	1.48	1.58	0.59	0.89	1.70
	durables	1,614	1.40	1.51	0.59	0.87	1.61
	non-durables	58	3.77	1.98	2.43	3.90	5.54
	Future (years)	1,672	1.35	0.11	1.37	1.37	1.37
	durables	1,614	1.35	0.11	1.37	1.37	1.37
	non-durables	58	1.37	0.00	1.37	1.37	1.37

Notes: The top panel contains the values for the Full sample, the bottom panel for the Dynamic sample, which is constrained by the availability of price time series. The lines with *Tests* in Column (1) refer to observations on the test level, those with *Products* refer to the level of products in the tests. *History* reports the time in years that a product is recorded on the price comparison website before the test publication date; *Future* reports the years the product remains listed on the website afterwards. The variable *Future* is censored at 500 days (1.37 years). *History* and *Future* have fewer observations in the Full sample because of non-availability on the price comparison website. Monetary values are in euro of January 2016.

5 Results

First, we estimate the cross-section relationship between quality and prices before quality information is published. This is a test of the main result of the signaling literature that firms signal quality through prices in experience goods markets. The second part of the results consists of the analysis of the Dynamic sample to answer what the price response is after the quality disclosure event, differentiated by quality. Following the theoretical literature, we differentiate between durables and non-durables. Then, we use Kaplan-Meier-estimates to see if quality affects market exit of products.

5.1 Cross-section relationship

The estimated relationship between price and quality in the cross-section is

$$\frac{P_{ij}}{\bar{P}_j} = \beta_0 + \sum_{g=2}^{5} \gamma_g \cdot 1[Grade_{ij} = g] + \nu_{ij} \tag{1}$$

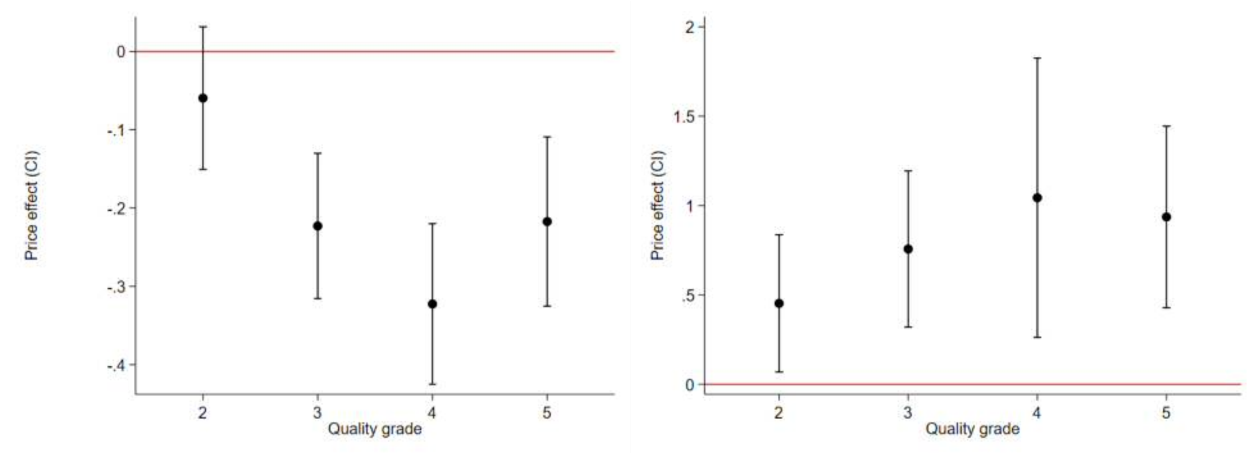
where the left-hand side denotes the price of product i in test j over the mean price of the products in the test. Prices used are those recorded before the tests are performed by Stiftung Warentest. γ_g are coefficients for the grades from 2 to 5 expressed as dummy variables, with grade 1 as the reference group.[8] Based on the discussion in Section 2.1, we expect high quality durable goods to have on average higher prices. For durable goods, reputation concerns play a less important role. For non-durable goods, we expect high quality goods to have lower prices. Note that we expect the coefficients γ_g to be negative for durable products and positive for non-durable products, because high quality means a low value in the test grade.

Figure 2 shows the estimates for the grades from 2 (second best quality) to 5 (worst quality) for durables (left) and non-durables (right); the numerical results are displayed in Table A.2. The left graph in the Figure shows that there is a strong positive relationship between quality and the relative price for durables. Products of grade 2 display the same negative tendency but the estimate is insignificant; for grade 3, prices are about one quarter lower, for grade 4, about one third. For products with the worst grade, prices increase relative to grade 4. For non-durable products, as shown on the right graph, the opposite pattern is observed. Relative prices are higher for lower quality products, except for the worst grade. Relative prices are 75 percent higher for grade 3 and twice as high for grade 4, before they slightly decrease relative

[8] We also included β_j, test fixed-effects, but the F-tests for both durables and non-durables where highly insignificant. This may not be surprising after dividing by the mean price in the test.

Figure 2: Price and quality regression coefficients and CI

Durables Non-durables



$\longleftarrow$ higher quality / lower quality $\longrightarrow$ $\longleftarrow$ higher quality / lower quality $\longrightarrow$

Notes: Regression estimates and 95 percent confidence intervals from regression of the prices relative to the test mean prices on grade dummies. Durable products are left, non-durable products right. Products of grade 1 (*Very good*) are the reference group.

to grade 4 for the worst grade.

In sum, both the descriptive results in Figure 1 and the results from the cross-sectional regressions confirm the theoretical predictions and suggest that prices fulfill a signaling role in an asymmetric information environment. The results show a positive relationship between prices and quality for durable goods; and the opposite pattern for non-durable products. Products of the worst quality imitate prices of products with higher quality. This makes the products of very bad quality difficult to distinguish from products of middle quality for consumers when simply relying on price. These results are largely consistent with separating equilibria of signaling models.

5.2 The Effect of Information on Prices

To investigate the informational content of the disclosure event, we apply an RD-like approach around the event of quality disclosure. Our implementation is similar to Kreiner et al. (2020). We use a dummy with parameter γ to estimate the event effect for each ordinal quality grade result. The publication of the test results defines the event. The basic specification which we estimate for each grade for durables and non-durables is given by

$$\frac{P_{it}}{P_{i,-105}} = \gamma \cdot 1[t > 0] + \sum_{k=0}^{k=4} \alpha_k t^k + \delta \cdot 1[0 \leq t \leq 4] + D_i\mu + \epsilon_{it} \tag{2}$$

with the dependent variable given by the price of each product relative to its price at the beginning of the event window. t is measured in relative time, going from -105 to 105 days after the event for each product i. α_k fits a fourth-order polynomial. δ makes it possible to allow for a time lag for the impact of the revealed quality information on prices. The time lag includes the day the results are published ($t = 0$), and 4 days afterwards. We view this as a possibility for the market to absorb the new information, in particular because it is assumed to take about 2-3 days to deliver the magazine to subscribers. D_i is a product fixed effect dummy. Based on the signaling literature we expect the parameter γ to be negative for high quality durable goods, because reducing information asymmetry about quality reduces the necessity to distort high quality products' prices upward. For high quality non-durable goods, we expect price increases — a positive estimate for γ — because of reduced necessity to distort prices downward.

