IoT Market Dynamics: An Analysis of Device Sales, Security and Privacy Signals, and their Interactions

Swaathi Vetrivel, Brennen Bouwmeester, Michel van Eeten, and Carlos H. Gañán, *Delft University of Technology*

https://www.usenix.org/conference/usenixsecurity24/presentation/vetrivel

This paper is included in the Proceedings of the 33rd USENIX Security Symposium.

August 14–16, 2024 • Philadelphia, PA, USA

978-1-939133-44-1

IoT Market Dynamics: An Analysis of Device Sales, Security and Privacy Signals, and their Interactions

Swaathi Vetrivel[*], Brennen Bouwmeester[†], Michel van Eeten[‡], and Carlos H. Gañán[§]

Delft University of Technology

Abstract

We explore the relationship between the Security and Privacy (S&P) of IoT devices and their sales, considering the S&P signals in the context of these sales. We obtained expert S&P ratings of IoT devices from a European consumer association and the corresponding sales data from a leading Dutch online store. We complemented this with additional information like user ratings, the number of reviews and update support duration from two Dutch online stores. Our regression model shows that, holding other variables constant, a one-standard-deviation increase in S&P ratings corresponds to a noteworthy 56% boost in sales. Crucially, we observe a possible correlation between price and demand for S&P; at lower prices, the sales of IoT devices are directly proportional to the S&P rating, but this relationship diminishes as price increases. Further, we find that the presence of update support duration information, intended as a security signal, corresponds to higher S&P ratings and, all else being constant, also corresponds to a 69% increase in sales. While the exact causal mechanisms for the boost in sales remain unclear, our findings suggest positive incentives might be at play for IoT devices offering S&P at affordable prices and presenting relevant S&P information at the point of purchase.

1 Introduction

A widespread consensus among security researchers states that many consumer Internet-of-Things (IoT) devices exhibit vulnerabilities that compromise user privacy and security. Despite the severe consequences of these vulnerabilities, ranging from individual breaches of privacy [11] to large-scale DDoS attacks [3, 28], there is no dent in the growth of IoT devices. The consumer IoT market is projected to grow steadily at 5% per annum from 2023 to 2028, reaching a market volume of around US$232 billion by 2028 [49].

[*]s.vetrivel@tudelft.nl
[†]b.j.bouwmeester@student.tudelft.nl
[‡]m.j.g.vaneeten@tudelft.nl
[§]c.hernandezganan@tudelft.nl

Observers argue that the sale of insecure IoT devices indicates market failures, namely externalities and information asymmetry [7, 12, 27, 44]. Externalities arise since manufacturers escape the consequences of poor security once the devices are in the market. Similarly, consumers might also undervalue security, since many of the negative effects of their compromised devices end up with third parties, as in the case of DDoS attacks. Despite this lack of incentives, recent studies show that consumers not only care for IoT security and privacy, but they are also willing to pay more for it [15, 19]. However, it is unclear if these preferences for Security and Privacy (S&P) will translate into actual purchase decisions due to the lack of relevant S&P information during purchase. This information asymmetry implies consumers cannot factor S&P in their decisions, which in turn, implies devices with better S&P are not rewarded with higher sales. However, is this conjecture supported by empirical evidence?

In this study, we empirically examine this conjecture by conducting a comprehensive three-fold analysis of consumer purchase decisions on online stores. We analyse the purchase decisions using sales as a proxy, investigate information asymmetry in the context of these decisions, and explore whether presence of information intended to act as a security signal is associated with higher or lower sales.

Specifically, we first examine whether consumer preferences for IoT S&P are reflected in their real life purchase decisions. Using expert ratings of IoT S&P, including expert update ratings, obtained from the main consumer association in the Netherlands and the corresponding monthly IoT sales data from a leading Dutch online retailer, we answer the research question '*To what extent does the expert security and privacy rating of an IoT device correlate with its sales*?' Next, we study the information asymmetry in the context of these purchase decisions by evaluating whether information presented to consumers on online stores contain S&P signals. The information on online stores, like user product ratings, already serves as signals for device quality and vendor reputation [41] and influence purchase decisions [10, 13, 54]. Moreover, some user reviews on online stores also contain

S&P related information [50]. In the absence of relevant S&P information, these signals of quality might act as proxies for S&P. To systematically examine if these signals are consistent with the S&P of an IoT device, we ask '*To what extent are expert ratings of S&P consistent with user ratings?'* We also analyze the presence of update support duration, presented as part of the product information on Dutch online stores due to a government mandate. It was intended to act as a security signal since IoT devices without updates can become unsafe [1]. We compare this to ratings from the consumer association asking '*To what extent are expert ratings of updates consistent with presence of update support information?*' Third, we also examine if the presence of update support duration is associated with a difference in sales through our final research question *'Does the availability of update support information of IoT devices on online stores correspond to higher sales?*'

To answer these research questions, we used three data sets. The Dutch consumer association, through its access to the technical labs commissioned by the European network of consumer associations, gave us access to expert ratings for four IoT device types: IP cameras, smart printers, smart speakers and smart watches. The expert ratings include the S&P rating, the update rating, and the overall device rating, an aggregate rating from various tests. From the leading Dutch online retailer mentioned earlier, we obtained the monthly sales data for the tested devices and the average prices they were sold at per month. The retailer opted to remain unnamed, and we will refer to them as "Winkel". Finally, we complemented the sales data with web scrapes of the product pages at Amazon.nl and at Winkel. We collected the average user rating, number of reviews and product description including update support duration information for the tested devices.

To the best of our knowledge, this is the first study to use ground truth data on sales from a large e-commerce retailer. Although we focus on the Dutch market, prior research shows that IoT device popularity is similar across a variety of countries [29]. Moreover, we want to emphasise that this is an observational study. While we study the relationship between IoT device sales and other factors, we do not claim to establish causality. Rather, we want to test whether the observational data is consistent with our current notions of lack of reward for IoT devices with better S&P.

To answer our first research question, we construct regression models. We find that, after controlling for other factors like price of the device, a one standard deviation or 1.5 unit increase in the expert S&P rating corresponds to a remarkable 56% increase in sales. In contrast, an increase in the expert overall rating corresponds to only a 17% increase in sales. Moreover, we observe that the relationship between sales and expert S&P ratings is moderated by price. Across all IoT device types, at lower prices, higher S&P ratings correspond to higher sales, but this relationship diminishes as the price increases.

For our second research objective of evaluating information asymmetry, we find mixed results. We don't find any evidence of correlation between the expert S&P ratings and the user ratings. With respect to the update support duration information, we find that a mere 4.3% of the product links under consideration contain complete update support duration information on both the online stores, while 27.8% of the links contain complete information on at least one of the stores. However, our findings indicate that the few IoT devices with complete update support information on both Amazon and Winkel exhibit a significantly higher expert rating for updates. This suggests that availability of update support duration information on online stores is reflective of the general update practices of manufacturers and can reduce information asymmetry.

For our third research question on the difference in sales due to presence of update support information, we extend the regression model, built to answer our first research question, with update information status on Amazon.nl and Winkel. We find that, after controlling for the other factors, devices with complete update support information on Winkel correspond to 69% higher sales. This suggests that there are some positive incentives at play for sellers and manufacturers that provide the additional information although the exact mechanism of the incentive is beyond the scope of our research. Our empirical study thus provides a nuanced perspective on IoT device interactions driven by real-world sales data and expert IoT ratings. Theoretically, our findings help understand how barriers to security from economic theory, like information asymmetry and externalities, play out in real markets.

Finally, although tangential to our research goals, our data allows us to capture one additional scientific benefit: we can quantify to what extent public data from online stores can be a proxy for ground truth on sales, since we obtained the latter from Winkel. We find that the number of reviews on Winkel is a reasonable proxy for the sales data from Winkel. In contrast, the number of reviews on Amazon.nl showed low correlations with the Winkel sales data. This suggests that although the proxy is useful, it specific to the retailer.

The paper is structured as follows: In Section 2, we review related research on consumers' IoT purchase decisions and information asymmetry. Section 3 provides details on the datasets used. Section 4 explores the relationship between expert S&P ratings and sales, answering our first research question, while Section 5 and Section 6 address our second and third research question respectively. Section 7 discusses our findings and offers recommendations, and Section 8 concludes the paper.

2 Related Work

In this section, we elaborate on related studies on various aspects of consumer purchase decisions in the market for IoT devices. We also present other work that evaluated the security and privacy of IoT devices using different techniques.

2.1 Consumer purchase decisions and signals for S&P

There has been a fair amount of prior work that studied consumer purchase decisions on online stores. Several studies have found evidence that high user ratings on online stores correlate with higher sales of products [10, 18, 34, 54]. Similarly, several studies have noted that the volume of reviews on the platforms also influence sales [2, 13, 31, 45].

With regard to S&P and purchase decisions, interviews with retail customers highlighted that the point of purchase is an opportune moment for informing consumers about the security of the devices purchased [37]. Other studies delved deeper into this in the IoT context, analysing purchase decisions when adequate information about the S&P of the device is provided. Blythe, Johnson, and Manning [5] find evidence that providing clear security information before purchase can promote secure device selection. Gopavaram et al. [20] found that consumers were willing to pay more for privacy when there were clear indicators of the privacy level. Emami-Naeini et al. [15] empirically confirmed these findings in an incentive compatible user study and showed that consumers are willing to pay more for S&P when relevant information is presented.

