MASTERARBEIT | MASTER'S THESIS

Titel | Title

Price Dispersion Puzzle: A Quantitative Analysis of Relative Thinking

verfasst von | submitted by

Aleksandar Vasic BSc.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 913
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Applied Economics
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Christine Zulehner Privatdoz.

Abstract

Classical search theory predicts that price dispersion for homogeneous goods should be independent of a product's base price, assuming search costs are constant. Yet empirical studies consistently reveal that price dispersion increases with product price—a contradiction known as the "Price Dispersion Puzzle". This thesis investigates this phenomenon in the context of the German online electronics market, where digital transparency should actually minimize search frictions. Leveraging a self-built web-scraping infrastructure, over 5 million price quotes for approximately 170,000 electronic products from a major price comparison platform have been collected. The analysis proceeds in three stages. First, a robust positive relationship between mean prices and price dispersion across and within product categories is documented, confirming the existence of the Price Dispersion Puzzle at scale. Second, consumer search costs are estimated from observed price distributions using a game-theoretic mixed-strategy equilibrium framework. Third, it is demonstrated that these inferred search costs increase systematically with product prices. Importantly, this systematic increase in search costs provides large-scale empirical support for the behavioral theory of "Relative Thinking," wherein consumers evaluate potential savings in relative rather than absolute terms. The findings suggest that Relative Thinking can partially explain the emergence of the Price Dispersion Puzzle, even in markets characterized by low technological search frictions.

Kurzfassung

Die klassische Suchtheorie prognostiziert, dass die Preisdispersion bei homogenen Gütern unabhängig vom Durchschnittspreis eines Produkts sein sollte, sofern die Suchkosten konstant sind. Empirische Studien zeigen jedoch durchgehend, dass die Preisdispersion mit dem Produktpreis zunimmt – ein Widerspruch, der als „Price Dispersion Puzzle“ bekannt ist. Diese Arbeit untersucht dieses Phänomen im Kontext des deutschen Online-Marktes für Elektronikprodukte, in dem digitale Transparenz Suchfriktionen eigentlich minimieren sollte. Mithilfe einer selbst entwickelten Web-Scraping-Infrastruktur wurden über 5 Millionen Preisangebote für rund 170.000 elektronische Produkte von einer großen Preisvergleichsplattform erhoben. Die Analyse erfolgt in drei Schritten. Erstens wird ein robuster positiver Zusammenhang zwischen Durchschnittspreisen und Preisdispersion innerhalb und zwischen Produktkategorien dokumentiert, was die Existenz des Price Dispersion Puzzle in bestätigt. Zweitens werden die Suchkosten der Konsument:innen anhand beobachteter Preisverteilungen unter Verwendung eines spieltheoretischen Modells mit gemischten Strategien geschätzt. Drittens wird gezeigt, dass diese geschätzten Suchkosten systematisch mit dem Produktpreis ansteigen. Entscheidend ist, dass dieser systematische Anstieg der Suchkosten groß angelegte empirische Evidenz für die verhaltensökonomische Theorie des „Relativen Denkens“ liefert, wonach Konsument:innen potenzielle Einsparungen relativ zum Basispreis und nicht in absoluten Beträgen bewerten. Die Ergebnisse legen nahe, dass Relatives Denken die Entstehung des Price Dispersion Puzzle zumindest teilweise erklären kann – selbst in Märkten, die durch geringe technologische Suchfriktionen gekennzeichnet sind.

Contents

List of Tables

List of Figures

1. Introduction

In classical economic theory, particularly within the framework of search models developed by Stigler (1961)[38], price dispersion in markets for homogeneous goods is attributed to imperfect information. A central theoretical prediction emerging from these models is that the absolute level of price dispersion should remain independent of the product's mean price. In other words, if search costs are constant across goods, then price dispersion should not systematically increase with the base price.

This prediction seems especially plausible in the context of digital markets. With price and product information readily accessible through search engines and comparison websites, the internet was expected to dramatically lower consumer search costs. As a result, many anticipated that online markets would become highly competitive, leading to more uniform pricing for identical goods. However, this expectation has not been borne out in practice. Despite the ease of obtaining price information online, substantial price dispersion persists. For example, Gorodnichenko et al. (2018)[18], using a large dataset of online prices across various consumer markets, find that the average ratio between the highest and lowest price for the same product is 1.65 in the United States and 1.52 in the United Kingdom.

Moreover, empirical research across numerous digital product categories from books and electronics to prescription drugs consistently documents a positive correlation between a product's mean price and its price dispersion (e.g., Baye et al., 2004[7]; Sorensen, 2000[36]; Xing, 2010[43]). This pattern, known as the "Price Dispersion Puzzle" (Azar, 2013[6]), challenges the predictions of standard search models and suggests the need for a more nuanced understanding of consumer search behavior in online markets.

One promising explanation for this puzzle is rooted in behavioral economics: consumers may not evaluate absolute price differences in isolation but rather respond to the relative magnitude of savings. This phenomenon is known as Relative Thinking (Azar, 2007[4]; 2011[5]). Under this behavioral lens, a $5 saving matters more when purchasing a $25 item than a $1,000 product, despite the identical absolute gain. This suggests that consumers have increasing search costs as the total price rises which might partially explain the Price Dispersion Puzzle.

Using a novel dataset of over five million online price observations across approximately 170,000 electronic products, this work pursues three objectives. First, it examines whether the Price Dispersion Puzzle can be observed at scale in German digital electronics markets. Second, it estimates effective consumer search costs by applying a game-theoretic model to a given price distribution. Third, it tests the hypothesis that search costs themselves scale with product price, thereby providing empirical evidence for the presence of Relative Thinking in digital consumer behavior.

Two main findings emerge from this thesis. First, price dispersion increases substantially with rising mean prices. Its magnitude varies after analyzing for category heterogeneity. Second, estimated search costs also rise with the average product price, consistent with the theory of Relative Thinking. While the estimation relies on a game-theoretical model with strong assumptions, the results align well with findings from alternative search cost estimation methods and experimental outcomes.

The thesis is structured as follows: It begins with a review of the relevant literature on price dispersion, search cost estimation methods and Relative Thinking. This is followed by a detailed description of the dataset. Next, I apply various established measures of price dispersion to examine whether the Price Dispersion Puzzle is present in the data. That is, whether price dispersion increases with mean price. The subsequent section introduces the estimation procedure for consumer search costs, which are then estimated across different product groups. Building on this, I test whether search costs themselves rise with mean prices, providing empirical evidence on the presence of Relative Thinking. The thesis concludes with a discussion that situates the findings within the broader literature, addresses limitations, and considers alternative explanations.

2. Related Literature

The following chapter synthesizes the most relevant theoretical and empirical contributions, structured around three closely connected research areas. First, the literature on price dispersion examines why identical goods are sold at different prices and seeks to uncover the underlying mechanisms, such as market structure, firm strategy, and consumer behavior. Second, research on search cost estimation explores how these costs can be inferred from observed price distributions, typically within the framework of formal theoretical models. Finally, the literature on relative thinking and context effects offers complementary insights, often through experimental methods, by investigating how consumers' perceptions and comparisons influence their willingness to search and pay—thereby affecting observed price variation. Together, these fields form the theoretical and empirical foundation for the questions addressed in this thesis.

2.1. Price Dispersion

Theoretical Foundations

The theoretical literature on price dispersion originates with Stigler (1961)[38], who was among the first to highlight that substantial price differences can persist even for homogeneous goods. Analyzing car and coal prices from 1959, he found considerable variation: car prices ranged from $2,350 to $2,515, while anthracite coal prices varied between $15.46 and $18.49 per unit. Stigler attributed this dispersion not only to service-related differences but also to consumers' limited access to information, introducing the idea that search costs play a fundamental role in shaping price variation. This insight was further formalized in the Diamond Paradox (Diamond, 1971)[14], which demonstrated that even minimal search costs could eliminate price competition entirely. In Diamond's model, if obtaining additional price quotes is costly, consumers may settle for the first observed price, enabling all firms to charge the monopoly price in equilibrium. This paradox triggered a wave of theoretical responses seeking to explain how price dispersion could persist under more realistic conditions. One such response came from Salop and Stiglitz (1977)[32], who introduced consumer heterogeneity into the model.

They distinguished between informed and uninformed consumers, showing that firms exploit this asymmetry by offering lower prices ("bargains") to informed shoppers while charging higher ("rip-off") prices to those lacking information. This differentiation leads to price dispersion as an equilibrium outcome. Varian (1980)[41] built on this idea but shifted focus to the temporal dimension of pricing. His model also incorporates informed and uninformed consumers, but he demonstrates that firms optimally employ a mixed strategy over time where they periodically lower prices to attract informed consumers and raising them at other times to profit from the uninformed. This intertemporal price discrimination leads to persistent price variation, even among identical firms. The emergence of formal search-theoretic models further advanced the understanding of price dispersion. A key distinction in this literature is between non-sequential (fixed-sample) and sequential search. Burdett and Judd (1983)[10] developed a non-sequential model in which consumers decide in advance how many price quotes to obtain, choosing the lowest among them. Even with identical firms and consumers, price dispersion arises endogenously from the costly nature of search. Stahl (1989)[37], in contrast, introduced a sequential search framework where consumers decide after each price whether to continue searching or to buy. By incorporating heterogeneity in search costs, assuming some consumers face no costs while others do, Stahl showed that the degree of competition and resulting price dispersion depends on the proportion of low-cost searchers. As the share of informed consumers increases, firms face more pressure to lower prices, leading to a narrower price distribution. Together, these foundational models establish a coherent theoretical basis for understanding price dispersion in markets with search frictions. They show that even in environments with identical products and many sellers, variation in consumer behavior and search technology can sustain persistent pricing differences.

Empirical Findings

Following the development of foundational theoretical models, empirical research has focused on measuring the extent of price dispersion and identifying its main drivers. A central distinction in this context is between online and offline markets, as key factors, especially search costs, differ substantially across these settings (Ratchford et al., 2022)[31]. An early empirical contribution by Pratt et al. (1979)[30] used data from the Boston telephone directory to examine prices in nearly competitive markets such as bicycles, peanuts, liquor, and pet grooming. Regressing the log of the standard deviation on the log of the mean price yielded a coefficient of 0.892, indicating that doubling the average price increases the standard deviation by roughly 86% ($2^{0.892} \approx 1.86$). The authors suggested that, beyond search costs, lower purchase frequency for more expensive goods could reduce the incentive to search, contributing to higher dispersion. Building on this, Sorensen (2000)[36] tested one of Stigler's (1961)[38] hypotheses where repeat buyers conduct more effective searches. Using a large dataset from the U.S. pharmaceutical market, he showed that price ranges for frequently purchased maintenance drugs were about 30% lower than for one-time-use prescriptions. This supports the idea that consumers are more likely to search when the benefits of recurring cost savings are greater. Kaplan and Menzio (2015)[22] analyzed over 300 million transactions from 50,000 U.S. households between 2004 and 2009. They found that the distribution of prices is typically symmetric and leptokurtic, with standard deviations for homogeneous goods ranging from 19% to 36%. Using a decomposition method, they showed that only 10% of price variation is attributable to differences in store expensiveness. The remaining variation is evenly split between within-store price fluctuations and differences in average prices across equally expensive stores. A significant share of this unexplained variation between 35% and 55% is attributed to search frictions, highlighting their central role in sustaining price dispersion.