Figure 3 shows the result of the polynomial and the shift of mean prices at publication/arrival of the quality information on the market. Grade 2 durable products, which have the largest sample size, exhibit a very strong and clear upward shift in mean price after the event, and a monotonic, almost linear downward trend both before and after the event. This downward trend, both before and after the event, is a well-known pattern in the theoretical literature of pricing of experience good (Bergemann and Välimäki (2006)). The upward shift is significant — see Table 4, Column (2), top panel — and estimated as .877 percentage points. A similar pattern holds for products of quality grade 1, with .787 percentage points upward shift. The effect for products of grade 3 is much lower, at .295 percentage points. For grade 4, we find a negative price effect with -.64 percentage points. For grade 5 products the effect is positive, which, as in the cross section relationship, reverses the expected price–quality relationship; it is insignificant, though. The differential effect of the event between the highest (grade 2) and the lowest (grade 4) significant effects is $.877 - (-.636) \approx 1.5$ percentage points.

For non-durables, shown in the graphs of the right column of Figure 3, and in the bottom panel of the regression results in Table 4, all significant results are the opposite of those for durables: for grade 2, the price effect is -1.6 percentage points, for grade 3 it is about -5 percentage points, and for grade 5 it is 1.4 percentage points. We include grade 1 and grade 4 for completeness, but give smaller weight in the interpretation to these results, because they are based on the time series of a single product each.

Overall, these estimates from the RD-model give mixed results compared to the predictions of the signaling literature regarding price dynamics. We see a significant increase in mean prices for durable products of very good, good and satisfactory quality, instead of a price decrease because of the reduction in information asymmetry. Also, we see a decrease in price for non-durables of satisfactory quality. These findings indicate that test results improve matching and consumers switch from low quality products to better quality products. The pattern of price changes for non-durables does not confirm theory, where high quality producers would start with low prices and when consumers become more informed, they would increase prices over time. An interesting observation is made for products of insufficient quality. Their prices rise both in the case of durable and non-durable products. One possible explanation is that producers of bad quality may imitate pricing of higher quality products to target consumers that are not aware of the product test results and that still rely on price as a signal. Hence, these producers exploit the uninformed to increase prices and imitate the behavior of good quality products.

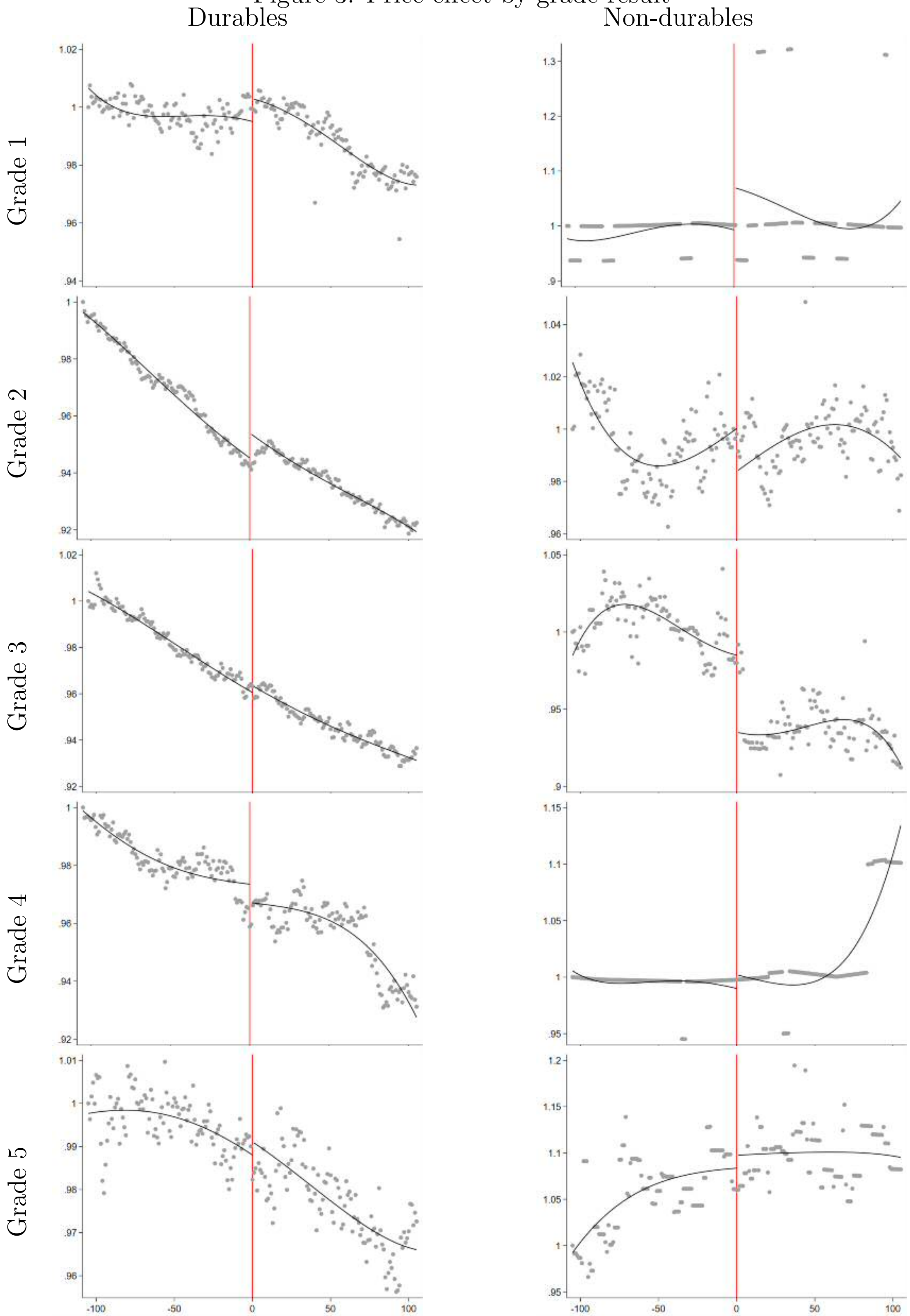
5.3 Market exit

Quality information disclosure can affect not only price setting, but also the decision to keep the product on the market. Bad test results could induce producers to take those products off the market. Because our price series are naturally censored, with more frequent censoring to be expected for younger products, we introduce a common censoring for all products at 500 days after the publication of the test. The design of the RD-window implies that the minimum survival time after the test was 105 days. We extend the data basis for the analysis in this Subsection to all observations that have market presence after the event, which corresponds to the 1,948 observations (see variable *Future* in Table 3).

Figure 4 shows the Kaplan-Meier survival estimates by grade on the left graph. Mostly, the hypothesis that a worse grade means earlier exit is confirmed. There is not a single exit for the products with the best quality, grade 1. Intermediate and very similar exit rates occur for grade 2 and 3, and an even higher rate for grade 4. Only grade 5, the worst grade, stands out from this negative survival–quality relationship, with no exits for about half of the time span considered and then ending up with the second-lowest exit rate.