There has also been prior work that evaluated different modes of presentation of S&P. Various forms of information labels have been proposed including graded labels, labels denoting S&P features, and labels indicating approval through independent assessment [23]. Emami-Naeini et al. [17] received positive feedback for labels with a similar design to food nutrition labels. Morgner et al. [35] found support for provisioning update support duration information, particularly from those perceiving higher risks from use of IoT devices. Although security labels have not yet found widespread adoption in the market, display of update support duration information has been in effect on Dutch online stores since 2020.

Based on these studies, we consider average user ratings, number of reviews and update support duration information as possible variables that influence purchase decisions. Moreover, to increase the generalizability of our findings, we collect these variables from two popular online stores in the Netherlands – Amazon.nl and Winkel.

2.2 Estimates of IoT device popularity

Earlier studies have estimated the popular devices in the market, which can be taken as an indicator of their sales, in a variety of ways. These studies are typically positioned as device identification in the wild and they use various techniques to identify the brand, make and model of deployed IoT devices. Several of these use network traffic to identify IoT devices [33, 43, 46, 51], however, we do not include them here since these analysis are local to the network from which the measurement was done and cannot be generalized. Other techniques that allow for more generalizability include passive sampling of network scans and active experiments done as ISPs and IXPs [42] and DNS fingerprints [38]. Taking a different approach, Kumar et al. [29] conducted a large-scale empirical study of IoT devices installed in homes of users of Avast antivirus software. They show that barring a few exceptions, the popularity of the four IoT device types that we consider – IP cameras, smart printers, smart speakers and smart watches – are largely similar across geographies.

2.3 Estimates of security of popular IoT devices

Junior et al. [24] estimated the security of popular IoT devices by analyzing the vulnerabilities present in their companion apps. For device selection, they started with the 100 most popular smart hubless devices on Amazon and then filtered it to devices that use WiFi for communication and then further to the categories of smart plugs, bulbs and infrared controllers. They find many devices share a common app and that 50% of apps corresponding to 38% of devices do not use proper encryption techniques. Moreover, there are papers [48, 53] that evaluate the security of popular IoT devices without specifying the metric used to define popularity for the device selection which lends their results incomparable with ours. In our work, we use sales figures for popularity and estimates of device security and privacy from the tests conducted by commissioned labs across Europe.

2.4 Correlation between S&P and sales

To the best of our knowledge, there are no studies that estimate the relationship between IoT device security and privacy, and its sales. However, earlier studies have analyzed the relationship at the level of a firm – between a firm's security and privacy posture and its stock market evaluation. Boroomand et al. [6] found that as a firm's investment in data privacy increases, its market valuation decreases. They attribute this to information asymmetry – firms' are not able to communicate their increased investment in privacy and consumers are not able to reward it despite valuing privacy. A similar study [52] evaluated the effect of investment in Data Privacy and Security (DPS) on the market value of two kinds of firms – those dealing with Business Data Analytics (BDA) and those that do not. They found DPS investment decreases a firms systematic risk and that these effects are higher for non-BDA firms when compared to BDA firms.

3 Dataset Description

In this section, we describe the three datasets from four sources that we used in our study. First, we received expert IoT device ratings from *Consumentenbond* (CB), a consumer welfare organization in the Netherlands. Next, the ground truth of sales data of IoT devices was generously provided by 'Winkel,' a leading online retailer in the Netherlands. Both of these datasets contain competitive data and are therefore typically confidential, sensitive, and strategically significant, which often restricts their availability for research. However, Consumentenbond and Winkel were exceptionally generous in providing these datasets for our study. Finally, we collected publicly visible attributes on online stores that prior work (Section 2) has identified as influencers in purchase decisions: user rating, number of reviews, and information on update support duration. We collected these from the IoT device product pages of two sources, the websites of Winkel and Amazon.nl, the Amazon store for the Netherlands. Table 1 provides an overview of the sources and the data obtained from each of them. In the following sections we describe these data sets. We use the term user ratings to refer to the average user ratings on Amazon and Winkel, and expert ratings to refer to the CB ratings.

Table 1: Overview of the collected data.

Dataset	Data Source	Data obtained
Expert IoT device ratings	Consumentenbond (CB)	Security and privacy ratings Update ratings Overall device rating
IoT sales data	Winkel	Monthly sales figures Average price per month
Information from product pages	Amazon and Winkel	User rating Number of reviews Update support duration

3.1 Expert IoT device ratings from Consumentenbond

The Consumentenbond is a non-profit consumer welfare organization in the Netherlands.[1] It works on a membership model and currently has about 420,000 members (about 5% of all Dutch households). Through a network of professional technical test labs commissioned by various European consumer associations, they conduct independent tests of various devices and provide valuable advice to consumers to enable informed purchases.

We received all the test results for the four IoT devices types tested by CB - IP Cameras, smart printers, smart speakers, and smart watches. All the tests were conducted between 2016, the start of IoT device testing by CB, and 2023. The testing process remained largely stable over this time period, with most devices being tested soon after their release date. In some cases, often due to the devices gaining popularity after their initial release, there were delays between device release date and the testing date.

The tests conducted and ratings given by CB are exhaustive and span a broad range, however, we only used the three expert ratings that were relevant for our study – the S&P rating, the update specific rating and the overall device rating. The S&P rating is an aggregate of the results of different sub-tests evaluating the password policy, number of vulnerabilities, updates and privacy concerns like data sharing. The heterogeneity in the device types, in terms of attack surface, privacy concerns and device functionality, leads to the necessity of device specific tests. Thus, the S&P sub-tests per device type vary both in number and type. Unfortunately, our confidentiality agreement with CB prevents us from sharing the test specifics and details on how the S&P rating is derived, but for each device type, the areas covered by the sub-tests are added in Appendix A.1. The update rating captures the clarity about update availability, and in the case of smart speakers, also the automatic update options available on the companion mobile apps. Although the update rating is included in the S&P rating, we also analyze it independently, as updates represent one of the most visible aspects of manufacturers' commitment to security from a consumer perspective. The overall rating is an aggregate of the S&P rating, update rating and other ratings like ease of use, and some device specific ratings like sound quality. Figure 1 shows the distribution of all three expert ratings across the different device types, the ratings are between 1 and 10. Although the S&P ratings are skewed towards higher values in some cases, we find enough variation among the ratings for a meaningful analysis of the relationship between S&P and market performance.

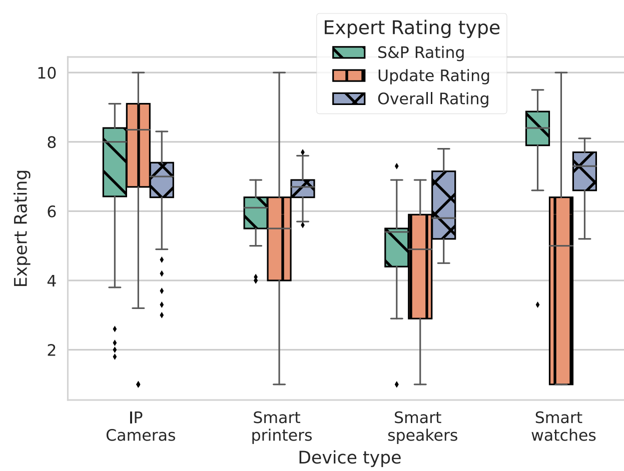


Figure 1: Distribution of expert ratings across the four device types

In total, we had the expert ratings for 469 IoT devices (130 IP cameras, 174 smart printers, 31 smart speakers and 134 smart watches) tested by the Consumentenbond. The average S&P rating across the four device types is 6.9 with the lowest being 1 for a smart speaker and the highest 9.5 for a smart watch. As seen in Figure 1, even within the same device type, there is a wide spread in the S&P ratings. This suggests that

[1] https://www.consumentenbond.nl

despite the potential disincentives for manufacturers to focus on S&P, at least some IoT devices have better S&P than others. We also observe variance across the four device types. The average S&P rating is the lowest for smart speakers (4.8) followed by smart printers (6) and IP cameras (7.3), with smart watches having the highest average S&P rating (8.3). The low average S&P rating for smart speakers is primarily due to the low scores of the companion mobile apps. It is important to note that these ratings are from the device population that was tested by CB and cannot be generalized to the device types.

The average update rating across all devices (5.2) is lower than the S&P rating (6.9). Smart watches have the lowest average rating for updates (4) closely followed by smart speakers (4.4) and smart printers (5), while IP cameras have the highest average of 7 for updates. The lower average update rating across IoT devices aligns with findings from other studies that indicate that update deployment tends to be slower for IoT devices [39]. The overall rating has an average of 6.8 across the device types with the lowest being 3 for an IP camera, and the highest 8.1 for a smart watch. In contrast to S&P rating and update rating, the average overall ratings across the device types do not show much variation.

3.2 Sales data of IoT devices

As mentioned earlier, we got sales data of IoT devices from a leading online retailer in the Netherlands. This was for a period of almost four years, from January 2019 to August 2023, and contained two datasets. The first dataset had the sales figures and average prices of IoT devices that were tested by CB. We used European Article Number (EAN), a unique 13 digit identifier for products, to match the CB test results with the Winkel sales data. The unit for the sales figures is the number of items of each device that were ordered and also reached the customer. Like most online retailers, Winkel also uses dynamic prices that are determined by a range of factors. For the purpose of this study, along with the monthly sales data, we obtained the monthly average prices at which the devices were sold for 385 devices (113 IP cameras, 132 smart printers, 28 smart speakers and 112 smart watches) that were tested by CB. We were not able to obtain the sales data for 18% of the devices tested by CB either due to the devices not being sold on Winkel or because we did not have the EAN code to match the devices with the sales data.