Focusing on online markets, Gorodnichenko et al. (2018)[18] examined around 50,000 price observations from the UK and the US (2010–2012), collected via a large price comparison website. Despite low search costs for consumers and low menu costs for firms, they documented substantial price dispersion—14.9 log-points in the UK and 17.5 in the US. Seller fixed effects accounted for 25–30% of dispersion in the US and about 40% in the UK. Explanations for these findings emerge from Aparicio et al. (2024)[3], who compared price dispersion for groceries in online and offline U.S. markets using scanner data. Contrary to expectations and earlier studies (e.g., Brynjolfsson & Smith, 2000)[9], they observed higher dispersion online, which they attribute to algorithmic pricing strategies enabling discrimination across time, location, and retailers. Zhuang et al. (2018)[44] further explain elevated online price dispersion by highlighting increased consumer risk and retailer type. In particular, they find that as the number of pure online retailers grows, so does price dispersion. Importantly, studies on consumer search behavior in online markets suggest that search frictions are still present even though search costs should be low compared to offline markets. De Los Santos (2018)[13] exploited web browsing data from consumers in the online book market. They find that 75% of consumers visit at most one online retailer before conducting a transaction. In the market of MP3-Players consumers search on average for two-three price quotas before engaging in a transaction (Santos et al., 2017)[34]. This indicates that limited search-effort is present contrary to the prediction that digital transparency removes search frictions.

2.2. Search Cost Estimation

In the empirical literature on consumer search, several estimation approaches have been developed to quantify search costs using observed market data. These methods aim to explain the presence of price dispersion with the predictions of search models. Nonetheless, they differ in terms of assumptions, data requirements, and modeling flexibility.

A pioneering contribution comes from Hong and Shum (2006)[19], who propose a method for estimating search cost distributions using only price data. Focusing on homogeneous goods and building on the Burdett and Judd (1983)[10] framework, they exploit the equilibrium implications of price distributions under both sequential- and non-sequential search. Their methodology identifies "critical" search cost cutoffs by linking each observed price to the share of consumers who search a given number of firms. This approach provides a foundation for estimating search costs without observing consumer behavior or quantities. However, it relies on the assumption that all firms are symmetric and that consumers either search sequentially or draw a fixed number of price quotes non-sequentially. Building on this, Moraga-González and Wildenbeest (2008)[28] propose a maximum likelihood estimation (MLE) approach for non-sequential search that improves numerical efficiency. Like Hong and Shum, they use price data only but integrate firm heterogeneity more flexibly by modeling search equilibria explicitly and allowing the data to inform the search cost distribution. This approach is particularly tractable for symmetric oligopoly markets. To improve flexibility, Moraga-González et al. (2013)[27] propose a semi-nonparametric (SNP) method that pools data from multiple articles with a common search technology. This approach enables robust identification of the full search cost distribution and uses flexible polynomial approximations that outperform earlier spline-based methods, which are usually prone to over- and underfitting. See chapter 5 for further explanations. Sanches et al. (2018)[33] introduce a minimum distance estimator that uses price distribution restrictions to identify search cost quantiles via non-linear least squares. It is simple, computationally efficient, and extendable to a sieve framework, making it practical when full structural models are infeasible.

As opposed to the analysis of homogenous goods, a parallel development in the literature incorporates differentiated product markets and consumer heterogeneity more explicitly. Hortacsu and Syverson (2004)[20] apply a structural model to mutual funds, where search and vertical differentiation coexist. They require both price and quantity data and assume consumers make discrete choices over product attributes and fees. Later work by Wildenbeest (2011)[42] and Lin and Wildenbeest (2020)[26] further integrates vertical and horizontal differentiation in estimating search costs, showing that ignoring quality heterogeneity can lead to significant overestimation of search frictions.

Finally, recent advances exploit consumer behavior data, such as browsing histories. De los Santos (2018)[13], Koulayev (2014)[25], and Kim et al. (2010)[24] model search using direct observations of which firm consumers visit and in what order. These models accommodate directed and sequential search, endogenous search set formation, and preference heterogeneity. While powerful, they require detailed behavioral data and often rely on strong assumptions about consumer rationality and expectations.

Overall, empirical methods for estimating search costs have evolved considerably. From early price-based equilibrium models to sophisticated structural and behavioral approaches. Hong and Shum (2006)[19] provided the crucial steppingstone by demonstrating how search costs could be inferred from price distributions alone. Subsequent work has expanded these ideas into more flexible, computationally efficient, and behaviorally rich frameworks suitable for both homogeneous and differentiated product markets.

2.3. Relative Thinking

Whereas search cost estimation methods attempt to estimate search costs for a given set of market data, the literature on relative thinking typically seeks to infer search cost differences across price levels using experimental settings. One of the most prominent examples is Thaler's (1980)[39] TV-calculator experiment: consumers exert greater effort to save \$5 on a \$25 calculator than on a \$500 TV. Another foundational study is by Tversky and Kahneman (1981)[40]. In their experiment, participants were asked whether they would drive 20 minutes to another store to save \$5 on a purchase. The context was varied: under one condition, the item's price was low (e.g., \$15); in another, it was high (e.g., \$125). The study found that a majority were willing to make the trip to save \$5 on the cheaper item, but far fewer would do so for the more expensive one.

Some authors, such as Azar (2007)[4], attempted to formalize this idea by developing a model in which consumers consistently consider relative price differences alongside absolute ones, even in contexts where only absolute prices should influence rational decision-making. Later, Azar (2011)[5] showed that some individuals are entirely guided by relative price differences ("full relative thinking"), while others also account for absolute price differences ("partial relative thinking").

A more recent controlled experiment was conducted by Karle et al. (2025)[23]. In this study, participants were fully informed about the true price distribution of 100 items. To access one additional price quote, they had to enter a 16-digit code requiring 60–85 seconds of effort, serving as a proxy for search costs. To measure relative thinking, the researchers kept the search effort constant but varied the price range. They found that search costs increased with the price scale; this effect was more pronounced among AMT workers than among students.

Similarly, Somerville (2022)[35] demonstrated in an experimental setting that consumers become less sensitive to prices as the overall scale increases, making them more likely to engage in a purchase. Bushong et al. (2021)[11] also showed that people judge price differences not only by their absolute size but also by how large they appear relative to the entire choice set. A price difference feels substantial when all other prices are within a similar range, but less so when the price range is broad.

3. Data

This chapter introduces the dataset used to empirically investigate the Price Dispersion Puzzle in the German digital electronics market. It outlines how the data was collected, what it comprises, and how it was processed. Given the central role of price comparisons and consumer search behavior in this study, careful attention is paid to the structure and composition of the sample, the characteristics of retailers, and the steps taken to prepare the data for the subsequent econometric analysis.

3.1. Source

The data was collected using a large-scale web scraping application developed in Python with the Selenium library. The target source was Idealo.de, one of the most popular price comparison platforms in Germany with over 30 million monthly visitors. Idealo was selected due to its extensive coverage, featuring over 50,000 retailers and 400 million price offers (Idealo, 2025)[21], making it a representative source across various industries. Scraping Idealo.de instead of Idealo.at or Idealo.fr ensures that only products available in the German market are included. This guarantees that the analysis is conducted at the national level and is not biased by retailers from Austria or France, for example, who may systematically offer different prices due to variations in VAT or transportation fees.

3.2. Sample

The dataset is limited to electronic articles from the German digital market. It comprises 170,633 unique products, each associated with multiple price observations, resulting in a total of 5,664,221 price quotes. To reduce noise and potential bias, prices above the 99.9th percentile have been excluded. This upper range includes outliers such as professional sound systems used for music and festival events, which can cost over €200,000 and are typically purchased by firms rather than individual consumers. These products also exhibit low standard deviation due to their infrequent availability in online retail shops.

Category	Subcategories	% of whole Sample
Computer	Drives, Graphics Cards, CPUs/Processors, RAM, Motherboards, PC Cases, ...	29.3%
Printers & Scanners	Multifunction Printers, Printer Cartridges, Toner, Laser Printers, Photo Printers	8.5%
Photography	Lenses, Camera Accessories, Digital Cameras, Action Cams, Instant Cameras, Camcorders	8.1%
Gaming Accessories	Gamepads, Steering Wheels, Joysticks, Gaming Consoles, Electronic Toys, Picture Frames, ...	7.4%
Household Electronics	Refrigerators, Washing Machines, Dishwashers, Clothes Dryers, Ovens, Cooktops, ...	18.2%
DIY	Electric Heaters, Dehumidifiers, Air Purifiers, Fans, Humidifiers, Cordless Screwdrivers, ...	8.3%
HiFi & Audio	Headphones, Musical Instruments, Children's Musical Instruments, Studio & Event Equipment, Radios, ...	7.8%
Telecommunication	Mobile Phones & Smartphones, Smartwatches, Mobile Phone Accessories, Telephones, Smartwatch Accessories, ...	8.6%
TV & Satellite	Televisions, Network Players, TV Accessories, Satellite Reception, Blu-ray Players, ...	3.9%

Figure 3.1: Category Structure

Excluding them helps avoid distortions in the analysis of price dispersion and consumer search costs. Following the literature on search models, which suggests that consumer search behavior varies by product type, the data is categorized into nine main groups and 274 subcategories (see Figure 2.1). Although the TV & Satellite category represents a relatively small share of the total sample, its absolute size remains sufficient. Specifically, it accounts for approximately 4% of the dataset, corresponding to around 6,800 articles and 225,000 price observations Table 2.1 provides an overview of key descriptive statistics,

Descriptive Statistics

	Mean	Median	SD	Min	Max	NAs
Prices without delivery costs	365.36	117.71	699.82	0.01	8,999	0
Prices with delivery costs	369.46	119.80	700.51	0.80	8,999	0
Shoprating	3.96	4.40	1.05	1	5	149,924
Number of ratings	18,288.08	638	107,796.70	1	8,029,113	149,924
Delivery days	4.84	5	1.48	1	12	2,281,321
Refund days	23.05	14	13.19	1	365	3,197,601

Notes: n = 5 664 221. The top 0.1% of articles, based on prices including delivery costs, have been excluded from the sample to mitigate the influence of outliers on the results.