Grade 4 and 5 have large confidence intervals for the Kaplan-Meier estimates that overlap with the point estimates of the other grades. The right graph of Figure 4 has the two groups, grade 2 and 3, which display very similar exit patterns, pooled together to show their 95 percent confidence interval in relation to the other grades. Grade 4 lies outside the confidence band; grade 5 approaches the band once exits start, but also lies outside the grade-2/3 band most of

Figure 3: Price effect by grade result



Notes: Mean prices, each product's price relative to its price at relative day -105. The black line is the result of the event estimation. The left column is for durable goods, the right column for non-durables. Each row of graphs is for one grade, starting from the best grade, 1, and going down to the worst grade, 5.

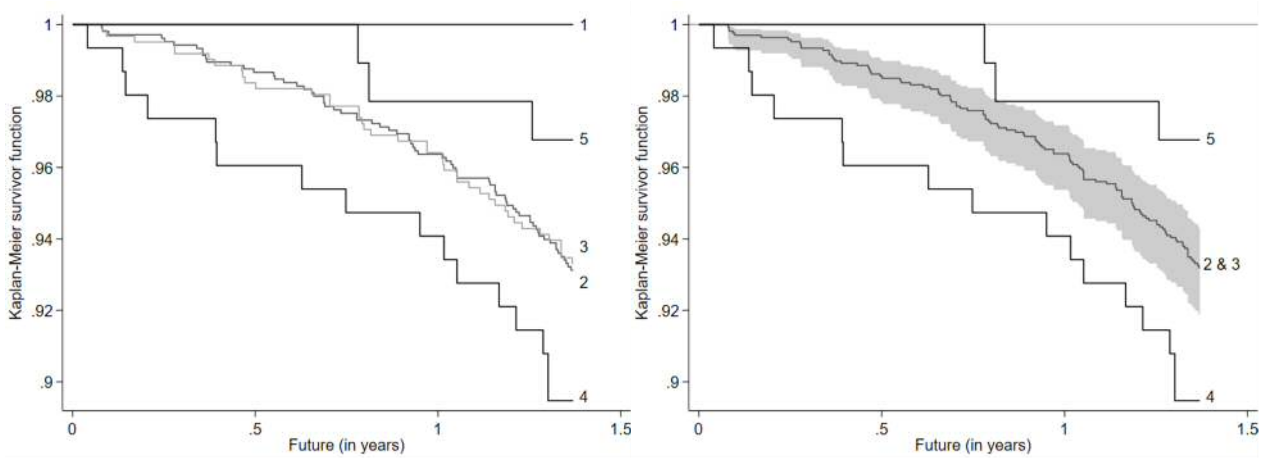
Table 4: Regression results for RD estimate

	(1) Grade 1	(2) Grade 2	(3) Grade 3	(4) Grade 4	(5) Grade 5
Durables					
$1[t > 0]$	**.00787****	**.00877****	**.00295***	**-.00636****	**.00305**
	(.00255)	**(.00091)**	**(.00130)**	**(.00231)**	**(.00326)**
t	-.00015**	-.00041**	-.00039**	-6.8e-05	-.00024**
	(4.2e-05)	(1.6e-05)	(2.2e-05)	(4.3e-05)	(5.3e-05)
t^2	-2.9e-06**	1.3e-06**	8.1e-07*	1.9e-07	-1.1e-06
	(5.0e-07)	(2.5e-07)	(3.4e-07)	(5.6e-07)	(7.1e-07)
t^3	-4.7e-09	5.5e-11	2.9e-09	-2.2e-08**	7.0e-09
	(3.9e-09)	(1.9e-09)	(2.4e-09)	(4.2e-09)	(5.4e-09)
t^4	1.9e-10**	-5.0e-11	-2.7e-11	-7.6e-11	3.2e-11
	(5.3e-11)	(2.8e-11)	(3.7e-11)	(5.9e-11)	(7.4e-11)
$1[0 \leq t \leq 4]$	-.00015	-.00802**	-.0018	-.0043	-.00656
	(.00313)	(.00137)	(.00189)	(.00258)	(.00362)
Constant	.99504**	.94522**	.96051**	.97354**	.98803**
	(.00156)	(.00056)	(.00077)	(.00137)	(.00190)
N	7,807	181,460	110,142	27,219	13,926
Non-durables					
$1[t > 0]$	.07699*	**-.01637****	**-.04989****	.01193**	**.01383**
	(.03555)	**(.00614)**	**(.01224)**	(.00352)	**(.01189)**
t	-.0008	.00043**	-.00024	-.00032**	.00013
	(.00042)	(.0001)	(.00016)	(6.0e-05)	(.00023)
t^2	-1.4e-05**	2.8e-07	8.3e-06**	-1.9e-06	-2.0e-06
	(5.2e-06)	(1.4e-06)	(2.0e-06)	(1.1e-06)	(3.5e-06)
t^3	6.9e-08*	-4.8e-08**	1.3e-08	7.9e-08**	2.7e-08
	(3.4e-08)	(1.1e-08)	(1.3e-08)	(6.3e-09)	(2.7e-08)
t^4	1.1e-09*	1.0e-10	-8.4e-10**	7.8e-10**	-2.0e-10
	(4.6e-10)	(1.6e-10)	(1.8e-10)	(1.1e-10)	(4.1e-10)
$1[0 \leq t \leq 4]$	-.08934*	.00645	.03698**	-.00106	-.03217**
	(.03542)	(.0067)	(.01124)	(.0033)	(.01079)
Constant	.99291**	1.0002**	.98499**	.99002**	1.0835**
	(.00821)	(.00356)	(.00533)	(.00174)	(.00738)
N	211	7,807	2,743	211	1,266

Notes: Each Column shows the regression results for the RD model for one grade, from best (grade 1) to worst grade (grade 5). The top panel is for durable goods, the bottom panel for non-durables. $1[t > 0]$, the main variable of interest, is put into bold face except where the estimate only relies on a single product. Product-fixed effects are included. Standard errors are clustered on calendar days. Standard errors are in parentheses below the coefficients. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$.

Figure 4: Kaplan-Meier survival by grades

All grades　　　　Aggregate Grade 2 & 3 with CI



Notes: The left graph shows Kaplan-Meier survival estimates by grade. The right graph pools grade 2 and 3 together and adds its 95 percent confidence interval.

the time, and also at the end of the censored time span. The pattern fits the idea that quality disclosure from the consumer protection agency is relevant, reduces asymmetric information and improves the average quality of products in the market, because bad products leave the market. Only for the worst grade products there is an effect reversion.