The second dataset had the sales figures and average prices of IoT devices within the same device type category that were not part of CB's testing. Comparing these two allowed us assess the selection bias in the devices tested by CB. To ensure the tests are relevant for a larger customer base, CB uses third party market research data to select and evaluate the more popular devices. When we compare tested vs. non-tested devices, we can indeed see that the tested devices have higher sales. Figure 2 shows the cumulative distribution of sales over the years of devices tested by CB and those not.

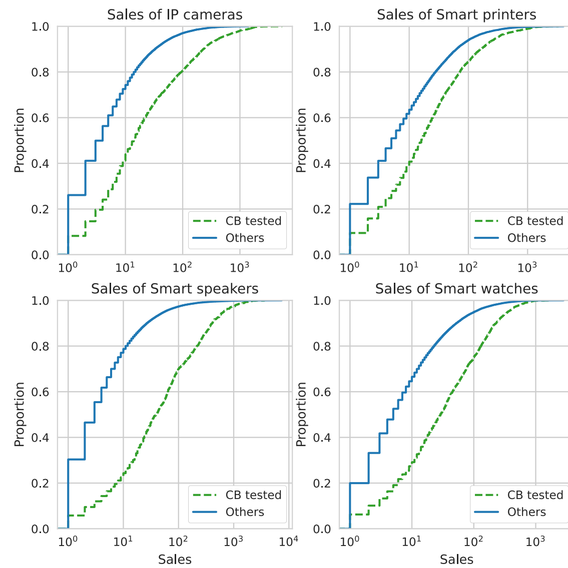


Figure 2: CDF of sales of devices tested by the Consumentenbond and those not

Across all the device types, the average sales per month of devices tested by CB is 91.36 (min: 1, median: 19, max: 5,353) while its 20.7 (min: 1, median: 4, max: 7,394) for others that were not in the CB data set. The comparison also underscores the rationale behind CB's choice to prioritize testing popular devices. As observed, approximately 20% or more of the devices across all the device types that CB has not tested have recorded only a single sale. We also find that the devices tested by CB are marginally more expensive than the other devices, especially for smart speakers and watches. The ECDF of the price comparison is added in the Appendix A.2.

3.2.1 Corroborating Winkel sales data with country level sales

To ensure that the sales data obtained from Winkel is representative of the national trends, we compared it to a commercially available Dutch market study data that is occasionally used by CB for device selection. Although we did not have this data for smart printers and speakers, for IP cameras and smart watches, a Spearman's Rank correlation test revealed a statistically significant and high positive correlation between the two data sets (0.72 and 0.75). This suggests that although there might be slight differences in the sales across specific stores, the sales data from Winkel is reflective of the national trends in sales of IoT devices.

3.3 Information from Product Pages

As noted earlier, we complemented the sales data with information scraped from two popular online stores, including user ratings, the number of reviews, and the duration of update support. We collected this data by scraping the product pages of the devices under consideration on Amazon.nl and Winkel. The preliminary collection was done between May and June 2023, with some missing details added in July 2023. Out of the 469 devices with expert S&P ratings, there were 14 devices that were not found on either platform. Of the remaining 455 devices, at the time of collection, 335 were found on both, 32 were found only on Amazon[2] and 88 were found only on Winkel.

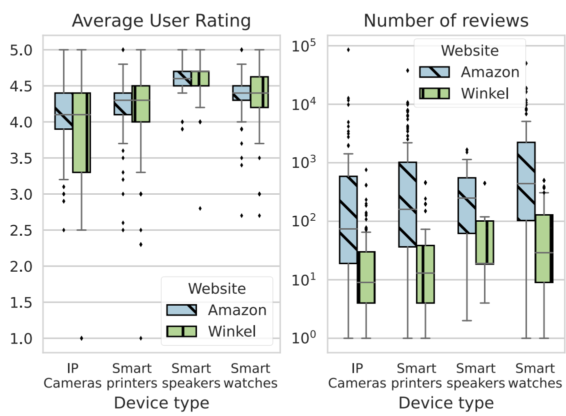


Figure 3: Distribution of user ratings and total reviews on Amazon

3.3.1 User ratings and number of reviews

In total, we scraped the details of 367 devices from Amazon, across 455 product links (122 IP cameras, 170 smart printers, 39 smart speakers, and 124 smart watches). On Winkel, we scraped details of 423 devices across 505 product links (140 IP cameras, 185 smart printers, 45 smart speakers, and 135 smart watches). On both Amazon and Winkel, the number of product links is higher than the number of devices because, for some devices, we found multiple product links due to variations in colours, sellers, or device bundles. In all such cases, we collected information from all the product links and treated them as separate instances rather than aggregating them per device to allow for a detailed analysis of update support information.

Figure 3 shows the distribution of user ratings and number of reviews on both the platforms. We find that the mean of user ratings are comparable across both the websites, 4.26 on Amazon and 4.17 on Winkel. However, the differences in number of reviews is staggering. The average number of reviews on Amazon is 1792, while it is 64.85 on Winkel, a mere 3.6% of the average number of reviews on Amazon. This is likely due to Amazon's policy of aggregating reviews across all of its websites rather than a difference in popularity between Amazon.nl and Winkel. For instance, Amazon.nl also has reviews from Amazon.com, the US Amazon store and Amazon.de, the German Amazon store.

[2] Unless otherwise specified, all references to Amazon are limited to Amazon.nl

3.3.2 Update support information

As mentioned in the Introduction, after a government intervention, sellers agreed to provide consumers with information about the availability of updates for smart devices at the time of purchase [1]. Note that there is a distinction here between seller and manufacturer. The seller acts as an intermediary between the manufacturer and consumer, listing the products on the e-commerce store and managing the sales. As per the intervention, manufacturers must provide the information to the sellers, who, in turn, update it on the online store to inform consumers. This was agreed upon in 2020. We started our data collection in May 2023, approximately three years after this policy came into effect allowing us to also analyse the presence of update support duration information as a security signal.

Our study reveals differences in how the update support duration information is displayed on each website. On Amazon, the update information is presented as a date (e.g., '13 April 2030'), and the update support field is part of the product's technical specifications. On Winkel, the update availability is specified using three fields under 'Introduction and support' within the product specifications table. These three fields are *Introduction year, Introduction month,* and *Support with updates.* The update support is specified in terms of the number of months from introduction (e.g., 'At least 24 months after the date of introduction'). While this clarifies the manufacturer's commitment to update support, some products have incomplete information as not all three fields are filled out.

Of the 455 products on Amazon, only 54 have a valid date; for the majority, the field is either empty or says 'Unavailable.' We categorise both empty and unavailable update information as invalid. Moreover, for six of the 54 products, the date is in the past (e.g., '30 June 2022'), and for one, it says ('1 January 2099'). However, we conservatively count these as valid since it is beyond this research's scope to verify if the updates for these devices continued past 2022 or will continue till 2099. Of the 505 products on Winkel, 231 had no information on updates whatsoever, 113 had valid information with all three fields complete, and 161 had partly valid information. That is, there was information about the duration of update support, but the Introduction year or month fields were empty or vice versa. We tag these as 'partly valid.'

Figure 4 shows the distribution of update statuses across the device types. On Amazon, we find that none of the smart printers have any valid information about updates; among the other device types, there is little variation – they all contain similarly minuscule percentages of valid information. Similarly, on Winkel, smart printers have the highest percentage of invalid information (78%). However, around 46.6% of IP cameras and 37.5% of smart speakers have valid information, while 71.5% of smart watches contain partly valid information – they provide information on the updates but fail to mention additional information on introduction year and month.

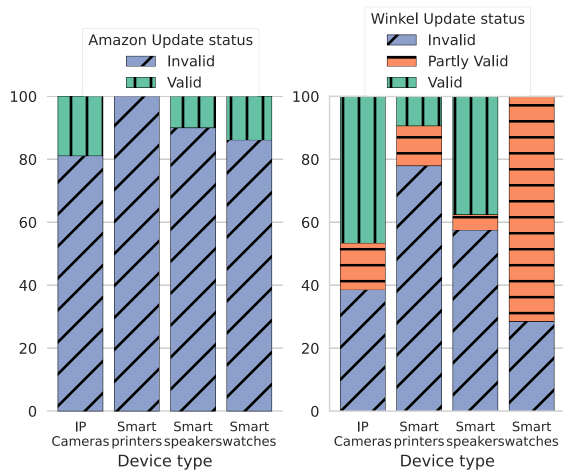


Figure 4: Distribution of Amazon and Winkel update statuses across device types

3.3.3 Comparison of Update support information on Amazon and Winkel

We find that for the 335 devices across 417 product links available on both websites, there is substantial divergence in the update statuses. Table 2 provides a comparison of the update field information. Only 18 devices, a meagre 4.3% of devices that are available on both platforms, provide valid update information on both websites, while an additional 27.8% provide valid information on one website but not on the other (7.9% have valid information on Amazon but not on Winkel and 19.9% on Winkel but not on Amazon). This suggests that the provisioning of update information is dependent on seller practices rather than the device-specific update support information provided by the manufacturer. Further, for none of the 18 devices with valid update statuses on both websites does the update support duration match. On Amazon, 14 of the 18 devices mention update support until 'April 2030,' whereas on Winkel, most devices mention update support for 24 months after introduction, with varying introduction years and months. This further highlights the role of sellers in provisioning this information and the difficulty in ensuring accuracy of such information.