Table 3.1: Descriptive Statistics

which remain largely consistent whether prices include delivery costs or not. Idealo also allows consumers to rate retailers within specific subcategories. In addition to total delivery time, the dataset includes information on refund days, which indicate the period during which a consumer can return a product to the retailer free of charge.

However, data on delivery and refund days is only available for approximately half of the sample. To examine how price levels differ across categories, Figure 3.2 provides

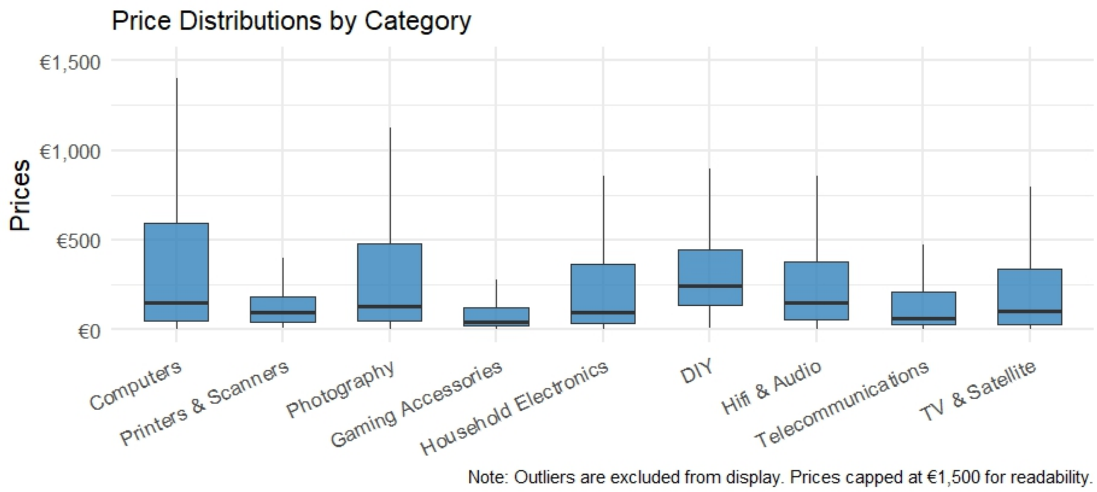


Figure 3.2: Price distributions by category

additional insights.

3.3. Shops

The dataset includes price information from 14,233 different retailers. However, most of these shops list prices for no more than 50 distinct articles (see Table 3.2).

Amount of articles offered per shop

Category	Count
Total Shops	14,233
Shops with 0 - 50 articles	12,238
Shops with 51 - 100 articles	663
Shops with 101 - 499 articles	868
Shops with 500 - 999 articles	177
Shops with 1000+ articles	287

Table 3.2: Available prices per shop

To ensure data quality, the final dataset used in the main analysis includes only retailers that exceed a minimum threshold of total price quotes. This restriction is necessary because Idealo cannot fully guarantee that its platform is free from fake shops and from their associated manipulated prices. Importantly, fake shops have a strong incentive to post numerous low-priced offers to increase their visibility and attract more customers to their own websites. By flooding the platform with artificially low prices across many products, they can appear more competitive than legitimate retailers. However, it is expected that Idealo is more likely to detect and remove such fraudulent shops if their number of listed prices becomes excessively large. Therefore, setting a minimum activity threshold helps strike a balance between including a broad range of retailers and filtering out potentially unreliable data. This restriction does not significantly reduce the overall dataset size, as shops with more than 500 price quotes still account for 83.8% of all available observations. Additionally, focusing on more active retailers improves the estimation of e-tailer-specific effects in the hedonic pricing model. For some price quotes, data on available payment and delivery methods is also included. These are summarized in Table 3.3. Payment and delivery services can serve as indicators of available buyer protection. Including them helps account for differences in payment security in the search cost analysis.

Payment Methods and Delivery Services

	Payment methods	offered by in %	Delivery services	offered by in %
1	Paypal	88.56	DHL	66.39
2	Visa	81.65	DPD	33
3	Mastercard	81.48	Spedition	28.72
4	Payment in Advance	74.73	UPS	17.20
5	Order on Account	59.74	Hermes	11.85
6	Immediate Transfer	44.80	GLS	11.37
7	Debit Note	35.10	Deutsche Post	5.90
8	American	30.97	Click and Collect	5.19
9	Amzpay	20.75	trans o flex | TOF	2.28
10	Giropay	11.41	GoGreen	2.03
11	Maestro	7.84	German Express Logistics	0.91
12	Diners	5.64	Fedex	0.37
13	Cash on Delivery	5.51	TNT	0.33
14	eps	2.94	ThinkGreen	0.03
15	Kreditkarte	1.77		
16	Google Checkout	0.84		
17	Jcb	0.69		
18	Carte	0.19		
19	Click and Buy	0.02		
20	Paysafecard	0.02		
21	PostPay	0.02		

Notes: n = 21 different payment methods; n = 14 different delivery services.
Certain payment and delivery services were excluded from the analysis due to their low frequency.

Table 3.3: Payment methods and delivery services

3.4. Sample Construction

During the search cost estimation procedure, different subsamples are employed to mitigate potential sample selection biases that may arise, for example, when only price observations with complete covariate information are used. In addition, conducting a robust store fixed-effects regression requires a substantial number of observations per shop. To address these issues, a series of restrictions is introduced. Alongside the full sample containing all available prices, the following three restricted subsamples are considered:

Restriction 1: Only price points for which all covariates are available.

Restriction 2: R1 + only shops with at least 500 price points.

Restriction 3: R1 + R2 + only articles for which at least 30–40 prices are available.

Shops that no longer meet the requirement of Restriction 2 after applying Restriction 3 are removed from the sample corresponding to Restriction 3.

4. Price Dispersion Puzzle

This subchapter analyzes the presence of the Price Dispersion Puzzle in the available dataset. Hence, several measures of price dispersion are compared. In addition to testing different sample specifications, the estimates are also examined for heterogeneity across product categories.

4.1. Price Dispersion Analysis

Before analyzing how search costs behave in the data, the first goal is to assess whether the Price Dispersion Puzzle is present in the scraped dataset. To this end, several measures from the price dispersion literature are employed to examine whether an increase in mean price leads to greater price variation for a given article. One of the standard measures is the standard deviation (see, e.g., Kaplan & Menzio, 2015[22]; Zhuang et al., 2018[44]). The use of log-transformed values of both the standard deviation and the mean price is also common in price dispersion analyses (e.g., Pan et al., 2002[29]). Additionally, Zhuang et al. (2018)[44] propose the price range as an alternative measure of dispersion (see also Brynjolfsson & Smith, 2000[9]). Furthermore, relative measures such as the coefficient of variation are considered, particularly in the context of price differences for homogeneous goods (Gorodnichenko et al., 2018[18]; Kaplan & Menzio, 2015[22]). The following regressions are used to illustrate how price dispersion responds to increases in the (log) mean price of an article:

$$SD_i = \beta_0 + \beta_1 \cdot \text{Mean Price}_i + \varepsilon_i \tag{4.1}$$

$$\log(SD_i) = \beta_0 + \beta_1 \cdot \log(\text{Mean Price}_i) + \varepsilon_i \tag{4.2}$$

$$\log(\text{range}_i) = \beta_0 + \beta_1 \cdot \text{Mean Price}_i + \varepsilon_i \tag{4.3}$$

$$CV_i = \beta_0 + \beta_1 \cdot \text{Mean Price}_i + \varepsilon_i \tag{4.4}$$

These regressions are applied not only to the full dataset for illustration, but also to various subsets defined by Restriction (1) - (3), which are relevant for the subsequent search cost estimation. This approach ensures that any positive relationship between absolute price dispersion and mean price is robust and not driven by the specific sample construction. Table 4.1 provides an overview of how different dispersion measures respond when the (log) mean price increases. Columns (1), (3), (5), and (7) are based on the full sample, whereas the remaining columns use the subset defined by Restriction (1). The imposed restriction leads to a noticeable difference in the regression of standard deviation on mean price: in the full sample, a 1€ increase in mean price is associated with an average increase of 0.097€ in the standard deviation. Under Restriction (1), the same price increase results in only a 0.058€ increase, on average. However, for the remaining dispersion measures, the effects are similar in magnitude. A 1% increase in mean price leads, on average, to a 0.79–0.82% increase in the log-standard deviation and a 0.78–0.82% increase in the total log-price range. Interestingly, the coefficient of variation (CV) remains fairly stable: there is only a very small negative effect, which is statistically significant due to the large sample size. Nonetheless, when comparing model predictions for a 10 € product and a 500€ product, the respective CVs would be 17.14% $(17.17 + (-0.003) \cdot 10 = 17.14)$ and 15.67% $(17.17 + (-0.003) \cdot 500 = 15.67)$ according to the model based on the full sample.

Price Dispersion Puzzle - Restr. (2) + Restr. (3)

	Dependent variable:							
	SD		ln SD		ln Range		CV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Price	0.054***	0.072***					-0.002***	-0.004***
	(0.0003)	(0.001)					(0.0001)	(0.0004)
ln Mean Price			0.787***	0.814***	0.776***	0.801***		
			(0.003)	(0.008)	(0.004)	(0.009)		
Constant	5.997***	9.932***	-1.523***	-1.310***	-0.621***	-0.130***	10.154***	15.985***
	(0.229)	(0.826)	(0.016)	(0.040)	(0.018)	(0.044)	(0.041)	(0.236)
Observations	101,111	4,345	77,083	4,329	77,083	4,329	101,111	4,345
R^2	0.227	0.399	0.429	0.692	0.381	0.651	0.010	0.026
Adjusted R^2	0.227	0.398	0.429	0.692	0.381	0.651	0.010	0.026

Note: *p<0.1; **p<0.05; ***p<0.01
This table analyzes four distinct measures of price dispersion. Each regression is estimated using the sample under Restr. (2) and (3)
The models in the left columns are based on sample Restr. (2), while the models in the right columns restrict the sample to sample Restr. (3).

Table 4.1: Price dispersion analysis for full sample and restr. (1)

If larger absolute price dispersion is partly driven by the effects of relative thinking, then an approximately constant CV would support the idea that consumers irrationally consider only relative savings rather than absolute ones. Since the search cost estimation is based on the subset defined by Restriction (3), Table 4.2 presents the results for the price dispersion measures under Restriction (2) and Restriction (3).