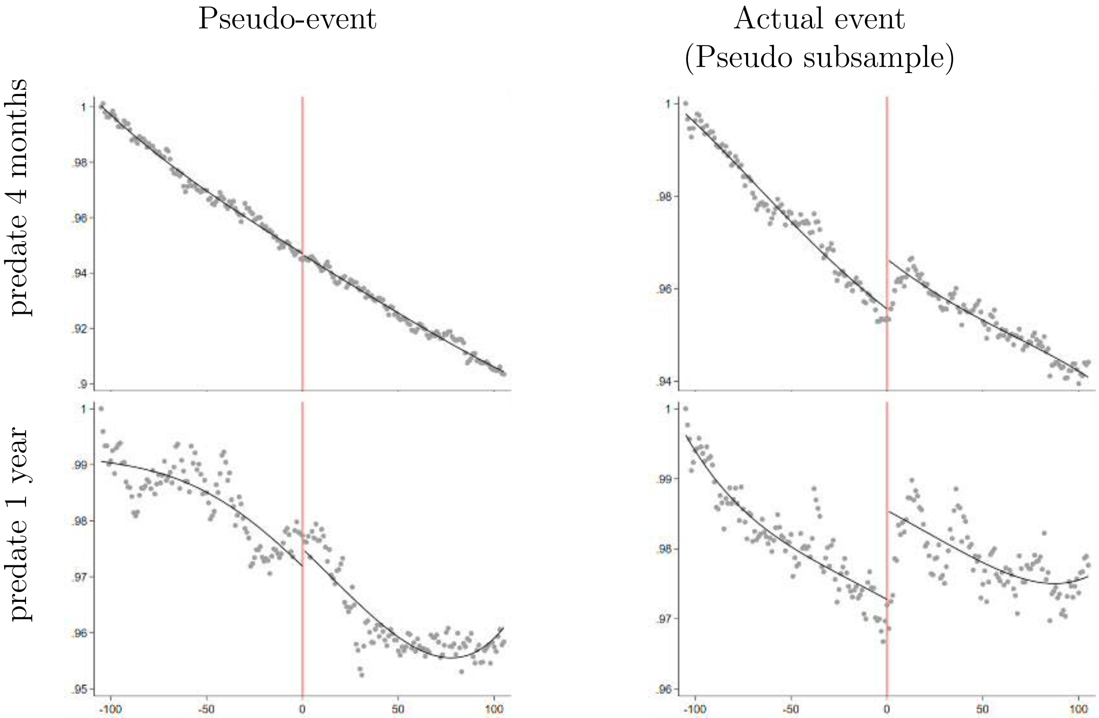
6 Robustness

Two pseudo-events are shown as evidence that the results are not driven by other factors than the event. Afterwards, we relax and vary some of the modelling choices that may drive the results. One varies the window length, the next one leaves out of the analysis observations immediately before and after the event, and finally, we experiment with different orders of the polynomial in time.

6.1 Pseudo-events

As a robustness check against confounding pricing-relevant artifacts, we create two pseudo-events. First, a pseudo-event predated to four months before the publication of the test magazine; second, predating one year. A shift of the event by four months is as close as possible to the actual event when shifting by full months, without the possibility that the actual event influences the results (105≈3.5 months). With the shift of one year we target to exclude any season-of-the-year effects that could affect pricing in general, or of a product group, or of a

Figure 5: Price effect for pseudo-events



Notes: Mean prices, each product's price relative to its price at relative day -105. The black line is the result of the event estimation. The left column shows the pseudo-events four months before the actual event (top panel) and one year before the actual event (bottom panel). All four graphs are for durables of grade 2. The sample of products differs from the main grade-2-durables results (Figure 3) and between the pseudo-events because of data availability of historical price series. To avoid sample selection effects, we reestimate each actual event with the subsample of products in the corresponding pseudo-event; the results for the actual event of quality publication with the same subsample as in the corresponding pseudo-event are shown in the right column. The top panel contains 609 products (128,499 obs.), the bottom panel 216 products (45,576 obs.).

specific product.

Figure 5 shows the two pseudo-events in the left column, compared to the actual event in the right column. Durables of grade 2, which is the largest group and displays the strongest effect of the event, are used in this robustness check. Because the number of products in the pseudo events is not the same as in the actual event due to price history data availability, we present for each pseudo-event also an additional graph for the actual event with the corresponding products to make the results more comparable. For the 4-months pseudo-event, the effect is close to zero and insignificant. Table 5 gives the numerical estimates for this pseudo-event in Column (1), and the actual event estimate for the same sample of products in Column (2). For the 1-year pseudo-event, the point estimate is a quarter of the actual event and also insignificant (Table 5, Column (3) and (4)). The low estimates and their insignificance for the pseudo-events support the validity of our event approach.

Table 5: Price effects for pseudo-events

	(1) Pseudo-event Predate 4 months	(2) Actual event (Pseudo subsample)	(3) Pseudo-event Predate 1 year	(4) Actual event (Pseudo subsample)
$1[t > 0]$	-.00038	.01076**	.00311	.01270**
	(.00108)	(.00117)	(.00180)	(.00227)
N	128,499	128,499	45,576	45,576

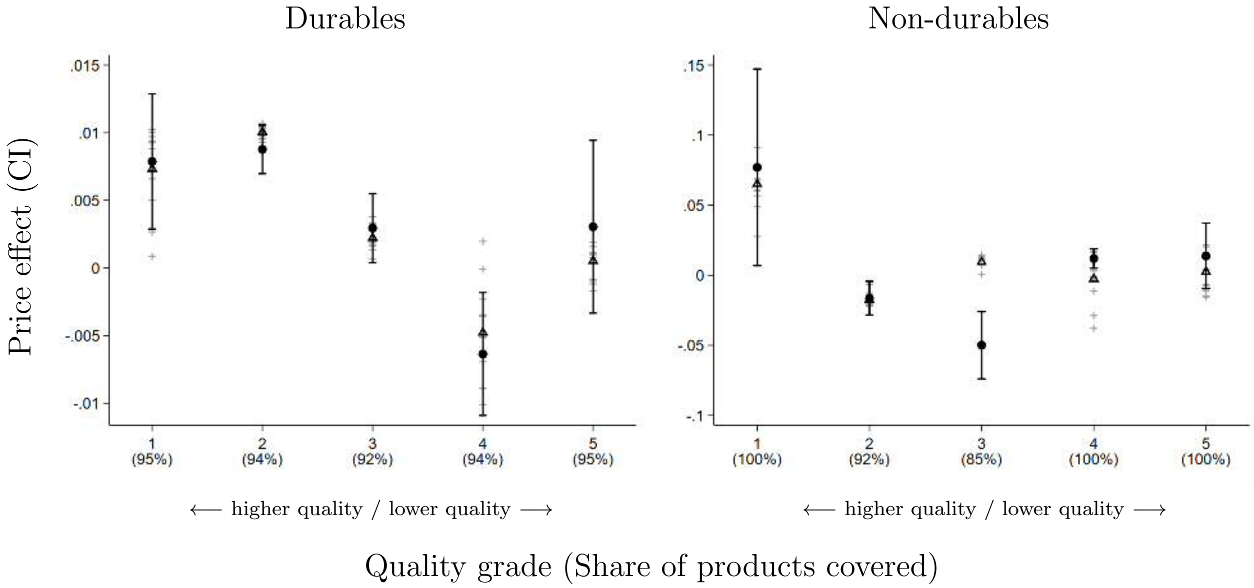
Notes: Price effects for predating to a pseudo-event four months and one year before the publication of the quality results. The specification is the same as in the main results, with the same covariates and clustering of the standard errors on calendar days. All four columns are for durables of grade 2. Column (1) shows the price effect for the pseudo-event 4 months before the actual event, Column (2) the main result constrained to the identical products as in the regression of Column (1); Column (3) contains the price effect for the pseudo-event 1 year before the actual event, Column (3) the main result constrained to the identical products as in Column (4). Significance levels: $*p < 0.05$, $**p < 0.01$.

6.2 Window length

The choice of the window length to be +/- 105 days (+/- 15 weeks) is a result of having a time span that is sufficiently long to estimate the event effect with precision and not losing too many products due to missing daily prices for longer time series. Also, 105 is just a multiple of 7, which was chosen as a way to preclude any artifacts from breaking weekly frequencies. In this Subsection, we vary the window length by changing it in units of full weeks. In particular, Figure 6 shows the event effect by grade and durability for variations of the window length from +/-10 weeks (+/-70 days) to +/-20 weeks (+/-140 days), instead of the original +/-15 weeks (in total, 11 window-lengths).