4 Relationship between S&P and Sales

In this section, we analyse the datasets described so far to answer our first research question on the relationship between security and privacy of an IoT device and consumer purchase decisions with sales as a proxy. To achieve a holistic understanding of the interplay among various factors, we also consider the variables previously identified in the literature as influencers on sales, rather than restricting ourselves solely to S&P. Moreover, in order to comprehensively assess the interplay and associations of the different variables with the sales, we used a Generalised Linear Regression (GLM) model modelled using the glmmTMB package in R [8]. Since our dependent variable, the sales of IoT devices, is a count data, we had two options for the GLM – Negative Binomial and Poisson Models. The sales data exhibited over dispersion, meaning the variance was larger than the mean. Therefore, we used the Negative Binomial model since it accommodates excess variability. Moreover, to generalise interactions among variables across different device types, we employed a Mixed Effects model. This approach treats device type as a random effect and considers other variables as fixed effects, allowing us to derive broad-level inferences independent of device-specific variations in the test protocols.

Table 2: Comparison of update status between Amazon and Winkel for devices found on both

Amazon update status	Winkel update status	Count
Invalid	Invalid	163 (39.1%)
Invalid	Partly Valid	120 (28.8%)
Invalid	Valid	83 (19.9%)
Valid	Invalid	11 (2.6%)
Valid	Partly Valid	22 (5.3%)
Valid	Valid	18 (4.3%)

4.1 Explanatory Variable Selection

Our explanatory variables of interest are price, the expert S&P, update and overall ratings from CB, user rating and number of reviews on Amazon and Winkel. Prior to running the model, we conducted multicollinearity tests to remove explanatory variables that are highly correlated with each other. This ensures stability and reliability of our parameter estimates by mitigating issues that arise from high correlations among explanatory variables. We used Generalised Variance Inflation Factor (GVIF) to check for multicollinearity. Following suggested guidelines, we excluded the expert update rating from our analysis due to its GVIF value exceeding 5 [36]. This left us with eight remaining variables, Device type, price, expert S&P rating, and Overall rating from CB, user rating and number of reviews from Amazon and Winkel. The GVIF of the variables included in the model is added in Appendix A.3.

Moreover, we conducted bi-variate correlation tests between the dependent variable and the explanatory variables to get a sense of how each explanatory variable relates to the dependent variable, and validate its inclusion in the model. To control for multiple comparisons, we adjusted our p-values using the False Detection Rate (FDR) approach, implemented through the Benjamini & Hochberg method in the stats package in R [40]. There were no changes in the significance levels of the correlation tests after the adjustment. We explore these correlations briefly before analysing the results of the model.

4.1.1 Correlation between sales and price

We first analyse the correlation between the price and sales to better understand the price sensitivity of IoT devices. Since neither data followed a normal distribution, we used Spearman's rank correlation test. The results, as shown in Table 3, reveal a statistically significant weak negative correlation between price and sales for IP cameras and smart printers and a moderate positive correlation for smart speakers. This suggests that consumers possibly perceive IP cameras and smart printers as utilitarian, where price and affordability plays a decisive role in their purchase decisions. This observation may also be influenced by the competitive dynamics in these markets, where companies often compete on price. In contrast, for smart speakers which often come with advanced features and innovations, consumers might be more willing to pay a premium for additional features. Moreover, smart speakers are often part of a larger ecosystem and consumers invested in these ecosystems might be willing to pay more for compatibility.

Table 3: Correlation between sales and price from Winkel

Device type	Spearman correlation: Total sales vs. average price
IP Cameras	-0.251^{**}
Smart printers	-0.339^{***}
Smart speakers	0.517^{***}
Smart watches	0.071
Aggregate	0.006

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

4.1.2 Correlation between sales and expert ratings from CB

Next, to understand if and to what extent S&P features and the overall device quality as indicated by the CB expert ratings correspond with the sales of a device, we performed pair-wise correlations between each of them (Table 4). We found no significant correlations between the expert S&P rating of a device and the sales. This implies that a direct linear relationship between a device's S&P posture and its sales is not readily evident. We find statistically significant correlations between sales and expert overall rating across all devices and also at the aggregate level. Due to the significant correlations between the expert overall ratings and sales, we include it in the model. Moreover, despite the lack of correlation, we also include the expert S&P rating to gain a deeper understanding of the underlying relationship between S&P and sales which might be more complex than a linear association.

Table 4: Correlation between sales and CB ratings

Device type	Spearman correlation: Total sales vs Expert S&P rating	Spearman correlation: Total sales vs Expert Overall rating
IP Cameras	0.088	0.216^{**}
Smart printers	-0.035	-0.286^{***}
Smart speakers	0.100	0.358^{**}
Smart watches	0.037	0.333^{***}
Aggregate	0.052	0.195^{***}

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

4.1.3 Correlation between price and expert ratings from CB

Next, we explore the relationship between the average price and the expert CB ratings. As seen in Table 5, the price of IP cameras and smart watches exhibit a modest positive correlation with the expert S&P rating. In contrast, for smart speakers, we find a moderate negative correlation between the price and S&P rating. However, smart speakers are a unique case. Within the 27 smart speakers in our sample, four smart speakers with the highest sales collectively account for approximately 56% of all smart speaker sales. These four devices have prices higher than average and expert S&P ratings lower than average. This indicates that the negative correlation is possibly influenced by the disproportionate presence of these devices, and cannot be generalised.

With respect to expert overall rating, we find moderate positive correlations between the sales and average price at the aggregate device level and also at the individual device level for all devices except IP cameras. The lack of correlation for IP cameras might be due to various factors like a more mature market, diverse manufacturers, differentiated market segments based on price and features, or other market factors. Nonetheless, the high degree of associations between the average price and expert CB ratings for the other device types suggests that these two variables exhibit interaction, which could, in turn, influence sales. To accommodate this interaction, we incorporate an interaction term between price and expert CB ratings in the model.

4.1.4 Correlation between sales and user rating and number of reviews on Amazon and Winkel

We find many significant positive correlations between sales and both user ratings and number of reviews from Amazon and Winkel (Table 6), validating their inclusion in the model.

Table 5: Correlation between average device price and expert CB ratings

Device type	Spearman correlation	
	Average price vs	
	Expert S&P rating	Expert Overall rating
IP Cameras	0.259**	0.160
Smart printers	0.112	0.633***
Smart speakers	−0.486***	0.756***
Smart watches	0.429***	0.559***
Aggregate	0.092	0.44***

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Moreover, this analysis allows us to determine to what extent publicly available information like number of reviews can serve as a proxy for sales. We find that the sales of IoT devices show a statistically significant positive correlation with the number of reviews on both Amazon and Winkel. This is true at both the aggregate level, across all the device types, and also at the individual device level. The only exception is the correlation between the sales of IP cameras and the number of reviews on Amazon.

Furthermore, we find that the correlation between the sales and the number of reviews on Winkel (0.8) is much higher than on Amazon (0.259). This implies that the number of reviews on a platform can act as a reliable indicator for sales on that platform. Since Amazon aggregates its review count across all of its stores, the number of reviews on Amazon may not be a reliable proxy for sales.

Table 6: Correlation between total device sales and consumer metrics from Amazon and Winkel

Device type	Spearman correlation			
	Total sales vs			
	Amazon user rating	Winkel user rating	Amazon total reviews	Winkel total reviews
IP Cameras	0.529***	0.439***	0.094	0.762***
Smart printers	0.189	−0.001	0.278***	0.777***
Smart speakers	0.488**	0.397**	0.376**	0.651***
Smart watches	0.213**	−0.191*	0.338***	0.767***
Aggregate	0.415***	0.193***	0.259***	0.802***

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In contrast, with respect to the user rating, the reverse is true. The correlation between the sales and user rating on Amazon (0.415) is higher than the correlation with the user rating on Winkel (0.193). This suggests that even on the same platform, user rating is a poor indicator of popularity. However, since, similar to the number of reviews, the user ratings on Amazon are aggregated from different stores like Amazon.de, Amazon.com, etc., the Amazon ratings seem to be a marginally more reliable indicator of popularity. It is worth noting that the devices in our data set have a selection bias of being popular in the Netherlands and Europe, and within this set of devices Amazon user rating shows a moderate correlation with sales. Without further research, we cannot conclude if the finding will hold true in other geographies or for other devices.

4.2 Model Evaluation

Next, we outline the steps involved in Modelling. As mentioned earlier, we use a Mixed Effects Negative Binomial Model, treating the explanatory variables as fixed effects and the device type as a random effect. This helps us understand the overall trends that affect all devices while also capturing device type specific variations. To validate our decision to include the device type as a random effect, we compared the fit of two models using likelihood ratio test [30]. The first model contained device type as a fixed effect, and the second as a random effect. The p-value from the likelihood ratio test was significant ($p < 0.001$), indicating that the model with device type as a random effect was a better fit.