Price Dispersion Puzzle - Restr. (2) + Restr. (3)

	Dependent variable:							
	SD		ln SD		ln Range		CV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Price	0.054***	0.072***					-0.002***	-0.004***
	(0.0003)	(0.001)					(0.0001)	(0.0004)
ln Mean Price			0.787***	0.814***	0.776***	0.801***		
			(0.003)	(0.008)	(0.004)	(0.009)		
Constant	5.997***	9.932***	-1.523***	-1.310***	-0.621***	-0.130***	10.154***	15.985***
	(0.229)	(0.826)	(0.016)	(0.040)	(0.018)	(0.044)	(0.041)	(0.236)
Observations	101,111	4,345	77,083	4,329	77,083	4,329	101,111	4,345
R^2	0.227	0.399	0.429	0.692	0.381	0.651	0.010	0.026
Adjusted R^2	0.227	0.398	0.429	0.692	0.381	0.651	0.010	0.026

Note: *p<0.1; **p<0.05; ***p<0.01
This table analyzes four distinct measures of price dispersion. Each regression is estimated using the sample under Restr. (2) and (3)
The models in the left columns are based on sample Restr. (2), while the models in the right columns restrict the sample to sample Restr. (3).

Table 4.2: Price dispersion analysis for restr. (2) and restr. (3)

Note that only a small fraction of the total articles available remain after imposing the restrictions under Restriction (3). Nevertheless, 4.300 articles still constitute of approximately 150 000 price observations. Comparing the measures under Restriction (3) in columns (2), (4), (6) and (8) is very similar to its full sample counterpart in Table 4.1.

4.2. Category Heterogeneity

Since the increase in absolute price dispersion may vary across categories, the relationship between the price dispersion measures and the (log) mean price is analyzed while accounting for heterogeneous effects. Therefore, interaction terms between the (log) mean price and product categories are included. The following regression will be estimated:

4.2.1. Model

$$PD\ Measure_i = \sum_{i=1}^{9} \beta_i^{(\text{cat.})} + \sum_{i=10}^{19} \beta_i^{(\text{cat.})} \cdot MeanPrice_i \cdot category_i + \varepsilon_i \tag{4.5}$$

Note that for the standard deviation and the coefficient of variation as dependent variables, the linear version of $Mean\ Price_i$ is used. When the logarithm of the standard deviation or the range is considered, $Mean\ Price_i$ is also log-transformed.

Cat. is an index for the nine main categories (see Figure 3.1) available, i.e.,

$$\text{cat.} \in \{\text{Computer}, \ldots, \text{TV \& Satellite}\},$$

For the sake of clarity, no reference category is omitted in this regression. Instead, there is an individual intercept for each category and a corresponding estimate for the interaction term.

4.2.2. Results

Price Dispersion Puzzle under Category Heterogeneity -- (linear) Mean Price

	Dependent variable:					
		SD			CV	
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Price* *Computer*	0.126***	0.066***	0.072***	-0.002***	-0.002***	-0.004***
	(0.001)	(0.001)	(0.002)	(0.0001)	(0.0001)	(0.001)
Mean Price* *Printers*	0.168***	0.054***	0.074***	-0.004***	-0.004***	-0.005***
	(0.003)	(0.003)	(0.006)	(0.0005)	(0.001)	(0.002)
Mean Price* *Photography*	0.056***	0.026***	0.043***	-0.003***	-0.002***	-0.004***
	(0.001)	(0.001)	(0.003)	(0.0001)	(0.0002)	(0.001)
Mean Price* *Gaming Accessories*	0.088***	0.044***	0.053***	-0.003***	-0.002***	-0.006***
	(0.001)	(0.001)	(0.007)	(0.0002)	(0.0002)	(0.002)
Mean Price* *Household Electronics*	0.073***	0.054***	0.128***	-0.003***	-0.002***	-0.003**
	(0.001)	(0.001)	(0.005)	(0.0001)	(0.0001)	(0.001)
Mean Price* *DIY*	0.073***	0.043***	0.206***	-0.002***	-0.003***	0.008
	(0.001)	(0.002)	(0.028)	(0.0002)	(0.0004)	(0.008)
Mean Price* *Hifi*	0.102***	0.063***	0.152***	-0.001***	-0.001***	0.002
	(0.001)	(0.001)	(0.009)	(0.0001)	(0.0001)	(0.003)
Mean Price* *Telecommunication*	0.148***	0.071***	0.074***	-0.005***	-0.003***	-0.009***
	(0.003)	(0.002)	(0.007)	(0.0004)	(0.0004)	(0.002)
Mean Price* *TV*	0.129***	0.087***	0.203***	-0.002***	-0.001***	0.0001
	(0.001)	(0.001)	(0.022)	(0.0002)	(0.0002)	(0.006)
Observations	170,650	108,235	4,348	170,650	108,235	4,348
R^2	0.454	0.314	0.562	0.558	0.422	0.552
Adjusted R^2	0.454	0.314	0.560	0.558	0.422	0.550
Note:	*p<0.1; **p<0.05; ***p<0.01 SD = Standard deviation, CV = Coefficient of variation. Columns (1) + (4) based on full sample; Columns (2) + (5) based on Restr. (1); Columns (3) + (6) based on Restr. (3). Category-specific intercepts are omitted to conserve space.					

Table 4.3: Category heterogeneity for SD and CV

Columns (1)–(3) in Table 4.3 use the standard deviation (SD), and Columns (4)–(6) use the coefficient of variation (CV) as the dependent variable. The explanatory variable is the mean price in linear form. For each price dispersion measure, the regressions are estimated using the full sample as well as the subsets defined by Restriction (1) and (3). Heterogeneous effects are quite substantial: in the full sample, the estimated effects range from 0.056 to 0.168. Under Restriction (3), which is the sample used for the search cost estimation, the variation across categories is even more pronounced, with coefficients ranging from 0.043 to 0.206. Restriction (1) yields somewhat lower estimates and less heterogeneity, between 0.026 and 0.087. In contrast to the average-effect analysis in the previous section, the inclusion of category-level heterogeneity reveals that results are sensitive to the imposed sample restrictions. Nevertheless, the effects for the CV remain largely stable, and under Restriction (3), some category-specific coefficients even become statistically insignificant.

Next, in Table 4.4, the log-standard deviation and the log-range are used as dependent variables, with the log-mean price serving as the explanatory variable. The corresponding results are reported in Table 4.4. Columns (1)–(3) and (4)–(6) present the results for the respective dependent variables under different sample restrictions. Within each group, the left column is based on the full sample, the middle column on Restriction (1), and the right column on Restriction (3).

Price Dispersion Puzzle under Category Heterogeneity - (log) Mean Price

	Dependent variable:					
		ln SD			ln Range	
	(1)	(2)	(3)	(4)	(5)	(6)
ln Mean Price* *Computer*	0.832***	0.735***	0.752***	0.841***	0.743***	0.747***
	(0.003)	(0.006)	(0.013)	(0.004)	(0.006)	(0.014)
ln Mean Price* *Printers&Scaners*	0.871***	0.811***	0.851***	0.813***	0.751***	0.750***
	(0.013)	(0.018)	(0.029)	(0.014)	(0.020)	(0.031)
ln Mean Price* *Photography*	0.739***	0.752***	0.810***	0.747***	0.775***	0.785***
	(0.007)	(0.013)	(0.032)	(0.008)	(0.014)	(0.035)
ln Mean Price* *Gaming Accessories*	0.813***	0.767***	0.756***	0.798***	0.776***	0.740***
	(0.007)	(0.013)	(0.036)	(0.008)	(0.014)	(0.038)
ln Mean Price* *Household Electronics*	0.855***	0.804***	0.882***	0.847***	0.780***	0.857***
	(0.003)	(0.006)	(0.018)	(0.004)	(0.007)	(0.020)
ln Mean Price* *DIY*	0.897***	0.952***	1.081***	0.897***	0.915***	1.125***
	(0.007)	(0.018)	(0.064)	(0.008)	(0.020)	(0.069)
ln Mean Price* *Hifi&Audio*	0.815***	0.872***	0.968***	0.783***	0.847***	0.948***
	(0.005)	(0.009)	(0.037)	(0.006)	(0.010)	(0.040)
ln Mean Price* *Telecommunication*	0.860***	0.820***	0.846***	0.883***	0.853***	0.899***
	(0.007)	(0.012)	(0.029)	(0.008)	(0.013)	(0.031)
ln Mean Price* *TV&Sat*	0.828***	0.782***	0.908***	0.800***	0.768***	0.928***
	(0.007)	(0.014)	(0.063)	(0.008)	(0.015)	(0.067)
Observations	166,327	84,041	4,332	166,327	84,041	4,332
R^2	0.894	0.804	0.936	0.923	0.857	0.960
Adjusted R^2	0.894	0.804	0.936	0.923	0.857	0.960

Note: *p<0.1; **p<0.05; ***p<0.01
SD = Standard deviation, CV = Coefficient of variation. Columns (1) + (4) based on full sample; Columns (2) + (5) based on Restr. (1); Columns (3) + (6) based on Restr. (3). Category-specific intercepts are omitted to conserve space.

Table 4.4: Category heterogeneity for ln SD and ln Range

The log-transformation of the variables results in lower overall heterogeneity across categories. Under the full sample, as shown in Column (1), a 1% increase in mean price leads to a 0.739–0.871% increase in the standard deviation. Restriction (1) and Restriction (3) show slightly more variation in their estimates, ranging from 0.735–0.952% and 0.752–1.081%, respectively (see Columns (2) and (3), Table 4.4). Similarly, the log-range as a measure of price dispersion yields effects of comparable magnitude in the full sample (Column 4) and under Restriction (1) (Column 5). However, under Restriction (3) (Column 6), the estimated coefficients vary more widely, ranging from 0.74% to 1.125%. To illustrate the magnitude of this difference: consider a product with a mean price of €100 and a standard deviation of €20. If the mean price increases by 20%, a category with an elasticity of 1.125 would see its dispersion rise to €24.86 (a 24.3% increase), whereas a category with an elasticity of 0.74 would only see it rise to €22.84 (a 14.2% increase).

5. Search Cost Analysis

The previous chapter confirmed that price dispersion increases with mean prices, challenging the predictions of standard search models with fixed search costs. In this chapter, a game-theoretic model is introduced that allows search costs to be inferred from observed price distributions. The estimation procedure is first illustrated using a single article group, and then applied across multiple groups to test whether estimated search costs systematically increase with average prices - proving the presence of Relative Thinking. Additionally, a hedonic pricing regression is used to control for product characteristics, and the search cost estimation is repeated on its residuals to ensure robustness.