Most of the results are within the 95 percent confidence interval of the original estimate for the group. The solid circles shows the original estimates and the solid lines the confidence interval for each original estimate. Each grey plus-sign represents one alternative estimate for each of the 11 window-lengths from 10 to 20 weeks. The hollow triangles show the mean of the 11 estimates per grade/durability (the 11 grey plusses, one for each variant of the window length). The sample size is different from the original estimates, because with increasing window length, the number of products with missing daily prices increases. Within the results in Figure 6, the number of observations is kept constant at the level implied by the largest window length, +/-20 weeks. Below each grade in Figure 6 is shown the share of products that remains in this robustness check. For products of grade 3 we observe that the alternative estimates lie outside the original confidence interval. These does not necessarily imply that the results for products of sufficient quality are not robust to the window length. Note that

Figure 6: Price effect for windows from 10 to 20 weeks



Notes: On the left, for each grade of durable products, the event effect is measured by using a window from length 10 to length 20 weeks instead of the original 15 weeks (105 days window). The solid circle and the symmetric error bars around it show the original event estimate and its 95 percent CI. Grey plus-signs represent one alternative estimate for each of the 11 window-lengths from 10 to 20. Black hollow triangles show the mean of the 11 alternative estimates. The samples become smaller when longer windows are considered; the sample size is fixed at the longest window length of 20 weeks. The percentage numbers below each grade show the sample that remains relative to the original sample. On the right, the same analysis is shown for non-durables.

only 85 percent of grade 3 products remain in the subsample when varying the window length. Moreover, when one looks at the smaller window lengths which are not limited by the number of products — that is, lengths 10-15 weeks, where 100 percent of group's products are available per construction — then the alternative estimates are almost identical to the original estimate.

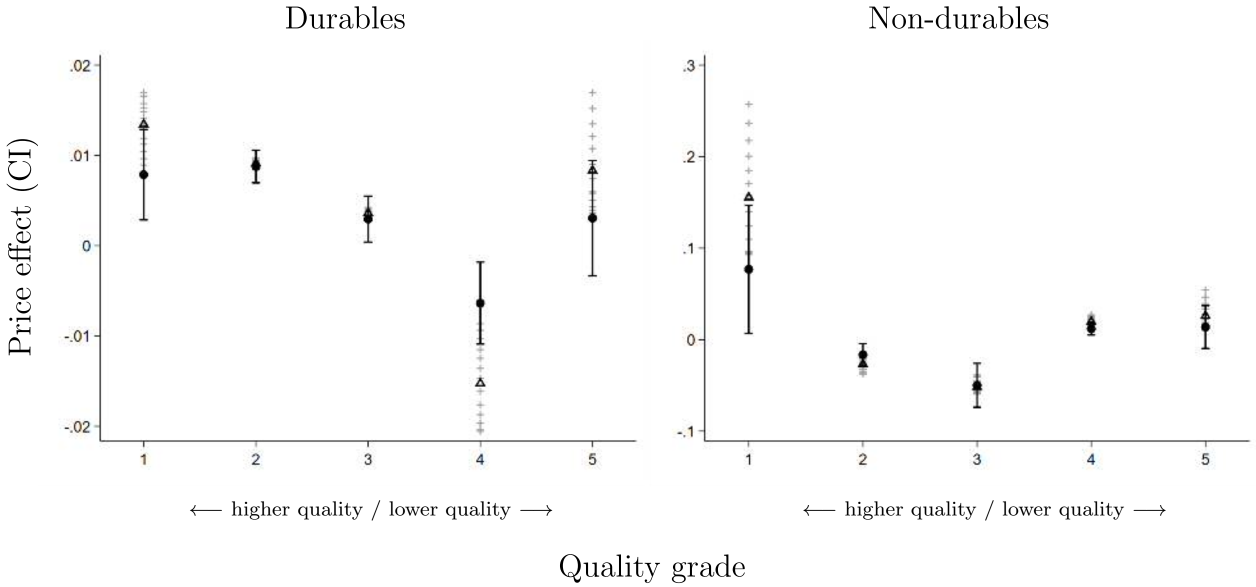
6.3 Cutting out close observations

The pattern of price changes in particular for the largest subsample, durables of grade 2, suggests that the price reaction takes a few days to diffuse. While we do include a 4-day post-event dummy to allow for adjustment time, it may be that a different lag is more appropriate. We relax this by dropping observations for a symmetric window of 1-14 days around the event. This "donut RD" is also useful to avoid any close-to-the-event effects before it occurs.[9]

Figure 7 shows the results in the same way as for the robustness check on window length above. For each grade/durability, 1 to 14 observations were dropped for each product and new

[9]This is similar to the robustness check in Barreca et al. (2011) or Dahl et al. (2014).

Figure 7: Price effect with 1-14 observations cut out around event



Notes: On the left, for each grade of durable products the event effect is measured by using a "donut RD" that symmetrically drops 1 to 14 daily prices around the event for all products. The solid circle and the symmetric error bars around it show the original event estimate and its 95 percent CI. Grey plus-signs represent one alternative estimate for each of the 14 donut RDs. Black hollow triangles show the mean of the 14 alternative estimates. On the right, the same analysis is shown for non-durables.

estimates obtained. The grey plus-signs represent one donut sample each. The solid circles and the lines show the original estimates. A hollow triangle gives the mean of all 14 donut sample estimates per grade and type of product. Several results are outside the 95 percent CI of the original estimate, and even the average result lies outside in some cases. However, the direction of the price change remains the same and the size of the effect is stronger than in the original estimates.

6.4 Degree of polynomial

Some of the graphs of the main results in Figure 3 exhibit complex patterns, which could make a fourth-order polynomial necessary. As a robustness check to this choice, we have estimated the model alternatively with polynomials of degree 1 to 5. Table 6 shows by grade and durability the results for the post-event shift, $1[t > 0]$, resulting from each specification. Column (4) reproduces the original estimates. Specifications where the highest-order polynomial term loses significance are in normal font, those where the highest-order term is significant are put in bold face. We give less weight in interpreting the robustness check to the specifications where the highest degree of the polynomial is insignificant.

Overall the results confirm our main results, although there is some variation. The term of order 5 is only significant for durables grade 1; all other specifications in the for this group give estimates that are very similar to the original estimate. Grade 2 durables, the largest group, has only order up to term 2 significant, but the estimates do not vary much between them. Grade-3-durables have estimates half as large for order 1 and 2, and are similar for 3 and 4, although in those cases the largest order polynomial of the specification is not significant. For durables grade 4, the estimates differ for order 1 and 2, but up to order 3 are significant, the last one having again a very similar estimate as order 4. Grade 5 durables show no significant results for the event effect.

For non-durables, all estimates are similar to the original ones when compared to the polynomial specification where the highest-order is significant.

7 Discussion and Conclusions

Our analysis empirically evaluates the effect of exogenous product quality information provided by a consumer protection agency in experience goods markets. Our identification approach allows us to attribute the price changes to the quality disclosure by the consumer agency. Theories of price signaling under asymmetric information predict that price distortions will be reduced when quality is disclosed. Our results show a substantial effect of quality disclosure on market outcomes for products which are tested by the agency.