Moreover, as a preprocessing step, we scaled and centered the variables to ensure uniform influence and improve model convergence and interpretability. Scaling adjusts variables to have similar scales, preventing larger values from dominating. Centering sets variable means to zero, easing interpretation and providing a meaningful intercept.

We added the eight variables identified in the previous section in a step-wise forward manner. We explored different ways of ordering the fixed effect variables during the step-wise forward training but found that the ordering does not make any difference to the results. We therefore picked an intuitive ordering for ease of presentation. In the model presented, we start with the intercept only model, then add the device type, the average price, the expert S&P rating and the expert overall device rating, followed by user ratings and number of reviews on Winkel and Amazon. Further, as mentioned earlier, we added interaction terms between price, expert S&P rating and expert overall rating of CB. This enabled us to evaluate how the relationship between these variables and sales varies with changes in the other variables. This resulted in eight distinct models, which are shown in Table 13 in Appendix A.6.

To evaluate which model is best suited for our data, we checked the goodness of fit using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Log-likelihood, following best practices from literature [9, 14, 22, 26]. AIC balances fit and complexity, guarding against over fitting, while BIC penalises complexity, promoting simpler models. Lower AIC and BIC values signify better model fit [9, 14, 26]. Log-likelihood quantifies how well a statistical model explains observed data, with higher values indicating better fit [22]. Moreover, in a mixed effects model, two R^2 types are used to assess the model: Conditional R^2, which gauges the explanatory power of the fixed effects and Marginal R^2, which measures the combined explanatory power of both fixed and random effects. Higher values for both indicate improved model fit.

Based on these criterion, amongst the models in Table 13 (Appendix A.6), we identified Model 8 to be the best fit for our data. It has a lower AIC, BIC and higher log likelihood. Moreover, although the conditional R^2 is slightly lower than the previous model, it has the highest marginal R^2 amongst all the models. This indicates that the entire model, including the fixed and random effects, best explains the variation in our dependent variable, the total sales.

4.3 Model Interpretation

The Incidence Rate Ratios (IRR) of each of the explanatory variables in our model is shown in Figure 5. The IRR quantifies the multiplicative change in the rate of occurrence of an outcome (in our case, the sales) when the explanatory variable changes by one standard deviation, after controlling for the other variables. An IRR of 1 suggests no change, while greater and lesser than one denotes an increase and decrease in sales respectively. The interaction terms added between price, expert S&P rating and expert overall rating allows us to see the combined effect of these variables on the sales. Table 7 shows the mean and standard deviation of all the explanatory variables in the model.

Table 7: Mean and Standard Deviation of Explanatory Variables

Explanatory Variable	Mean	Standard Deviation
Avg Price	€179.10	€112.70
Expert S&P Rating	7.05	1.48
Expert Overall Rating	6.94	0.65
Winkel User Rating	4.18	0.64
Amazon User Rating	4.27	0.38
Winkel Total Reviews	67.21	120.11
Amazon Total Reviews	1966.33	6665.94

To address our first research question regarding the relationship between expert S&P ratings and consumer purchase decisions of IoT devices, we found that, when holding other variables constant at their mean values, an increase in the expert S&P rating by one standard deviation accounts for a 56% increase in sales (IRR of 1.56). In contrast, an increase in the expert overall rating by one standard deviation corresponds to only a 17% increase in sales (IRR of 1.17). Additionally, a one standard deviation increase in price corresponds to a 27% decrease in sales (IRR 0.73). Moreover, the significant interaction term between price and expert S&P rating indicates that the interaction between them dampens the effect of either variable on sales by 20% (IRR 0.8). On the other hand, the interaction between expert overall rating and S&P rating, corresponds to an increase in sales by 18% (IRR 1.18).

Figure 6, generated using the sjPlot package [32] in R, shows the interaction between price, expert overall rating and S&P rating. We observe that an increase in expert S&P rating corresponds to marginally higher sales for higher values of overall rating (the rightmost plot of Figure 6). Moreover, across all three plots, we observe that effect of S&P rating on sales is higher for lower values of price, and this effect gradually diminishes as the price increases. After a certain price point, changes in the S&P rating have little impact on sales.

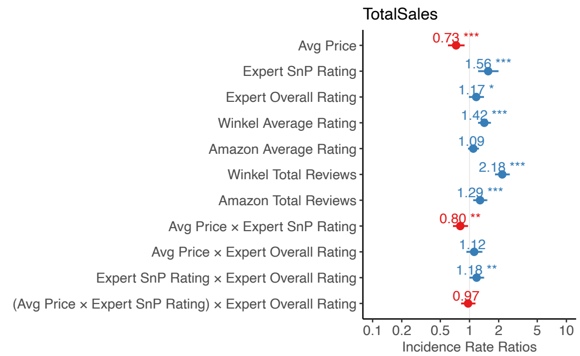


Figure 5: IRR from the Mixed effect Negative Binomial Model (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$)

This finding highlights the crucial relationship between price and S&P of a device. From our bi-variate correlation analysis (Table 5) we observe a positive correlation between the price and expert S&P rating for IP cameras and smart watches. The model results reveals the context of these positive correlations and suggests that an increase in S&P rating, while possibly corresponding to a marginally higher price, also corresponds to higher sales. Moreover, as observed via the interaction effects, the amplifying effect of S&P on sales is higher for lower prices, and gradually decreases as price increases. A crucial takeaway from this is that at the same price level, IoT devices with better S&P perform better in the market in terms of sales. In the next section, we evaluate whether this better performance can be attributed to the presence of S&P signals.

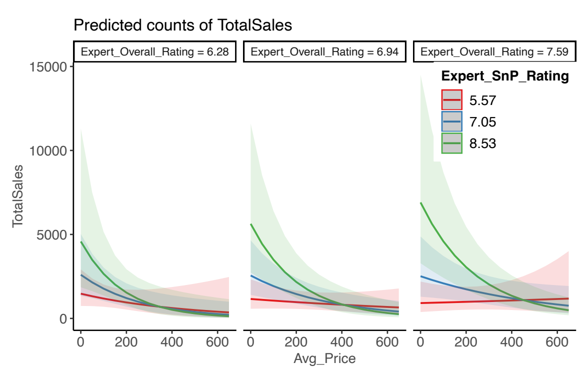


Figure 6: Interaction effects of Price, expert S&P rating and expert Overall rating on Total sales

With regard to the other explanatory variables, we see that the highest IRR (2.18) is for total reviews on Winkel. An increase in the number of reviews on Winkel by one standard deviation corresponds to a 118% increase in sales. In contrast, an increase in the number of reviews on Amazon by one standard deviation corresponds to only a 29% increase. This aligns with the results from our bi-variate correlation analysis: the number of reviews on Winkel is better proxy for the Winkel sales data than the number of reviews on Amazon.

With respect to user ratings, we observe no significant effect of the Amazon user rating on total sales, but the IRR for the Winkel user rating is 1.42. This indicates that while a change in the user rating on Amazon does not correspond to any change in the sales, an increase in the Winkel user rating by one standard deviation corresponds to a 42% increase in sales. This contradicts with our bivariate correlation analysis (Table 6), which shows positive correlation for both Amazon and Winkel user ratings, suggesting that the linear relationship between Amazon user rating and sales is possibly influenced by other confounding variables. Moreover, the increase in user rating on Winkel corresponding to an increase in sales underscores the role of consumer metrics in influencing the purchase decisions of consumers. In an online store, such mechanisms play a crucial role in signalling device quality and stimulating trust [41]. With respect to the random effect of the device type, we observe a variance of 0.234. This suggests that only a minimal amount of additional variability in sales can be explained by the IoT device type.

5 Relationship between expert ratings and information from product pages

In this section, we address our second research question assessing whether user ratings and update support duration information align with expert ratings. This allows us to evaluate the presence of S&P signals and the extent of information asymmetry in the context of the purchase decisions we analysed in Section 4.

5.1 Relationship between expert S&P ratings and average user rating

We first analyse the correlation between the expert S&P rating – which also includes the expert update rating – and the user ratings on Amazon and Winkel. Table 8 provides an overview of the Spearman correlations between the expert ratings and the Amazon and Winkel user ratings. Since our primary aim is to analyse if these ratings act as signals for S&P, we do not analyse the relationship with the expert overall rating. However for the interested reader, we present the correlation between user rating and overall ratings in the Appendix A.4.

At the aggregate level, across all the device types, we observe a low negative correlation between the expert update rating and the user rating on both Amazon and Winkel. Moreover, at the individual device level, although the expert update rating for smart speakers shows a low positive correlation with both the user ratings, the update ratings of both IP cameras and smart printers show a negative correlation with the Winkel user rating. The negative correlations could be a reflection of consumers' negative experiences with the update process [21] or a consequence of consumers not being able to account for good update practices into the user rating due to lack of visibility or information. Although invisibility is considered a desirable feature in security design [4], it can also lead to consumers not being able to account for security in their rating of the device.