5.1. Setup

To estimate consumer search costs in a market for homogeneous goods, this thesis applies the semi-nonparametric (SNP) method developed by Moraga-González, Sándor and Wildenbeest (2013)[27]. Unlike earlier approaches (e.g., Hong and Shum, 2006[19]), which estimate search cost distributions on a market-by-market basis, this SNP-approach allows for the pooling of price data from similar products within the same market, as long as they share the same underlying search technology. While consumers in different markets may vary in their valuations, firm costs, and number of competitors, their search behavior is assumed to follow a common cost distribution. This pooling facilitates a more granular estimation of search cost cut-off points, leading to greater precision in recovering the full cumulative distribution function (CDF) of consumer search costs.

5.1.1. Theoretical Model

The model builds on the non-sequential search framework of Burdett and Judd (1983)[10] and its extension by Hong and Shum (2006)[19], which introduces heterogeneous search costs across consumers. There are K firms, each producing a homogeneous product at unit cost r. A continuum of consumers (normalized to unit mass) each wishes to purchase at most one unit. Consumers face search costs when observing prices but are allowed to observe the first price for free, consistent with most of the literature.

Each consumer has a privately known search cost c, drawn from a continuous, atomless distribution $G(c)$ with density $g(c)$. A consumer who observes k prices incurs a total cost of $(k-1)c$. The maximum price any consumer is willing to pay is v. After observing k prices, a consumer purchases from the firm offering the lowest observed price. Firms and consumers engage in a simultaneous move game. Each firm sets a pricing strategy F_i, which is a distribution over prices. In equilibrium, firms choose prices to maximize profits given the pricing strategies of competitors and consumers' search behavior. Symmetric mixed-strategy equilibria are of interest here. The most important equilibrium conditions are as follows:

1. Firms best respond to each other's strategies.
2. Firms are indifferent across all prices in the support of F, implying constant expected profit across this support where μ_k is the share of consumers searching for k prices:

$$(p-r)\sum_{k=1}^{K} k\mu_k(1-F(p))^{k-1} = \mu_1(v-r)$$

3. Consumers optimize search: Each consumer chooses a number $k \in \{1,\dots,K\}$ that maximizes expected utility where $E[p_{(1:k)}]$ is the expected minimum price if the consumer searches for k different prices:

$$k(c) = \arg\max_k \left[v - E[p_{(1:k)}] - (k-1)c\right]$$

A mixed-strategy equilibrium arises when some consumers search only one firm (thus firms can charge high prices), while others search multiple firms (imposing competitive pressure). This mix avoids both monopoly pricing and Bertrand undercutting.

5.1.2. Estimating Search Costs

To estimate the distribution $G(c)$, a series of intermediary steps is required:

- **Cutoff values** c_k are defined such that consumers are indifferent between searching k vs. $k+1$ firms and $E[p_{(1:k)}]$ will be computed based on the empirical CDF:

$$c_k = E[p_{(1:k)}] - E[p_{(1:k+1)}]$$

- **The share of consumers** μ_k searching k firms is given by:

$$\mu_1 = 1 - G(c_1)$$
$$\mu_k = G(c_{k-1}) - G(c_k), \quad \text{for } k = 2, \ldots, K$$

Instead of estimating $G(c)$ by interpolating across cutoff values, this approach employs semi-nonparametric (SNP) maximum likelihood estimation following Gallant and Nychka (1987)[17]. This method allows for a flexible approximation of the density $g(c)$ while pooling data across similar articles.

- **Search cost density** is approximated using a Hermite polynomial expansion

$$g(c; \gamma, \sigma, \boldsymbol{\theta}) = \frac{\sum_{i=0}^{N} \theta_i u_i(c)}{\left(\sum_{i=0}^{N} \theta_i^2\right)^{1/2}}$$

with the transformation $c = \exp(\gamma + \sigma x)$. This transformation, where x is a standard random variable, follows the SNP estimator framework of Fenton and Gallant (1996)[16]. It ensures positivity of the search cost distribution and aligns with the parametric form used in their original density specification. The θ_i are weights applied to the basis functions u_i, and the parameter vector $(\gamma, \sigma, \theta_1, \ldots, \theta_N)$ is estimated via maximum likelihood. So the task is to find the set of parameters which is able to describe the price density given the presence of search costs the best.

The price density given its search cost for an article m is derived by the equilibrium conditions mentioned in 5.1.1 and is defined as:

$$f^m(p_i \mid g) = \frac{\sum_{k=1}^{K^m} k\mu_k^m (1 - F^m(p_i \mid g))^{k-1}}{(p_i - r^m) \sum_{k=1}^{K^m} k(k-1)\mu_k^m (1 - F^m(p_i \mid g))^{k-2}}$$

Since this approach estimates the relevant parameters using information not from a single article but from a group of articles that share the same search cost technology, the log-likelihood is constructed using all pooled data:

$$\mathcal{L}(g \mid p_1, \ldots, p_M) = \frac{1}{M} \sum_{m=1}^{M} \left(\frac{1}{K^m} \sum_{i=1}^{K^m} \log f^m(p_i \mid g) \right)$$

To determine the optimal number of SNP terms N, the Integrated Squared Error (ISE) is minimized via cross-validation, following the approach of Coppejans and Gallant (2002)[12]. This method balances the flexibility of the approximation against the risk of overfitting. Additionally, the Conditional Kolmogorov (CK) test developed by Andrews (1997)[2] is used to assess how well the price density is captured by the current model parameters. After estimating the model for various values of N, the search cost estimates corresponding to the best CK/ISE combination are selected.

5.1.3. Example

Before computing the search cost density and its respective moments, three key observations from Moraga-González et al. (2013)[27] application example must be emphasized: (1) They use ten different products in their article pool, (2) market size heteroscedasticity is limited, with approximately 30 to 40 prices per article, and (3) the articles lie within a similar price regime.

These criteria have also been applied to the dataset used in this thesis. Since the aim is to test whether search costs increase with rising mean prices, several groups were constructed that satisfy the three conditions mentioned above. Unlike the original authors, this thesis does not have access to product-specific characteristics to enforce similarity in pricing (condition 3). However, since products with similar features typically fall within a comparable price range, a quantile-based price ranking is used as a proxy. The number of quantiles is determined by the number of articles available within each product category. For example, if a category contains 50 articles, it is divided into five quantiles to ensure each group contains ten articles. Applying this procedure to the dataset results in 306 distinct groups.

Unlike Moraga-González et al. (2013)[27], who illustrate their method using memory chips, this thesis presents a custom application example based on a group of ten CPU processors priced above €200 which can be found in Table 4.1. For each processor, between 30 and 40 different price quotations are available. The analysis is based on prices including transportation fees, as these represent the final costs relevant to consumers when making purchasing decisions. The presence of significant price dispersion implies that the savings between a fully informed consumer and one who searches only once are non-negligible.

Next, the relevant parameters for the search cost distribution are estimated. As in Moraga-González et al. (2013)[27], it is not possible to determine a priori how many SNP parameters (denoted N) should be included in the search cost density function. Therefore, the density is estimated for various values of N. For each value, the log-likelihood (LL) is computed, alongside the two model-fit measures: the Integrated Squared Error (ISE) and the Conditional Kolmogorov (CK) statistic. The CK statistic, following Andrews

Descriptive Statistics per Article (CPU Processors)

	Articlename	No. of Stores	Mean Price	SD Price	Min Price	Max Price
1	Intel Core i5-14600KF Tray	36	234.18	18.71	213.69	294.8
2	AMD Ryzen 7 5700X3D Boxed WOF	40	240.06	21.78	204.91	305.36
3	AMD Ryzen 7 7700 Tray	40	259.74	11.7	243.19	289.22
4	Intel Core i5-13600KF	36	262.28	44.95	216.29	400
5	Intel Core i7-12700KF Boxed (BX8071512700KF)	38	267.33	17.94	248.77	331
6	AMD Ryzen 7 5800XT Boxed	30	272.57	28.13	240.94	369.8
7	Intel Core i7-12700F Tray (CM8071504555020)	30	275.69	26.13	253.78	335
8	Intel Core i5-13600K Boxed	33	288.44	21.59	258.81	345
9	Intel Core i5-13600K Tray	32	294.76	27.21	251.85	349
10	Intel Core i5-13600KF Boxed	39	298.71	30.87	259.48	359.78

Table 5.1: Descriptives for CPU processors

(1997)[2], tests the null hypothesis that the model is correctly specified, conditional on the estimated SNP parameters and observed market covariates. A low CK statistic implies better model fit, as it increases the likelihood of maintaining the null hypothesis. The results are presented in Table 4.2.

Fit SNP estimates for different values of N

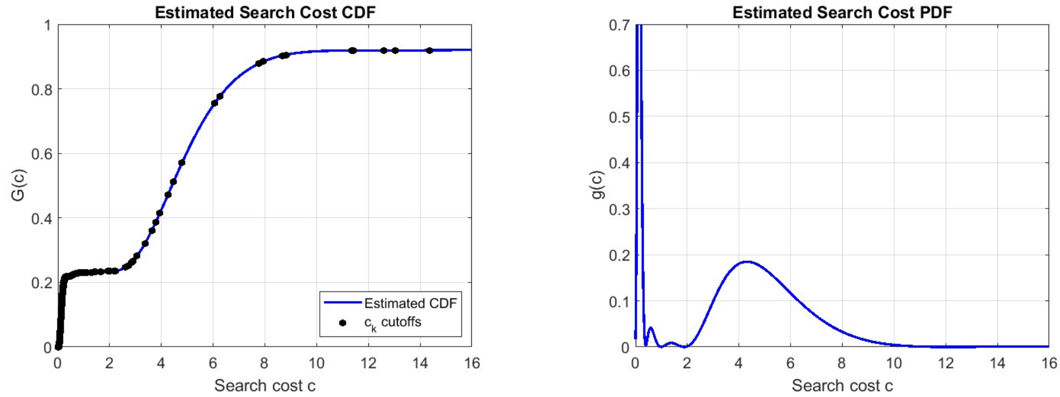
N	LL	ISE	Conditional.Kolmogorov
1	45.3740	-0.0228	3.3020
2	43.0700	-0.0316	0.8750
3	43.0540	-0.0317	0.8770
4	43.0510	-0.0318	0.8310
5	43.0540	-0.0317	0.8370
10	42.2610	-0.0364	0.6100
15	42.1930	-0.0377	0.6140
20	42.1410	-0.0374	0.6550
23	42.1640	-0.0382	0.5800
25	42.1680	-0.0376	0.5820

N = Number of SNP-Parameters, LL = log-likelihood value, ISE = integrated squared error

Table 5.2: SNP-estimates for CPU processors example

Due to increasing computational burden—especially for larger N—the analysis is limited to values of N up to 25. The SNP specification yielding the lowest CK statistic (and typically also the lowest ISE) is chosen as the best-fitting model. For this CPU processor group, the CK statistic decreases steadily up to around N=20, after which it levels off. The best fit is observed at N=23, and thus, the corresponding search cost PDF and CDF, along with the distributional moment estimates, are computed based on this specification.