Our cross-sectional results suggest that quality is reflected in prices, supporting the theoretical models that emphasize price as an important quality signal that firms employ to overcome asymmetric information problems. The dynamic estimates show that firms significantly change the price when the test results are published. However, the increase of the price after the event contradicts some of the main theoretical predictions of the asymmetric information literature that predict a price decrease, especially for the high-quality products.

Results for products that are revealed to be of very bad quality have some tendency to reverse the relationships we find between price and quality. Producers of bad quality could imitate pricing of higher quality products to target consumers that are uninformed — in the sense of not being aware of the product test results — and that still rely on price as a signal. Product survival, which is higher than that for the worst products only for the best products, is in line with the idea of such an imitation strategy being profitable.

Overall these results suggest positive welfare effects caused by large-scale quality disclosure of information in experience goods markets. The increasing prices for products of very good and

Table 6: Robustness check on polynomial

	(1) Order 1	(2) Order 2	(3) Order 3	(4) Order 4	(5) Order 5
<u>Durables</u>					
Grade 1	**0.00919****	**0.00982****	0.00731**	**0.00787****	**0.00059**
	(0.00198)	(0.00187)	(0.00264)	(0.00255)	(0.00316)
Grade 2	**0.00929****	**0.00883****	0.00892**	0.00877**	0.00938**
	(0.00078)	(0.00076)	(0.00092)	(0.00091)	(0.00124)
Grade 3	**0.00190**	**0.00158**	0.00303*	0.00295*	0.00112
	(0.00098)	(0.00098)	(0.00129)	(0.00130)	(0.00161)
Grade 4	**0.00405****	**0.00438****	**-0.00614****	-0.00636**	-0.00727**
	(0.00157)	(0.00158)	(0.00229)	(0.00231)	(0.00275)
Grade 5	**-0.00084**	**-0.00042**	0.00295	0.00305	-0.00210
	(0.00244)	(0.00244)	(0.00326)	(0.00326)	(0.00393)
<u>Non-durables</u>					
Grade 1	0.03974	0.04162	**0.07367***	**0.07699***	0.05295
	(0.02555)	(0.02609)	(0.03409)	(0.03555)	(0.03899)
Grade 2	0.00735	**0.00668**	**-0.01667****	-0.01637**	-0.01452
	(0.00460)	(0.00465)	(0.00612)	(0.00614)	(0.00771)
Grade 3	-0.05456**	-0.05469**	-0.04739**	**-0.04989****	-0.05176**
	(0.00832)	(0.00848)	(0.01207)	(0.01224)	(0.01469)
Grade 4	**-0.02443****	**-0.02772****	**0.00962***	**0.01193****	0.00793*
	(0.00636)	(0.00475)	(0.00474)	(0.00352)	(0.00318)
Grade 5	**-0.00118**	**0.00107**	0.01442	0.01383	0.00210
	(0.01065)	(0.01035)	(0.01216)	(0.01189)	(0.01371)

Notes: Each Column shows the regression results for the RD model for the variable $1[t > 0]$, the main variable of interest, for increasing polynomial order in variable time, t. Each line is for the subsample of one grade. Column (4) contains the original estimates. Estimates resulting from specifications where the highest-order term is insignificant are in normal font, those where the highest-order term is significant are put into bold face. The top panel is for durable goods, the bottom panel for non-durables. Product-fixed effects are included. Standard errors are clustered on calendar days. Standard errors are in parentheses below the coefficients. Significance levels: *< 0.05, **< 0.01.

good quality durables and decreasing prices for products of sufficient quality can be explained by improved matching of consumers: they switch to products of better quality. The price increase is not necessarily welfare reducing if consumers that value quality are willing to pay for high quality. More accurate welfare assessment would require a model where we specify the preferences of the consumers. However, the fact that bad quality products are overpriced emphasizes once more the positive welfare effects of the tests.

Another aspect that should be considered when evaluating the overall welfare effects is related to the changes in behavior of producers other than price. As shown in the discussion of the literature, such quality disclosure is associated with quality improvements (Dranove and Jin, 2010). Even though our data does not allow us to evaluate how quality evolves after the test event, our exit results suggests that bad test grades drive some bad products out of the market. This aspect is important especially when considering the importance of reputation in signaling quality. This can be expected to hold especially for multi-product firms or firms that sell many products under the same brand (umbrella branding), because they want to avoid that the bad image transmits to the other products sold under the same name.

References

Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review*, 44(3):1007–1040.

Anderson, M. and Magruder, J. (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *Economic Journal*, 122(563):957–989.

Armstrong, M. (2011). Economic models of consumer protection policies. *MPRA Paper*, 34773.

Armstrong, M., Vickers, J., and Zhou, J. (2009). Consumer protection and the incentive to become informed. *Journal of the European Economic Association*, 7(23):399–410.

Bagwell, K. and Riordan, M. H. (1991). High and declining prices signal product quality. *American Economic Review*, 81(1):224–239.

Barreca, A. I., Guldi, M., Lindo, J. M., and Waddell, G. R. (2011). Saving babies? Revisiting the effect of very low birth weight classification. *Quarterly Journal of Economics*, 126(4):2117–2123.

Beales, H., Craswell, R., and Salop, S. C. (1981). The efficient regulation of consumer information. *Journal of Law & Economics*, 24(3):491–539.

Bergemann, D. and Välimäki, J. (2006). Dynamic pricing of new experience goods. *Journal of Political Economy*, 114(4):713–743.

Board, O. (2009). Competition and disclosure. *Journal of Industrial Economics*, 57(1):197–213.

Brown, J. and Jansen, M. (2020). Consumer protection laws in auto lending.

Cabral, L. and Hortacsu, A. (2010). The dynamics of seller reputation: evidence from ebay. *Journal of Industrial Economics*, 58(1):54–78.

Caves, R. E. and Greene, D. P. (1996). Brands' quality levels, prices, and advertising outlays: empirical evidence on signals and information costs. *International Journal of Industrial Organization*, 14(1):29–52.

Curry, D. J. and Riesz, P. C. (1988). Prices and price/quality relationships: A longitudinal analysis. *Journal of Marketing*, 52(1):36–51.

Dahl, G. B., Løken, K. V., and Mogstad, M. (2014). Peer effects in program participation. *American Economic Review*, 104(7):2049–74.

Daughety, A. F. and Reinganum, J. F. (2008). Imperfect competition and quality signalling. *RAND Journal of Economics*, 39(1):163–183.

Dellavigna, S. and Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *Journal of Finance*, 64(2):709–749.

Destatis (2019). *Consumer price index for Germany, Weighting pattern for base year 2015.* Statistisches Bundesamt, Germany.

Dranove, D. and Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4):935–963.

Dranove, D., Kessler, D., McClellan, M., and Satterthwaite, M. (2003). Is more information better? The effects of Report Cards on health care providers. *Journal of Political Economy*, 111(3):555–588.

Farrell, J. v. R. (1981). *Prices as signals of quality.* PhD thesis, University of Oxford.

Freedman, S., Kearney, M., and Lederman, M. (2012). Product recalls, imperfect information, and spillover effects: Lessons from the consumer response to the 2007 toy recalls. *Review of Economics and Statistics*, 94(2):499–516.