Table 8: Correlation between CB S&P and update ratings and user ratings on Amazon and Winkel

Device type	Spearman correlation			
	Expert update rating vs		Expert S&P rating vs	
	Amazon user rating	Winkel user rating	Amazon user rating	Winkel user rating
IP Cameras	-0.007	-0.215^{**}	0.012	-0.104
Smart printers	-0.134	-0.376^{***}	0.141	0.416^{***}
Smart speakers	0.380^{**}	0.324^{*}	-0.118	-0.323^{*}
Smart watches	-0.005	0.098	0.204^{**}	0.008
Aggregate	-0.152^{***}	-0.23^{***}	-0.011	-0.077

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Regarding the expert S&P rating, we find that, at the aggregate level, across all device types, there are no significant correlations between the expert S&P rating and the user ratings on both Amazon and Winkel. At the individual device level, we observe a modest positive correlation between the expert S&P rating for smart watches and smart printers and the user ratings on Amazon and Winkel respectively. In contrast, the S&P rating of smart speakers shows a moderate negative correlation with the Winkel user rating. The negative correlation, taken together with the couple of low positive correlations shows that user rating cannot serve as a reliable indicator of the S&P posture of an IoT device. Rather, the user ratings likely reflect the overall consumer experience rather than an evaluation of the S&P posture of the device. Moreover, the differences in the correlations between the expert ratings and the user ratings across the device types is inline with earlier work that found that percentage of references to S&P varied across reviews for different IoT device types [50].

5.2 Relationship between update ratings and update support information

Next, we turn to analyzing the provisioning of the update support information and the expert update ratings of CB, to better understand if the update statuses are reflective of general update practices of manufacturers. If they are, they can serve as signals for S&P and help consumers make more informed purchase decisions with respect to IoT S&P. Table 9 shows the mean values of expert update rating for the different update information statuses on Amazon and Winkel. We observe that on both Amazon and Winkel, devices with valid update support information have a higher update rating. To check if the differences between the two groups on Amazon (Invalid and Valid), and the three groups on Winkel (Invalid, Partly valid and Valid) are significant, we use two different tests. For Amazon, we use the Mann-Whitney U test and for Winkel, we use Kruskal-Wallis test with an additional post-hoc Dunn's test if the results are significant.

Table 9: Mean values of update rating for different values of update status on Amazon and Winkel

Update Status	CB Update rating
Amazon	
Invalid (397)	5.2
Valid (48)	5.9
Winkel	
Invalid (191)	5.1
Partly valid (151)	4.7
Valid (103)	6.4

Our results shows that the differences on both Amazon and Winkel, among the different groups are significant. The statistically significant higher expert update ratings for devices with valid update support duration information imply that the mere availability of update support duration information reflects manufacturers' practices regarding updates. This might then serve as a signal for security, reducing the information asymmetry for IoT devices. Note that the this signal is only available for a minor percentage of devices under consideration: only 22.4% of devices on Winkel and 11.8% of devices on Amazon had valid information. Moreover, for the 18 devices for which the information was available on both websites, the update support duration was not consistent. Nonetheless, we verify if there is a correlation between the presence of valid update support information and higher device sales in the next section.

6 Relationship between Update Information Status and sales

In this section, we address our third research question on the relationship between provisioning update support information and sales. To do so, we extend the model and add categorical variables for update information status for Amazon and Winkel, in addition to the other explanatory variables. The results are shown in Table 12 in Appendix A.6. As mentioned earlier, on Amazon, the update status has two values, Invalid and Valid while on Winkel it has three values Invalid, Partly Valid and Valid. Table 10 shows the IRR for the update statuses. We find that, after controlling for the other factors in our model, the status of invalid on Amazon corresponds to a 34% decrease in sales (IRR 0.66) while a valid update status on Winkel corresponds to a 69% increase in sales. This implies that the presence of update support duration information, which serves as a reasonable signal for the S&P of an IoT device (from section 5), also corresponds to a positive change in sales. Although the underlying causal mechanisms are unclear, this suggests that there are positive incentives at play that will reward sellers and manufacturers who adopt initiatives like IoT security labels [16, 23, 47]. Moreover, our results align with other research [35] indicating that the availability of update support duration information has an impact of 8% to 35% on consumers' purchase decisions.

Table 10: IRR for Update Statuses on Amazon and Winkel

Update Status	IRR
Amazon Invalid (397)	0.66^{**}
Amazon Valid (48)	0.87
Winkel Invalid (191)	0.79
Winkel Partly valid (151)	1.1
Winkel Valid (103)	1.69^{*}

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

7 Discussion and Recommendations

Before we discuss our results and offer recommendations, we want to emphasise that this is an observational study. We have analysed the relationships between the variables but this is insufficient to establish causality. We found that despite the lack of S&P signals for a majority of devices, there is an alignment between consumer preferences for S&P – as established in prior work [15, 20] – and their purchase decisions, although the underlying causal mechanisms are unclear. We also find that the relationship is moderated by the price of the device. At lower prices, devices with higher S&P ratings correspond to higher sales while this relationship diminishes as the price increases.

Previous studies on security labels [15, 20] have found that consumers are willing to pay more for S&P when provided with information about it. Our results corroborate these findings and additionally suggest that, in the absence of security labels, the willingness to pay more for S&P exists up to a certain price point, beyond which the demand for S&P decreases. This dual emphasis on both security and price highlights the nuanced nature of consumer preferences in the IoT market.

This also underscores the challenge for manufacturers to balance better S&P at affordable prices. As we observe in our study, at the aggregate level, across all device types, and also for IP cameras and smart watches, the expert S&P rating positively correlates with price, suggesting that increase in S&P corresponds to an increase in price. This highlights a dilemma for manufacturers. They can focus on S&P at the risk of increasing the price and losing out on sales, or they can offer devices at comparable price points and enjoy higher sales. Note that the model results show that a one standard deviation increase in price (112.7 EUR or 121.35 USD), is associated with a 27% decrease in sales. While it is beyond the scope of this study to determine what types of security controls are feasible to implement within 112.7 EUR, a crucial takeaway is that at the same price points, IoT devices with one standard deviation or 1.5 unit increase in S&P correspond to a 56% increase in sales. This suggests that manufacturers who manage the balancing act between S&P and price stand to gain significant benefits.

With regard to the user ratings, we find limited correlation between the user ratings and expert S&P ratings suggesting that only a minority of the reviews and associated ratings might express S&P concerns that align with expert evaluations. This indicates that most consumers may either lack the ability to discern the S&P features of a device or may not consider them significant factors in their evaluations and ratings. Alternatively, even when factored into ratings, their S&P concerns might not align with expert evaluations, possibly due to friction with S&P configurations like 2FA [15, 50].

This emphasises an important aspect that needs to be balanced with S&P – usability. As the results of our correlation with expert update ratings highlight, better update ratings correspond to lower user ratings (Table 8). This could be a consequence of consumers not being able to account for good update practices into the user rating due to lack of information. It could also be a reflection of consumers' negative experiences and friction with the update process [21]. This emphasises the importance of designing S&P features with a focus on user experience and aiming to decrease such friction.

Moreover, increased transparency about the S&P features of an IoT device that informs users about the steps needed for the added S&P will enable better management of consumer expectations. As the results from our third research question show, on Winkel, devices for which complete update support information was available correspond to a 69% increase in sales. Although it requires further research to understand the underlying causal mechanisms for the boost in sales, our results, in line with other user studies [35], suggest that there might be positive incentives at play for IoT devices that contain S&P information at the purchase point. Crucially, availability of S&P information at purchase might increase the price point till which consumers are willing to pay more for better S&P. Thus, initiatives like security labels will yield benefits not only for the consumer, in terms of more informed purchase decisions, but may also for reward manufacturers and sellers with higher sales.

To that end, as our study results show, sellers play a key role in publishing relevant information on online stores. Three years after the mandate on provisioning update support duration information, we find that only 4.8% of devices analysed contain valid, albeit different update support duration on both stores. This implies that any effort aimed at improving manufacturer compliance, for security labels and the like, should also consider the various sellers who operate as intermediaries between manufacturers and consumers on e-commerce platforms. Although the onus is on the manufacturers to provide the information to the sellers and on the online store to enable provisioning of the information on the product pages, the task of updating the relevant information in the product pages still falls to the sellers. In the ongoing discussions on stakeholders in the realm of S&P IoT devices, these intermediaries are often overlooked. This oversight causes us to miss their perspective, which is important for the successful implementation of security labels. Unlike traditional brick-and-mortar stores, where these labels can be physically displayed on products with minimal intervention from store owners, online stores rely on sellers to present information as they deem appropriate. As we see with the case of update information, even if the online store has the provision to display this information, it often remains empty or has incomplete information. Therefore, it is vital to emphasise the necessity of transparency regarding the S&P features of a device, not only to the manufacturers, but also to the sellers.

7.1 Limitations

The expert ratings used in our study come from the tests conducted by the network of European technical test labs commissioned by various consumer associations in Europe. Although other tests using different constructs of S&P might arrive at different ratings, we believe these trusted ratings serve as a good indicator of the overall S&P posture of the device. A limitation of the testing of consumer associations is that they maximize for consumer benefit and therefore focus on devices that are popular in the market. While this introduced some selection bias in our data, the observed S&P ratings were diverse enough to meaningfully answer our research question.

Further, although our sales data is from a single e-commerce store in the Netherlands, we find that it is well aligned with the national market study data (refer subsection 3.2.1). Moreover, while any study based on a single country raises questions on generalizability, prior work estimating the popularity of IoT devices across geographies found that across all regions, 100 vendors account for nearly 90% of all IoT devices [29]. This suggests that while there might be minor differences in device popularity across regions, there is an over arching uniformity in vendor distribution which makes the implications from our findings applicable to e-commerce stores across geographies. Moreover, although including user reviews would have led to a richer analysis, we excluded them since prior work found that only 9.8% of user reviews of IoT devices contain references to S&P [50].