Figure 4.1 shows the estimated PDF and CDF of the search cost distribution.



(a) Cumulative distribution function   **(b)** Probability density function

Figure 5.1: G(c) and g(c) of search costs distribution

Following the estimation of SNP parameters, the distribution's moment conditions are derived. Consistent with Moraga-González et al. (2013)[27], the moments are calculated up to the 95th percentile to mitigate distortions due to extreme tail behavior often present in flexible SNP estimators. This threshold choice is validated by testing the method on the original authors' data. For the CPU processor group analyzed here, the mean search cost is approximately €5.85, with a standard deviation of €5.33. The median consumer exhibits a search cost of around €4.24. Importantly, although CK statistics may be similar across nearby SNP specifications, the resulting moment estimates can still differ considerably. While median estimates remain relatively stable, the mean search cost varies between €4.37 and €8.11 for models estimated with $N \in [5, 25]$.

5.2. Relative Thinking

To illustrate the presence of relative thinking, the goal is to show that an increase with mean price also leads to higher search cost moments. Hence, as in the described example, the search cost estimates will be calculated to a larger group of article pools to check wether an increase in the mean price of the group leads to larger search cost estimates. Besides the search cost mean, the median and standard deviation are also taken into account. Price dispersion doesn't only occur due to search frictions. As mentioned earlier in the literature overview, determinants such as payment risk, shop expensiveness or delivery conditions can play a crucial role in pricing. Therefore, a hedonic pricing equation can be used to control for these factors. After that, the residuals of this regression will be extracted and used to estimate the search costs again.

5.2.1. Hedonic Pricing Equation

To ensure reliable estimation of shop fixed effects and minimize incidental parameter bias, the sample is restricted to shops with at least 500 price observations Restriction (2). Next, the dataset gets restricted to articles that have between 30–40 price observations (Restriction (3)) to enable a good model-fit. This filtering process reduced the sample to 146,480 observations covering 118 shops and 4272 articles. The hedonic pricing regression that is run looks as follows:

$$\begin{aligned}\text{price}_i = {} & \beta_0 + \beta_1 \cdot \text{page}_i + \beta_2 \cdot \text{deliverydays}_i + \beta_3 \cdot \text{refunddays}_i \\ & + \beta_4 \cdot \text{numratings}_i + \beta_5 \cdot \text{shoprating}_i \\ & + \sum_{j=1}^{20} \gamma_j \cdot \text{payment}_{ij} + \sum_{k=1}^{15} \delta_k \cdot \text{delivery}_{ik} \\ & + \varphi_i \cdot \text{shopname}_i + \varepsilon_i \end{aligned} \tag{5.1}$$

The regression consists of all scrapable information that was accessible on the price comparison website. The dependent variable *price* refers to the price including delivery costs for a given article. *Page* is a numeric variable indicating on which page number within a subcategory an article appears. Since articles are ordered by popularity, those on the first page are more popular than those on subsequent pages. The effects of popularity on pricing and price dispersion have been extensively studied in the economic literature (see e.g., Baye et al.(2004)[7] or Pan et al. (2002)[29]). Next, *deliverydays* indicates how many days an article takes to arrive if ordered at the scraping point in time. *Refunddays* describes the number of days a customer is allowed to return a product to the retailer free of charge. Visitors on Idealo can rate shops within each category. The total number of ratings and the average shop rating are captured by *numratings* and *shoprating*, respectively. Several payment methods are included as dummy variables, since each method offers a different level of security for the customer. For example, payments via Mastercard or Visa offer comprehensive buyer protection, while methods like immediate transfer or payment in advance are considered less secure. Including payment method dummies helps mitigate the influence of payment risk on prices. To account for delivery convenience—and to some extent payment-related risk—a dummy for different delivery services is also included. Since some services are less popular, they may be perceived as

riskier by consumers. Lastly, the variable *shopname* controls for shop-specific effects such as reputation, branding, or customer service.
To estimate search costs while controlling for the covariates described above, only the residuals from regression 4.1 are relevant. For readers interested in the results of this regression, the corresponding results are represented in figure A1 in the appendix. Columns (1) and (2) are based on sample restriction (1), while columns (3) and (4) correspond to restriction (3). After including shop fixed effects, most estimates for payment and delivery services are dropped due to perfect multicollinearity.

5.2.2. Results

To illustrate relative thinking, several pooled article groups are constructed in the same manner as described in the example in 5.1.3. For each group, the average price and the search cost moments are determined. If relative thinking is present, then an increase in the average price should be associated with a larger search cost mean/median. Hence, the following regressions will be estimated:

$$\log(\text{SC Mean}_i) = \beta_0 + \beta_1 \cdot \log(\text{MeanPrice}_i) + \varepsilon_i \tag{5.2}$$

$$\log(\text{SC Median}_i) = \beta_0 + \beta_1 \cdot \log(\text{MeanPrice}_i) + \varepsilon_i \tag{5.3}$$

SC Mean reflects the overall search cost average of an article group i. *MeanPrice* is the respective price average of i. The results for the effects on the mean can be seen in table 5.3.

Column (1) uses the full sample, where each product typically has between 30 and 40 price observations. Columns (2) to (4) rely on subsamples defined by Restriction (3). Column (2), like Column (1), analyzes raw prices and serves as a robustness check to address potential sample selection bias—specifically, whether the subset of price observations with complete covariate information differs systematically from those with missing covariates. Column (3) is based on residuals from an OLS regression, while Column (4) uses residuals from a fixed-effects regression.

Consistent with the theory of relative thinking, Columns (1) to (3) support the hypothesis that search cost estimates increase with the average price level. Based on Column (2), a 1% increase in the average price is associated with a 0.35% increase in the estimated mean search cost. This effect is similar in magnitude when using the full sample in Column (1). Even after controlling for observables and estimating search costs from the residuals

Relative Thinking for Search Cost Mean				
	Dependent variable:			
	ln(Search Cost Mean)			
	(1)	(2)	(3)	(4)
ln(Mean price)	0.401^{***}	0.345^{***}	0.263^{***}	0.023
	(0.034)	(0.100)	(0.046)	(0.024)
Constant	-0.567^{***}	-1.022^{**}	1.933^{***}	2.602^{***}
	(0.167)	(0.480)	(0.220)	(0.115)
Observations	1,133	306	306	306
R^2	0.110	0.037	0.097	0.003
Adjusted R^2	0.110	0.034	0.094	-0.0003
Note:	*p<0.1; **p<0.05; ***p<0.01			
	SC = Search Costs. Results are based on pricing data with delivery costs.			

Table 5.3: Relative Thinking - Search cost mean

in Columns (3), a significant positive relationship between mean search costs and average market price persists, though the magnitude of the effect is smaller. This attenuation is theoretically plausible: accounting for store-level heterogeneity reduces the expected gains from search for consumers. Consequently, to explain observed price differences, consumers would engage in more extensive search, leading to lower estimated search costs. Finally, when controlling for shop fixed effects in Column (4), the relationship becomes statistically insignificant. This is expected, as including store fixed effects removes all between-store variation, leaving only within-store variation. Next, the same analysis is conducted on the median of search cost estimates. The results are stated in table 5.4:

Relative Thinking for Search Cost Median				
	Dependent variable:			
	ln(Search Cost Median)			
	(1)	(2)	(3)	(4)
ln(Mean price)	0.427^{***}	0.362^{***}	0.350^{***}	0.010
	(0.036)	(0.090)	(0.074)	(0.032)
Constant	-1.262^{***}	-1.696^{***}	1.025^{***}	2.317^{***}
	(0.179)	(0.431)	(0.352)	(0.155)
Observations	1,133	306	306	306
R^2	0.108	0.050	0.069	0.0003
Adjusted R^2	0.108	0.047	0.066	-0.003
Note:	*p<0.1; **p<0.05; ***p<0.01			
	SC = Search Costs. Results are based on pricing data with delivery costs.			

Table 5.4: Relative Thinking - Search cost median

The column specifications are the same as in table 5.3. Similar as in the analysis for the search cost mean, a positive and significant relationships for the sample specifications (1) – (3) can be found. The strength of the relative thinking effect is similar for (1) and (2) compared to table 5.3. As opposed to the search cost mean, the median is less affected than the mean after estimating the search costs for the residuals of the OLS-case (compare (3) in table 5.3 and table 5.4). Also, the results for the residuals of the fixed-effects regression are insignificant in this case as well.

Finally, to observe how search cost heterogeneity in consumer search behavior changes if the price varies, the standard deviation of the search cost estimates has also been extracted and used as a dependent variable in the following regression:

$$\log(\text{SearchCostSD}_i) = \beta_0 + \beta_1 \cdot \log(\text{MeanPrice}_i) + \varepsilon_i \tag{5.4}$$

SearchCostSD describes the corresponding search cost standard deviation in a an article group. In table 5.5, besides an increase in the search cost mean and median, search cost heterogeneity also seems to increase.

Search Cost Heterogeneity

	Dependent variable:			
	ln(Search Cost SD)			
	(1)	(2)	(3)	(4)
ln(Mean price)	0.356***	0.313***	0.200***	0.016
	(0.030)	(0.089)	(0.034)	(0.021)
Constant	-0.186	-0.553	2.098***	2.569***
	(0.148)	(0.425)	(0.165)	(0.101)
Observations	1,133	306	306	306
R^2	0.111	0.039	0.100	0.002
Adjusted R^2	0.110	0.036	0.097	-0.001
Note:	*p<0.1; **p<0.05; ***p<0.01			
	SD = Standard Deviation. Results are based on pricing data with delivery costs.			

Table 5.5: Search cost heterogeneity

A 1% increase in the mean price of an article group, increases the standard deviation of search costs by 0.31–0.36 % if no covariates are considered on average. If residuals from an OLS regression are utilized, a 1% increase in average price increases search cost standard deviation by 0.2%. For the residuals of a fixed-effects regression there are no significant findings. Therefore, not only do larger average prices lead to higher search cost means and medians, but they do also affect the corresponding search cost uncertainty.