Friberg, R. and Grönqvist, E. (2012). Do expert reviews affect the demand for wine? *American Economic Journal: Applied Economics*, 4(1):193–211.

Gerstner, E. (1985). Do higher prices signal higher quality? *Journal of Marketing Research*, 22(2):209–215.

Greenstone, M., Oyer, P., and Vissing-Jorgensen, A. (2006). Mandated disclosure, stock returns, and the 1964 Securities Acts amendments. *Quarterly Journal of Economics*, 121(2):399–460.

Grossman, S. (1981). The informational role of warranties and private disclosure about product quality. *Journal of Law & Economics*, 24(3):461–83.

Grossman, S. J. and Hart, O. D. (1980). Disclosure laws and takeover bids. *Journal of Finance*, 35(2):323–334.

Hilger, J., Rafert, G., and Villas-Boas, S. (2011). Expert opinion and the demand for experience goods: An experimental approach in the retail wine market. *Review of Economics and Statistics*, 93(4):1289–1296.

Hollenbacher, A. and Yerger, D. (2001). Third party evaluations and resale prices in the US used vehicle market. *Applied Economics Letters*, 8(6):415–418.

Ippolito, P. M. and Mathios, A. (1990). Information, advertising and health choices: A study of the cereal market. *RAND Journal of Economics*, 21(3):459–480.

Janssen, M. C. and Roy, S. (2010). Signaling quality through prices in an oligopoly. *Games and Economic Behavior*, 68(1):192 – 207.

Jin, G. Z. and Kato, A. (2006). Price, quality, and reputation: evidence from an online field experiment. *RAND Journal of Economics*, 37(4):983–1005.

Jin, G. Z. and Leslie, P. (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. *Quarterly Journal of Economics*, 118(2):409–451.

Jovanovic, B. (1982). Truthful disclosure of information. *Bell Journal of Economics*, 13(1):36–44.

Klein, T., Lambertz, C., and Stahl, K. O. (2016). Market transparency, adverse selection, and moral hazard. *Journal of Political Economy*, 124(6):1677 – 1713.

Koning, P. and van der Wiel, K. (2013). Ranking the schools: How school-quality information affects school choice in the Netherlands. *Journal of the European Economic Association*, 11(2):466–493.

Kreiner, C. T., Reck, D., and Skov, P. E. (2020). Do lower minimum wages for young workers raise their employment? Evidence from a Danish discontinuity. *Review of Economics and Statistics*, 102(2):339–354.

Lewis, G. (2011). Asymmetric information, adverse selection and online disclosure: The case of ebay motors. *American Economic Review*, 101(4):1535–1546.

Mathios, A. D. (2000). The impact of mandatory disclosure laws on product choices: An analysis of the salad dressing market. *Journal of Law & Economics*, 43(2):651–678.

Milgrom, P. and Roberts, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy*, 94(4):796–821.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *Bell Journal of Economics*, 12(2):380–391.

Pope, D. G. (2009). Reacting to rankings: Evidence from 'America's Best Hospitals'. *Journal of Health Economics*, 28(6):1154–1165.

Reinstein, D. A. and Snyder, C. M. (2005). The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *Journal of Industrial Economics*, 53(1):27–51.

Riesz, P. C. (1979). Price-quality correlations for packaged food products. *Journal of Consumer Affairs*, 13(2):236–247.

Shapiro, C. (1983). Optimal pricing of experience goods. *Bell Journal of Economics*, 14(2):497–507.

Stiftung Warentest (2014). So testet und bewertet die Stiftung Warentest [This is how Stiftung Warentest tests and evaluates]. `https://www.test.de/unternehmen/testablauf-5017344-0/`.

Stiftung Warentest (2019). Jahresbericht 2019 [Annual report]. `http://www.test.de`.

Tirole, J. (1988). *The Theory of Industrial Organization*. MIT Press.

Vickers, J. (2004). Economics for consumer policy. In *Proceedings of the British Academy Volume 125, 2003 Lectures*, pages 287–310.

Werner, R. M. and Asch, D. A. (2005). The unintended consequences of publicly reporting quality information. *Journal of the American Medical Association*, 293(10):1239–1244.

Wolinsky, A. (1983). Prices as signals of product quality. *Review of Economic Studies*, 50(4):647–658.

A Appendix

Figure A.1: Sample Test Issue January 2017

Elektrische Zahnbürsten

Elektrische Zahnbürsten: Eine putzt sehr gut

Produkt	Braun Oral-B Pro 3000[1] D20.524.3M	Braun Oral-B Pro 600[1] D16.513	Philips Sonicare FlexCare Platinum 6 series HX9111/20	Panasonic Sonic Vibration Toothbrush[3] EW DL 83	Rossmann/ Perlodent med Akku-Zahnbürste Premium[4]	Philips Sonicare Clean-Care+ 1 series HX3212/01	dm/ Dontodent Active Professional Akku-Zahnbürste	AEG Elektrische Akku-Zahnbürste EZ 5623	Müller/ Sensident Sensi Sonic Schallzahnbürste 7146	Go Travel Toothbrush Sonic Action[7]
Art der Bürstenbewegung	Oszillierend-rotierend und pulsierend	Oszillierend-rotierend und pulsierend	Vibration (Schall)	Vibration (Schall)	Oszillierend-rotierend	Vibration (Schall)	Oszillierend-rotierend	Oszillierend-rotierend	Vibration (Schall)	Vibration (Schall)
Mittlerer Preis ca. (Euro)	76	29	152	179	18	31	16	23	25	14
Mittlerer Preis für eine Ersatzbürste ca. (Euro)	3.90	3.90	7.50[2]	6.50	1.60	5.00[5]	1.74	2.49	1.98	5.00
test-QUALITÄTSURTEIL 100%	**GUT (1,7)**	**GUT (1,9)**	**GUT (1,9)**	**GUT (2,3)**	**GUT (2,4)**	**BEFRIEDIGEND (2,9)**	**BEFRIEDIGEND (3,2)**	**BEFRIEDIGEND (3,3)**	**AUSREICHEND (4,0)**	**AUSREICHEND (4,4)**
Zahnreinigung 60%	sehr gut (1,5)	gut (1,8)	gut (1,7)	gut (2,5)	gut (2,3)	befried. (2,9)*)	befriedigend (3,2)*)	befriedigend (3,3)*)	ausreichend (4,0)*)	ausreichend (4,4)*)
Handhabung 30%	gut (2,4)	gut (2,4)	gut (2,5)	gut (2,3)	befriedigend (2,6)	gut (2,5)	befriedigend (2,6)	befriedigend (3,1)	befriedigend (2,6)	ausreichend (4,1)
Zähneputzen (z. B. Handlichkeit, Benutzung an schwer erreichbaren Stellen, Geräusch)	+	+	○	+	○	○	○	○	○	○
Putzzeitsignal (Timer)	+	+	+	+	+	+	+	–[6]	+	–[6]
Gebrauchsanleitung	+	+	+	+	+	+	+	○	+	–*[8]
Reinigung des Geräts	○	○	○	+	○	+	○	⊖	○	○
Stand- und Liegefestigkeit	○	○	○	○	○	+	+	+	○	–
Betriebszeit mit einer Akku- oder Batterieladung (min)	+/48	○/42	++/128	++/113	+/74	++/95	+/52	++/95	++/93	++/756
Betriebsart/Batteriewechsel	Akku/Entfällt	Akku/Entfällt	Akku/Entfällt	Akku/Entfällt	Akku/Entfällt	Akku/Entfällt	Akku/Entfällt	Akku/Entfällt	Akku/Entfällt	Batterie/○
Haltbarkeit, Umwelteigenschaften 10%	sehr gut (1,0)	sehr gut (1,0)	sehr gut (1,1)	sehr gut (1,1)	befried. (3,0)	sehr gut (1,1)	sehr gut (1,1)	sehr gut (1,0)	sehr gut (1,2)	befriedigend (3,0)
AUSSTATTUNG / TECHNISCHE MERKMALE										
Ausstattung, Zubehör	2 verschiedene Bürstenköpfe, Ladestation, optische Andruckkontrolle	1 Bürstenaufsatz, Ladestation	1 Bürstenaufsatz, Schutzkappe, Ladestation, Reisebox	3 verschiedene Bürstenaufsätze, Ladestation, Reisebox	1 Bürstenaufsatz, Schutzkappe, Ladestation	1 Bürstenaufsatz, Schutzkappe, Ladestation	1 Bürstenaufsatz, Ladestation, optische Andruckkontrolle	4 Bürstenaufsätze, Ladestation, separate Wandhalterung mit Befestigungsmaterial	1 Bürstenaufsatz, Ladestation	2 Bürstenköpfe, 1 AAA-Batterie
Gewicht der Zahnbürste ca. (g)	129	122	141	99	119	148	134	104	44	41
Bürstenaufbewahrung/Anzahl der Bürsten	■/2	□	□	■/4	■/2	□	■/1	■/4	■/2	□
Anzahl der Putzeinstellungen	3 (Reinigen, Sensitive, Zahnfleischreinigung)	1	3 Putzeinstellungen (Clean, White, Deep Clean) in je 3 Intensitätsstufen	2 (Normal, Sanft)	3 (Aktiv, Sensitiv, Massage)	1	2 (Reinigung, Sensitive)	1	3 (Normal, Soft, Pulse)	1
Putzzeitsignal: Intervalltimer	■	■	■	■	■	■	■	□	■	□
Putzzeitsignal: Ende der Putzzeit	■	■	■	■	■	■	■	□	■	□
Ladekontrollleuchte vorhanden	■	■	■	■	■	■	■	■	■	□