As we observe in our study, there is variations in the results among these four device types, which might limit their relevance to other IoT device types. We account for these variations in the model by treating the device types as random effects while studying the fixed effects of the other factors. Moreover, although the factors we included were based on prior literature, we acknowledge that there might be other confounding variables that influence consumer purchase decisions which we have not considered. Future work can evaluate the influence of other factors and deepen our understanding of the underlying causal mechanisms at play.

7.2 Research Ethics

We got approval for the study from the Ethics Review Board in our institution. We assessed our methodology against the ethical guidelines outlined in the Menlo Report for ethical practices in computing studies [25]. We only scraped publicly available information from Amazon and Winkel, and did not collect the associated usernames. Moreover, to ease the load on the servers, rather than use web crawlers, we fed specific pages to the scraping scripts. We also added delays and distributed the scraping over a longer duration to minimise server strain.

8 Conclusion

In this work, we analysed consumer purchase decisions and signals for S&P in the market context within which these decisions are taken. Our results showed that despite lack of information about S&P for a majority of the devices, a one standard deviation increase in the S&P rating of a device is associated with a 56% increase in sales. However, the relationship is moderated by the price of the device. The effect is stronger at lower prices and decreases corresponding to an increase in price. In addition, we find that availability of complete update support information on online stores reflects the update practices of manufacturers and can therefore act as a signal for S&P. Further, we also find that on one of the online stores, devices with complete update support information correspond to a 69% increase in sales. This suggests that there are positive incentives at play that will reward manufacturers and sellers who adopt S&P transparency initiatives like security labels with higher sales. Our results also highlight the crucial role of sellers in ensuring the success of such initiatives since they are responsible to update the relevant information on online stores.

Acknowledgments

This publication is part of the RAPID project (Grant No. CS.007), financed by the Dutch Research Council (NWO), and has also been partially supported by the INTERSCT project (Grant No. NWA.1160.18.301), also funded by NWO. The authors would like to thank the Consumentenbond and Winkel for generously providing the datasets for the analysis.

References

[1] ACM (Autoriteit Consument & Markt). Consumenten beter geïnformeerd over updates bij aankoop slim apparaat na actie acm, 2020. [Accessed: October 16, 2023].

[2] Naveen Amblee and Tung Bui. Harnessing the Influence of Social Proof in Online Shopping: The effect of Electronic Word-of-Mouth on Sales of Digital Microproducts. *International journal of electronic commerce*, 16(2):91–114, 2011.

[3] Manos Antonakakis, Tim April, Michael Bailey, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J Alex Halderman, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, Yi Zhou, Manos Antonakakis Tim April, Matthew Bernhard Elie Bursztein, Jaime J Cochran Zakir Durumeric Alex Halderman Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever Zane Ma, Joshua Mason, and Nick Sullivan Kurt Thomas. Understanding the Mirai Botnet. *USENIX Security '17*, 2017.

[4] Giampaolo Bella, Bruce Christianson, and Luca Viganò. Invisible security. In *Security Protocols XXIV: 24th International Workshop, Brno, Czech Republic, April 7-8, 2016, Revised Selected Papers 24*, pages 1–9. Springer, 2017.

[5] John M Blythe, Shane D Johnson, and Matthew Manning. What is security worth to consumers? investigating willingness to pay for secure Internet of Things devices. *Crime Science*, 9(1):1–9, 2020.

[6] Farzam Boroomand, Aija Leiponen, and Gurneeta Vasudeva. Does the market value attention to data privacy? evidence from US-listed firms under the GDPR. Technical report, Wharton Mack Institute working paper, 2022.

[7] Irina Brass, Leonie Tanczer, Madeline Carr, and Jason Blackstock. Regulating IoT: Enabling or Disabling the Capacity of the Internet of Things? *Risk&Regulation*, 33:12–15, 2017.

[8] Mollie E. Brooks, Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Maechler, and Benjamin M. Bolker. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, 9(2):378–400, 2017.

[9] Peter Bruce and Andrew Bruce. *Practical Statistics for Data Scientists*. O'Reilly Media, 2017.

[10] Judith A Chevalier and Dina Mayzlin. The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of marketing research*, 43(3):345–354, 2006.

[11] Barry Collins. Google Home Hub Shows Random Nest Cam Footage on Family's Device, May 2020. [Accessed: October 16, 2023].

[12] European Commission, Joint Research Centre, E Leverett, R Clayton, R Anderson, and G Baldini. *Standardisation and certification of safety, security and privacy in the 'Internet of Things'*. Publications Office of the European Union, 2017.

[13] Geng Cui, Hon-Kwong Lui, and Xiaoning Guo. The Effect of Online Consumer Reviews on New Product Sales. *International Journal of Electronic Commerce*, 17(1):39–58, 2012.

[14] Henry De and Graft Acquah. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics*, 2:1–6, 02 2010.

[15] P. Emami-Naeini, J. Dheenadhayalan, Y. Agarwal, and L.F. Cranor. Are Consumers Willing to Pay for Security and Privacy of IoT Devices? In *Proceedings of the 32nd USENIX Security Symposium*, January 2023.

[16] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. Ask the experts: What Should Be on an IoT Privacy and Security Label? In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 447–464. IEEE, 2020.

[17] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. Exploring How Privacy and Security Factor into IoT Device Purchase Behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[18] David Godes and Dina Mayzlin. Using Online Conversations to Study Word-of-Mouth Communication. *Marketing science*, 23(4):545–560, 2004.

[19] S. Gopavaram, J. Dev, S. Das, and L. J. Camp. IoT Marketplace: Willingness-To-Pay vs. Willingness-To-Accept. In *Proceedings of the 20th Annual Workshop on the Economics of Information Security (WEIS 2021)*, June 2021.

[20] Shakthidhar Reddy Gopavaram, Jayati Dev, Sanchari Das, and Jean Camp. IoTMarketplace : Informing Purchase Decisions with Risk Communication. Technical report, Working Paper, 2019, ftp://svn. soic. indiana. edu/pub/techreports/TR742. pdf, 2019.

[21] Julie M Haney and Susanne M Furman. Work in Progress: Towards Usable Updates for Smart Home Devices. In *International Workshop on Socio-Technical Aspects in Security and Trust*, pages 107–117. Springer, 2020.

[22] David W. Hosmer, Trina A. Hosmer, Saskia le Cessie, and Stanley Lemeshow. A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in medicine*, 16 9:965–80, 1997.

[23] Shane D Johnson, John M Blythe, Matthew Manning, and Gabriel TW Wong. The impact of IoT security labelling on consumer product choice and willingness to pay. *PloS one*, 15(1):e0227800, 2020.

[24] Davino Mauro Junior, Luis Melo, Hao Lu, Marcelo d'Amorim, and Atul Prakash. A Study of Vulnerability Analysis of Popular Smart Devices Through Their Companion Apps. In *2019 IEEE Security and Privacy Workshops (SPW)*, pages 181–186. IEEE, 2019.

[25] Erin Kenneally and David Dittrich. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. *Available at SSRN 2445102*, 2012.

[26] David R. Anderson Kenneth P. Burnham. *Model Selection and Inference: A Practical Information-Theoretic Approach*. JSTOR, 2002.

[27] Wolfgang Kerber. Governance of IoT Data: Why the EU Data Act Will not Fulfill Its Objectives. *GRUR International*, 72(2):120–135, 2023.

[28] Constantinos Kolias, Georgios Kambourakis, Angelos Stavrou, and Jeffrey Voas. DDoS in the IoT: Mirai and Other Botnets. *Computer*, 50(7):80–84, 2017.

[29] Deepak Kumar, Kelly Shen, Benton Case, Deepali Garg, Galina Alperovich, Dmitry Kuznetsov, Rajarshi Gupta, and Zakir Durumeric. All Things Considered: An Analysis of {IoT} Devices on Home Networks. In *28th USENIX security symposium (USENIX Security 19)*, pages 1169–1185, 2019.

[30] Fraser Lewis, Adam Butler, and Lucy Gilbert. A unified approach to model selection using the likelihood ratio test. *Methods in ecology and evolution*, 2(2):155–162, 2011.

[31] Yong Liu. Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of marketing*, 70(3):74–89, 2006.

[32] Daniel Lüdecke. *sjPlot: Data Visualization for Statistics in Social Science*, 2023. R package version 2.8.15.

[33] Markus Miettinen, Samuel Marchal, Ibbad Hafeez, N Asokan, Ahmad-Reza Sadeghi, and Sasu Tarkoma. IoT SENTINEL: Automated Device-Type Identification for Security Enforcement in IoT. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pages 2177–2184. IEEE, 2017.

[34] Wendy W Moe and Michael Trusov. The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, 48(3):444–456, 2011.

[35] Philipp Morgner, Christoph Mai, Nicole Koschate-Fischer, Felix Freiling, and Zinaida Benenson. Security Update Labels: Establishing Economic Incentives for Security Patching of IoT Consumer Products. In *2020 IEEE Symposium on Security and Privacy (S&P)*, pages 429–446. IEEE, 2020.

[36] Robert M O'Brien. Dropping Highly Collinear Variables from a Model: Why it Typically is Not a Good Idea. *Social Science Quarterly*, 98(1):360–375, 2017.