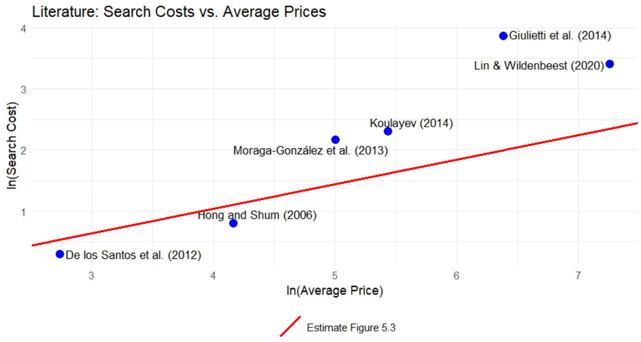
6. Discussion

Literature Comparison

This thesis aimed to understand how an increase in absolute price dispersion, accompanying a rise in mean prices, can be partially explained by a behavioral bias known as relative thinking, where consumers tend to exhibit higher search costs as the price scale increases. First, it was demonstrated that the German digital market for electronics is characterized by substantial and rising price dispersion. Across all sample specifications, absolute price dispersion measures are positively correlated with the respective mean price. Estimates from the two log-log models yield coefficients of approximately 0.8. This is notably similar to the early findings of Pratt et al. (1979)[30], who reported an estimate of 0.89 for the relationship between log standard deviation and log mean price despite analyzing prices from brick-and-mortar stores and across content-wise different product categories (e.g., bicycles or liquor). Similarly, Pan et al. (2001)[29] found estimates around 0.9 for both log-log regressions after analyzing 6,739 price quotes for 581 products from 105 e-tailers. The lin-lin model using the standard deviation as the price dispersion measure produced estimates ranging from 0.54 to 0.97. These values are substantially larger than the estimate reported by Sorensen (2000)[36] for prescription drugs, which was 0.18. However, Sorensen used wholesale costs as the dependent variable and controlled for purchase frequency, whereas this thesis focuses on final consumer prices, including transportation fees, and does not account for purchase frequency due to data limitations. In contrast, Aalto-Setälä (2003)[1] found a coefficient of 0.81 when analyzing grocery store prices in Finland, which aligns closely with the findings of this thesis.

Comparing search cost estimates across the literature is not a straightforward task. Methodologies vary widely within the field, making it difficult to draw direct comparisons. For example, if one were to apply another search cost estimation techniques to the data used in this thesis the resulting estimates would likely differ. Nonetheless, it is notable that average search cost estimates tend to increase when more expensive goods are analyzed

Figure 6.1 presents various search cost estimates found in the economic literature. While the number of recent academic studies quantifying search costs is limited, the figure provides suggestive evidence of a general pattern. The red line in the figure represents the



(a) Comparison

Author (Year)	Product	Price Range	Avg. Search Cost
Hong and Shum (2006)	Books	\$34–95	\$2.23
Lin & Wildenbeest (2020)	Medigap insurance market	\$1030–1813	\$30.00
Koulayev (2014)	Hotel room booking	\$230	\$4–16
De los Santos et al. (2012)	Books	\$8–23	\$1.35
Moraga-González et al. (2013)	Memory chips	\$116–182	\$8.70
Giulietti et al. (2014)	Energy contracts	\$592	\$47.30

(b) Overview

Figure 6.1: Search cost estimates comparison

estimated degree of relative thinking derived from Figure 5.3. This highlights that, within the literature, higher perceived search costs are often associated with products of higher prices. Interestingly, this pattern is further supported by experimental studies on relative thinking. Although the parameter estimated in this thesis, which is based on a log-log regression, cannot be directly compared in magnitude to experimental estimates, the behavioral implications are aligned. For instance, Karle et al. (2025)[23] model consumer utility using a perceived price function discounted by the price range: $v_{rt}(p, F) = \frac{p}{(b-a)^{\rho}}$, where $\rho = 0$ implies no relative thinking. They show that ρ is not statistically different from 1, indicating the presence of relative thinking. Similarly, Sommerville (2022)[35] finds that when both the mean price and price range increase, consumers spend on average 17% more due to reduced search effort. Both Karle et al. and Sommerville artificially increased the price scale in their experimental designs to isolate the effect of relative thinking—a setup that mirrors the empirical pattern observed in this thesis (see the log(range)–log(mean price) regressions in Tables 4.1 and 4.2). While the methods used to identify relative thinking and estimate search costs differ across studies, a consistent pattern emerges: consumers tend to rely more on relative rather than absolute measures when making decisions which leads to inflated search costs

Limitations

Obviously, inferring search costs just from a given price distribution only works if strong assumptions are imposed. One of the main assumptions is the underlying search strategy in which consumers engage in. Search cost estimates can differ substantially when sequential search is assumed as opposed to non-sequential search. Given the online context, it is very likely that consumers engage in non-sequential search since they could simply copy the article name from a given price quota and search for additional price alternatives while opening separate tabs. However, the model also assumes that consumers obtain a random price quota from the given price distribution F. As Lin Wildenbeest (2020) already pointed out, consumers usually engage in directed search. This is a prominent way of searching where some firms are more dominant than others. Baye et al. (2016) explain that firms engage in search engine optimization (SEO). This affects the appeared order of firms after searching for a certain good. Consumers would then search for higher ranked firms first after retrieving additional price quotas from less present firms. Since this thesis has no access to click-stream data, one could account for the fact of directed search and the presence of SEO if this data is available.

Next, it is important to note that the data is only available for a single time period. Therefore, it is not possible to empirically verify whether firms actually set prices according to a mixed-strategy. However, facilitated by the low menu cost nature of digital markets, some studies have been able to show that firms do follow mixed-strategy pricing. For example, Moraga-González and Wildenbeest (2008) demonstrated this in the memory chip market, Lewis (2008) in the gasoline market, and Wildenbeest (2011) in supermarket product pricing. Importantly, the theoretical model assumes a symmetric equilibrium in which each firm draws prices from the same distribution. In reality, however, firms differ, for instance in reputation, reliability, or payment security, which likely allows some of them to draw from a higher segment of that distribution. This heterogeneity violates the symmetry assumption and can bias the respective search cost estimates and ultimately the magnitude of relative thinking.

Additionally, the data was collected during the Christmas season, a period typically characterized by promotional activity and discount pricing. This may have unique implications for the observed levels of price dispersion and further challenges the assumption of symmetric mixed-strategy pricing. Some firms may engage more aggressively in discounting than others, leading to asymmetric pricing behavior that deviates from theoretical assumption of firms setting prices symmetrically. To account for heterogeneity in firm characteristics, a hedonic pricing regression was used to control for observable differences across sellers. The resulting search cost estimates were attenuated, and the previously observed positive correlation between average prices and search costs, i.e. relative thinking, weakened. This suggests that part of the relationship was driven by systematic differences between firms. Applying a fixed-effects specification went one step further by controlling for all observed and unobserved store characteristics. As a result, the correlation between mean prices and search cost estimates vanished. This outcome makes sense: fixed effects absorb all between-store variation, and the residuals used to estimate search costs now reflect only within-store price variation across products. Since relative thinking primarily operates through cross-store price comparisons, i.e., consumers reacting to higher prices across different sellers, this behavioral mechanism is effectively shut down when only within-store variation is retained.

Alternative Explanations

Important to note is that relative thinking is only one potential source which could explain the higher search cost estimates. Several rational explanations have been proposed in the literature to account for why search costs tend to be high and increase with the price level.

One explanation is that firms may deliberately complicate the search for prices, making it more difficult for consumers to identify the best option (Ellison and Ellison, 2009)[15]. Additionally, consumers may underestimate the potential benefits of search and consequently exert insufficient effort in comparing alternatives. High-value products, such as energy contracts or insurance plans, often exhibit greater complexity, which can increase the cognitive burden of comparison.

In other cases, larger purchases may raise trust-related concerns, leading some consumers to avoid lower-priced options despite equivalent stated quality and service levels. Finally, it is possible that individuals purchasing more expensive goods systematically face higher search costs due to greater opportunity costs of time, as they may have higher income levels. However, one could also argue that wealthier individuals are, on average, more educated, more efficient in searching, and more likely to use price comparison tools (Karle et al., 2025[23]).

Beyond that, some authors suggest that purchase frequency helps explain the price dispersion puzzle. Expensive products are, on average, bought less frequently, which makes consumers less experienced and less familiar with the price distribution thus giving firms greater price-setting power. However, Sorensen (2000)[36] shows that in the market for prescription drugs, even after controlling for purchase frequency, price dispersion remains larger as prices increase.

7. Conclusion

Although digital markets are characterized by low search costs, empirical research using clickstream data has shown that consumers' willingness to obtain additional price quotes is limited (De Los Santos, 2012[13]; Bronnenberg et al., 2016[8]). Furthermore, both the search cost estimation and experimental literature suggest that estimated search costs increase as the price scale rises. Since price dispersion also tends to rise substantially with the base price, higher search costs—driven by the behavioral bias known as Relative Thinking—may partially explain the so-called Price Dispersion Puzzle.

The main findings of this thesis are as follows: First, price dispersion measures increase significantly in the German digital electronics market. Second, based on the applied search cost estimation procedure, search costs also tend to increase with higher base prices. These results suggest the presence of Relative Thinking in consumer behavior. However, given the strong assumptions necessary to derive search cost estimates solely from observed price distributions, the absolute magnitude of these estimates should be interpreted with caution. Nevertheless, when considered in relation with prior literature on search cost estimation and experimental evidence on Relative Thinking, there is a robust pattern indicating that search costs indeed rise with the price level.

These findings carry important implications for consumer behavior. Given that the average net hourly wage in Germany ranges between €15–20, searching for better prices is a low-effort strategy to achieve substantial savings. Consumers should be made aware of the Relative Thinking bias and be especially encouraged to invest additional search effort for higher-priced goods, where the potential for absolute savings increases due to higher price dispersion. Moreover, using price comparison websites to access multiple price quotes simultaneously is strongly recommended.

Future research could enhance the precision of search cost estimates by incorporating clickstream data, which allows modeling of search behavior in vertically differentiated or directed search environments. This would enable relaxation of some of the stronger assumptions currently required. Additionally, if consumer-level data can be matched with browsing behavior, it may be possible to uncover the extent of heterogeneity in search costs which helps to determine whether empirical estimates reflect actual cognitive effort or rather misspecified beliefs. Finally, this thesis also shows that price dispersion varies significantly across product categories. This suggests that, even in digital markets, search behavior could be category-specific. Alternatively, other explanations beyond search costs may be necessary to fully account for the different rise in price dispersion.

Bibliography

[1] Ville Aalto-Setälä. Price dispersion and search costs in grocery retailing. Technical report, Working paper, National Consumer Research Centre, Helsinki, Finland, 2003.