Bewertungsschlüssel der Prüfergebnisse:
++ = Sehr gut (0,5–1,5). + = Gut (1,6–2,5).
○ = Befriedigend (2,6–3,5). ⊖ = Ausreichend (3,6–4,5).
– = Mangelhaft (4,6–5,5).

■ = Ja.
□ = Nein.

***) Führt zur Abwertung** (siehe „So haben wir getestet" auf Seite 18).
Bei gleichem Qualitätsurteil Reihenfolge nach Alphabet.
1) Getesteter Bürstenaufsatz: Oral-B Cross Action.
2) In der 4er-Packung. Auch 2er-Packung (20 Euro) erhältlich.
3) Getesteter Bürstenaufsatz: Multifunktionsbürste.
4) Produkt laut Anbieter inzwischen verändert.
5) In der 8er-Packung. Auch andere Gebindegrößen (25 Euro/4 Stück und 15 Euro/2 Stück) erhältlich.
6) Kein Putzzeitsignal vorhanden.
7) Bezogen über Pro-Idee.
8) Keine deutschsprachige Gebrauchsanleitung vorhanden; Angaben zum Einlegen der Batterie nicht korrekt.

test 1/2017 1/2017 test Ernährung und Kosmetik 21

Notes: The Figure shows a sample table collecting all results from a test of electric toothbrushes. We have added a black frame around the row containing the aggregate test result. The yellow background, printing in bold face and the red emblem of the magazine in the beginning of the row, which draw attention to the grades are already in the original print.

Table A.1: Data coverage of CPI weights (per mil)

2-digit			3-digit			Sample	
	Name	Weight		Name	Weight	Weight	N
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
01	Food, non-alc. bev.	96.9	011	Food	84.87	0.13	15
			012	Non-alcoholic bev.	11.98		
02	Alcohol, tobacco	37.8			37.77		
03	Clothing and footwear	45.3	031	Clothing	35.56	1.46	56
			032	Footwear	9.78		
04	Housing	324.7			324.7		
05	Furniture, household	50.0	051	Furniture, lighting, floor cover	19.42	0.40	16
			052	Household textiles	3.95		
			053	Household appliances	8.8	7.46	521
			054	Glass-/tableware	3.68	0.57	44
			055	Tools/appliances house/garden	6.45	2.45	206
			056	Housekeeping goods/services	7.74		
06	Health	46.1	061	Medical products/equipment	19.42	3.75	43
			062	Outpatient services	20.22		
			063	Hospital services	6.49		
07	Transport	129.1	071	Purchase of vehicles	34.66	2.05	47
			072	Goods and services for vehicles	70.7	0.36	153
			073	Transport services	23.69		
08	Communication	26.7	081	Postal and parcel services	1.84		
			082	Telephones, communic. devices	2.66	2.66	175
			083	Telecommunication services	22.22	10.09	61
09	Recreation, entertain-, ment, culture	113.4	091	Photogr./info-processing equ.	14.18	10.12	854
			092	Other durables recreation/culture	2.34	0.29	10
			093	Oth. recreation/garden, pets	17.64	1.69	96
			094	Recreational/cultural services	37.41		
			095	Print, writing, drawing	15.17		
			096	Package holidays	26.62		
10	Education	9.0			9.02		
11	Restaurant, accommod.	46.8			46.77		
12	Miscellaneous	74.3	121	Personal care	22.88	3.53	244
			123	Personal effects n.e.c.	6.16	0.72	195
			124	Social protection services	14.17		
			125	Insurance services	24.68		
			126	Financial services n.e.c.	2.07		
			127	Other services n.e.c.	4.29		
Σ		1,000			1,000	47.72	2,736

Notes: The Table shows the representation of the products in our data in the German CPI. Columns (1)-(3) show code, name and weight for the categories of the top-level code; Columns (4)-(6) show the first subcategory for those 8 of 12 top-level categories where our products fall into. The underlying matching of our product types and the calculation of the corresponding weights representation in Column (7) was done on the lowest level of product codes available in the weighting pattern, which are 645 categories. Column (8) reports the number of observations that our data contain for each subcategory. Names of the categories are abbreviated (see Destatis, 2019).

Table A.2: Price-quality regression results

	Grade				Const.	R-sq.	N
	2	3	4	5			
Durables	-0.0595	-0.2229**	-0.3225**	-0.2173**	1.1415**	0.0565	2503
	(0.0465)	(0.0473)	(0.0523)	(0.0551)	(0.0451)		
Non-durables	0.4525*	0.7566**	1.0434**	0.9360**	0.4726**	0.0799	233
	(0.1949)	(0.2217)	(0.3965)	(0.2579)	(0.1803)		

Notes: Regression estimates (and standard errors in parentheses) from regression of the prices relative to the test mean prices on grade dummies. Separate regressions for durable and non-durable products were estimated. Products of Grade 1 (*Very good*) are the reference group. Significance levels: $^*p < 0.05$, $^{**}p < 0.01$.