[2] Donald WK Andrews. A conditional kolmogorov test. *Econometrica: Journal of the Econometric Society*, pages 1097–1128, 1997.

[3] Diego Aparicio, Zachary Metzman, and Roberto Rigobon. The pricing strategies of online grocery retailers. *Quantitative Marketing and Economics*, 22(1):1–21, 2024.

[4] Ofer H Azar. Relative thinking theory. *The Journal of Socio-Economics*, 36(1):1–14, 2007.

[5] Ofer H Azar. Relative thinking in consumer choice between differentiated goods and services and its implications for business strategy. *Judgment and Decision Making*, 6(2):176–185, 2011.

[6] Ofer H Azar. Firm strategy and biased decision making: The price dispersion puzzle. *Applied Economics*, 45(7):901–910, 2013.

[7] Michael R Baye, John Morgan, and Patrick Scholten. Price dispersion in the small and in the large: Evidence from an internet price comparison site. *The Journal of Industrial Economics*, 52(4):463–496, 2004.

[8] Bart J Bronnenberg, Jun B Kim, and Carl F Mela. Zooming in on choice: How do consumers search for cameras online? *Marketing science*, 35(5):693–712, 2016.

[9] Erik Brynjolfsson and Michael D Smith. Frictionless commerce? a comparison of internet and conventional retailers. *Management science*, 46(4):563–585, 2000.

[10] Kenneth Burdett and Kenneth L Judd. Equilibrium price dispersion. *Econometrica: Journal of the Econometric Society*, pages 955–969, 1983.

[11] Benjamin Bushong, Matthew Rabin, and Joshua Schwartzstein. A model of relative thinking. *The Review of Economic Studies*, 88(1):162–191, 2021.

[12] Mark Coppejans and A Ronald Gallant. Cross-validated snp density estimates. *Journal of Econometrics*, 110(1):27–65, 2002.

[13] Babur De los Santos. Consumer search on the internet. *International Journal of Industrial Organization*, 58:66–105, 2018.

[14] Peter A Diamond. A model of price adjustment. *Journal of economic theory*, 3(2):156–168, 1971.

[15] Glenn Ellison and Sara Fisher Ellison. Search, obfuscation, and price elasticities on the internet. *Econometrica*, 77(2):427–452, 2009.

[16] Victor M Fenton and A Ronald Gallant. Qualitative and asymptotic performance of snp density estimators. *Journal of Econometrics*, 74(1):77–118, 1996.

[17] A Ronald Gallant and Douglas W Nychka. Semi-nonparametric maximum likelihood estimation. *Econometrica: Journal of the econometric society*, pages 363–390, 1987.

[18] Yuriy Gorodnichenko, Viacheslav Sheremirov, and Oleksandr Talavera. Price setting in online markets: Does it click? *Journal of the European Economic Association*, 16(6):1764–1811, 2018.

[19] Han Hong and Matthew Shum. Using price distributions to estimate search costs. *The RAND Journal of Economics*, 37(2):257–275, 2006.

[20] Ali Hortaçsu and Chad Syverson. Product differentiation, search costs, and competition in the mutual fund industry: A case study of s&p 500 index funds. *The Quarterly journal of economics*, 119(2):403–456, 2004.

[21] Idealo.de. Über den idealo preisvergleich. Accessed: 2025-06-15.

[22] Greg Kaplan and Guido Menzio. The morphology of price dispersion. *International Economic Review*, 56(4):1165–1206, 2015.

[23] Heiko Karle, Florian Kerzenmacher, Heiner Schumacher, and Frank Verboven. Search costs and context effects. *American Economic Journal: Microeconomics*, 17(2):127–161, 2025.

[24] Jun B Kim, Paulo Albuquerque, and Bart J Bronnenberg. Online demand under limited consumer search. *Marketing science*, 29(6):1001–1023, 2010.

[25] Sergei Koulayev. Search for differentiated products: identification and estimation. *The RAND Journal of Economics*, 45(3):553–575, 2014.

[26] Haizhen Lin and Matthijs R Wildenbeest. Nonparametric estimation of search costs for differentiated products: Evidence from medigap. *Journal of Business & Economic Statistics*, 38(4):754–770, 2020.

[27] José Luis Moraga-González, Zsolt Sándor, and Matthijs R Wildenbeest. Semi-nonparametric estimation of consumer search costs. *Journal of Applied Econometrics*, 28(7):1205–1223, 2013.

[28] José Luis Moraga-González and Matthijs R Wildenbeest. Maximum likelihood estimation of search costs. *European Economic Review*, 52(5):820–848, 2008.

[29] Xing Pan, Brian T Ratchford, and Venkatesh Shankar. Why aren't the prices of the same item the same at me. com and you. com?: Drivers of price dispersion among e-tailers. *Com and You. Com*, 2001.

[30] John W Pratt, David A Wise, and Richard Zeckhauser. Price differences in almost competitive markets. *The Quarterly Journal of Economics*, 93(2):189–211, 1979.

[31] Brian Ratchford, Gonca Soysal, Alejandro Zentner, and Dinesh K Gauri. Online and offline retailing: What we know and directions for future research. *Journal of Retailing*, 98(1):152–177, 2022.

[32] Steven Salop and Joseph Stiglitz. Bargains and ripoffs: A model of monopolistically competitive price dispersion. *The Review of Economic Studies*, 44(3):493–510, 1977.

[33] Fabio Sanches, Daniel Silva Junior, and Sorawoot Srisuma. Minimum distance estimation of search costs using price distribution. *Journal of Business & Economic Statistics*, 36(4):658–671, 2018.

[34] Babur De Los Santos, Ali Hortaçsu, and Matthijs R Wildenbeest. Search with learning for differentiated products: Evidence from e-commerce. *Journal of Business & Economic Statistics*, 35(4):626–641, 2017.

[35] Jason Somerville. Range-dependent attribute weighting in consumer choice: An experimental test. *Econometrica*, 90(2):799–830, 2022.

[36] Alan T Sorensen. Equilibrium price dispersion in retail markets for prescription drugs. *Journal of Political Economy*, 108(4):833–850, 2000.

[37] Dale O Stahl. Oligopolistic pricing with sequential consumer search. *The American Economic Review*, pages 700–712, 1989.

[38] George J Stigler. The economics of information. *Journal of political economy*, 69(3):213–225, 1961.

[39] Richard Thaler. Toward a positive theory of consumer choice. *Journal of economic behavior & organization*, 1(1):39–60, 1980.

[40] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *science*, 211(4481):453–458, 1981.

[41] Hal R Varian. A model of sales. *The American economic review*, 70(4):651–659, 1980.

[42] Matthijs R Wildenbeest. An empirical model of search with vertically differentiated products. *The RAND Journal of Economics*, 42(4):729–757, 2011.

[43] Xiaolin Xing. Can price dispersion be persistent in the internet markets? *Applied Economics*, 42(15):1927–1940, 2010.

[44] Hejun Zhuang, Peter TL Popkowski Leszczyc, and Yuanfang Lin. Why is price dispersion higher online than offline? the impact of retailer type and shopping risk on price dispersion. *Journal of Retailing*, 94(2):136–153, 2018.

A. Appendix

Hedonic Pricing Regression Results

	Dependent variable:			
	Price			
	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)
Page	0.339***	0.418***	0.721***	0.654***
	(0.027)	(0.026)	(0.067)	(0.065)
Delivery Days	11.717***	21.042***	9.513***	17.869***
	(0.417)	(0.479)	(1.029)	(1.178)
Refund Days	-3.397***	-10.276***	-2.337***	4.579
	(0.055)	(2.571)	(0.123)	(5.979)
Number of Ratings	-0.0001***	-0.0001	-0.0001***	-0.001
	(0.00001)	(0.001)	(0.00002)	(0.002)
Shoprating	6.392***	-28.009*	7.547***	-47.628
	(1.000)	(14.608)	(2.498)	(35.585)
Paypal	-63.414***	271.023	-108.033***	
	(5.705)	(184.305)	(13.701)	
Visa	-593.473***		-466.246***	
	(6.047)		(14.821)	
Mastercard	484.949***		337.726***	
	(6.042)		(15.130)	
IT	-38.876***		-4.628	
	(1.611)		(3.930)	
PiA	48.933***		54.648***	
	(2.157)		(5.172)	
Amazon Pay	-26.373***		-1.554	
	(2.052)		(5.265)	
Debit Note	-128.417***		-109.809***	
	(1.682)		(4.023)	
Order on Account	-27.422***		-14.738***	
	(1.696)		(4.139)	
American	20.792***		-76.895***	
	(2.221)		(5.582)	
Diners	277.760***		171.513***	
	(3.232)		(8.279)	
Giropay	149.913***		93.989***	
	(2.648)		(6.708)	
Maestro	129.907***		318.593***	
	(2.712)		(6.976)	

Cash on Delivery	17.069^{***}	$-2,611.303^{***}$	-44.742^{***}	
	(3.304)	(149.026)	(8.494)	
eps	877.474^{***}		160.839^{***}	
	(4.741)		(14.702)	
Jcb	-316.403^{***}		-498.694^{***}	
	(17.493)		(80.459)	
Creditcard	-189.088^{***}		-115.143^{***}	
	(4.817)		(11.473)	
DHL	-47.386^{***}		-26.548^{***}	
	(2.076)		(4.979)	
DPD	45.642^{***}		6.245	
	(1.994)		(5.016)	
Spedition	121.948^{***}		89.411^{***}	
	(1.742)		(4.485)	
GLS	10.805^{***}		-31.986^{***}	
	(2.960)		(7.184)	
GoGreen	195.692^{***}		205.629^{***}	
	(4.089)		(9.793)	
Hermes	-109.957^{***}		-82.556^{***}	
	(2.474)		(6.081)	
Click and Collect	-248.415^{***}		-58.388^{***}	
	(3.458)		(8.504)	
UPS	4.731^{**}		22.774^{***}	
	(2.137)		(5.303)	
Deutsche Post	143.671^{***}		197.734^{***}	
	(2.528)		(6.535)	
TOF	-210.996^{***}		-214.115^{***}	
	(5.892)		(14.054)	
German Express Logistics	-54.951^{***}		51.846^{***}	
	(4.550)		(11.389)	
TNT	-72.475^{***}		-82.112	
	(17.178)		(50.043)	
Constant	466.267^{***}		431.745^{***}	
	(8.079)		(19.307)	
Observations	1,372,910	1,372,910	146,480	146,480
R^2	0.137	0.002	0.065	0.002
Adjusted R^2	0.137	0.002	0.064	0.002
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$			
	IT = Immediate Transfer, PIA = Payment in Advance			

A1: Hedonic Pricing Regression Results