[37] Simon Parkin, Elissa M Redmiles, Lynne Coventry, and M Angela Sasse. Security when it is welcome: Exploring device purchase as an opportune moment for security behavior change. In *Proceedings of the Workshop on Usable Security and Privacy (USEC'19)*. Internet Society, 2019.

[38] Roberto Perdisci, Thomas Papastergiou, Omar Alrawi, and Manos Antonakakis. IoTFinder: Efficient Large-Scale Identification of IoT Devices via Passive DNS Traffic Analysis. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 474–489. IEEE, 2020.

[39] Vijay Prakash, Sicheng Xie, and Danny Yuxing Huang. Inferring Software Update Practices on Smart Home IoT Devices Through User Agent Analysis. In *Proceedings of the 2022 ACM Workshop on Software Supply Chain Offensive Research and Ecosystem Defenses*, pages 93–103, 2022.

[40] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0.

[41] Alvin E Roth. *Who Gets What - And Why: The New Economics of Matchmaking and Market Design*. Houghton Mifflin Harcourt, 2015.

[42] Said Jawad Saidi, Anna Maria Mandalari, Roman Kolcun, Hamed Haddadi, Daniel J Dubois, David Choffnes, Georgios Smaragdakis, and Anja Feldmann. A Haystack Full of Needles: Scalable Detection of IoT Devices in the Wild. In *Proceedings of the ACM Internet Measurement Conference*, pages 87–100, 2020.

[43] Matias RP Santos, Rossana MC Andrade, Danielo G Gomes, and Arthur C Callado. An efficient approach for device identification and traffic classification in IoT ecosystems. In *2018 IEEE Symposium on Computers and Communications (ISCC)*, pages 00304–00309. IEEE, 2018.

[44] Bruce Schneier. We Need to Save the Internet from the Internet of Things, 2023. Accessed on November 29, 2023.

[45] KungHsin Shao. The Effects of Controversial Reviews on Product Sales Performance: The Mediating Role of the Volume of Word of Mouth. *International Journal of Marketing Studies*, 4(4):32, 2012.

[46] Rahul Anand Sharma, Elahe Soltanaghaei, Anthony Rowe, and Vyas Sekar. Lumos: Identifying and Localizing Diverse Hidden {IoT} Devices in an Unfamiliar Environment. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 1095–1112, 2022.

[47] Yun Shen and Pierre-Antoine Vervier. IoT Security and Privacy Labels. In *Privacy Technologies and Policy: 7th Annual Privacy Forum, APF 2019, Rome, Italy, June 13–14, 2019, Proceedings 7*, pages 136–147. Springer, 2019.

[48] Omer Shwartz, Yael Mathov, Michael Bohadana, Yuval Elovici, and Yossi Oren. Reverse Engineering IoT Devices: Effective Techniques and Methods. *IEEE Internet of Things Journal*, 5(6):4965–4976, 2018.

[49] Statista. Consumer IoT - Worldwide, 2023.

[50] Swaathi Vetrivel, Veerle Van Harten, Carlos H Gañán, Michel Van Eeten, and Simon Parkin. Examining Consumer Reviews to Understand Security and Privacy Issues in the Market of Smart Home Devices. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 1523–1540, 2023.

[51] Lingjing Yu, Bo Luo, Jun Ma, Zhaoyu Zhou, and Qingyun Liu. You Are What You Broadcast: Identification of Mobile and IoT Devices from (Public) WiFi. In *USENIX Security Symposium*, pages 55–72, 2020.

[52] Yueyue Zhang, Cheng Zhang, and Yunjie Xu. Effect of data privacy and security investment on the value of big data firms. *Decision Support Systems*, 146:113543, 2021.

[53] Binbin Zhao, Shouling Ji, Wei-Han Lee, Changting Lin, Haiqin Weng, Jingzheng Wu, Pan Zhou, Liming Fang, and Raheem Beyah. A Large-Scale Empirical Study on the Vulnerability of Deployed IoT Devices. *IEEE Transactions on Dependable and Secure Computing*, 19(3):1826–1840, 2020.

[54] Feng Zhu and Xiaoquan Zhang. Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of marketing*, 74(2):133–148, 2010.

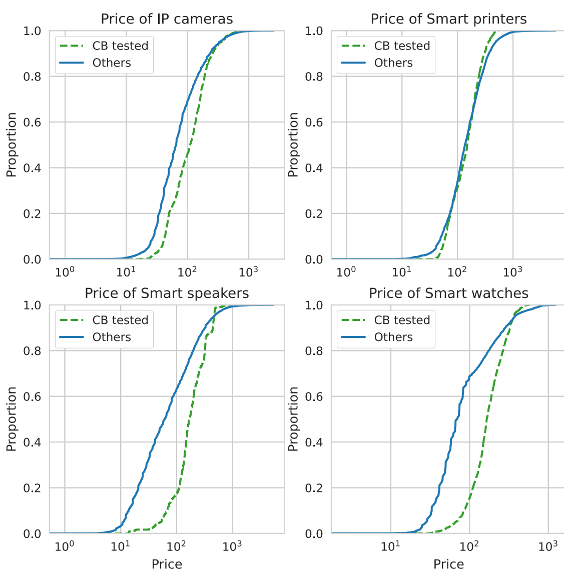
A Appendix

A.1 List of sub-tests for each device type

Table 11: List of test categories for each device type

IP Cameras	Smart Watches	Smart Printers	Smart Speakers
Password Policy	Standard Settings	Setup	Data Security
Standard Installation	Encryption	Access Controls	Decommissioning
Android App	Factory Reset	Password Policy	Password Policy
iOS App	Man-in-the-Middle	Updates	Network Security
Updates	Updates	Permissions	Update Policy
Known Vulnerabilities	Password Policy	Encryption	iOS App
	Account Management	Authentication	Android App
	Data Minimization	Known Vulnerabilities	Privacy Policy
	Options	Decommissioning	
	Choice Consequences		
	Vulnerability Hotline		

A.2 Price comparision between CB tested and non-tested IoT devices



A.3 GLM Variable Selection

Generalised Variance Inflation Factors

	GVIF	Df	GVIF^(1/(2∗Df))
Device type	4.623	3	1.291
Average price	1.534	1	1.239
Expert S&P Rating	2.881	1	1.698
Expert Overall Rating	1.804	1	1.343
Winkel user rating	1.345	1	1.596
Winkel total reviews	1.116	1	1.057
Amazon user rating	1.463	1	1.209
Amazon total reviews	1.071	1	1.034

A.4 Correlation between CB device ratings and average ratings on Amazon and Winkel

Device type	Spearman correlation: Expert Overall rating vs Amazon user rating	Spearman correlation: Expert Overall rating vs Winkel user rating
IP Cameras	0.144	0.141
Smart printers	0.159	0.054
Smart speakers	0.664^{***}	0.289
Smart watches	0.360^{***}	0.124
Aggregate	0.221^{***}	0.127^{**}

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A.5 GLM Results with Update Statuses

Table 12: Mixed Effect Negative Binomial Model with Update Statuses

Dependent variable: Total sales *Predictors*	*Incidence Rate Ratios*	*p*
(Intercept)	1807	< 0.001
Avg Price	0.74	0.001
Expert SnP Rating	1.63	< 0.001
Expert Overall Rating	1.23	0.014
Winkel User Rating	1.42	< 0.001
Amazon User Rating	1.15	0.029
Winkel Total Reviews	2.09	< 0.001
Amazon Total Reviews	1.25	0.006
Bol update status [Partly Valid]	1.1	0.768
Bol update status [Unavailable]	0.79	0.444
Bol update status [Valid]	1.69	0.065
Amazon update status [Unavailable]	0.66	0.042
Amazon update status [Valid]	0.87	0.59
Avg Price × Expert SnP Rating	0.91	0.333
Avg Price × Expert Overall Rating	1.15	0.125
Expert SnP Rating × Overall Rating	1.04	0.702
(Avg Price × Expert SnP Rating) × Expert Overall Rating	0.89	0.184
Random Effects		
σ2	0.63	
τ00 DeviceType	0.32	
ICC	0.34	
N DeviceType	4	
Observations	302	
Marginal R2 / Conditional R2	0.588/0.727	

A.6 GLM Results

Table 13: Mixed effect Negative Binomial GLM regression results

Dependent variable: Total sales									
	Model 0	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Predictors	*IRR*	*IRR*	*IRR*	*IRR*	*IRR*	*IRR*	*IRR*	*IRR*	*IRR*
(Intercept)	2654.54 (< 0.001)	2630.75 (< 0.001)	2692.34 (< 0.001)	2651.34 (< 0.001)	2102.05 (< 0.001)	1729.2 (< 0.001)	1743.41 (< 0.001)	1508.9 (< 0.001)	1537.83 (< 0.001)
Avg Price			0.66 (< 0.001)	0.62 (< 0.001)	0.56 (< 0.001)	0.59 (< 0.001)	0.6 (< 0.001)	0.72 (< 0.001)	0.73 (< 0.001)
Expert S&P Rating				1.32 (0.055)	1.17 (0.333)	1.2 (0.078)	1.24 (0.053)	1.42 (0.006)	1.56 (0.001)
Avg Price x Expert S&P Rating				0.86 (0.024)	0.85 (0.173)	0.87 (0.146)	0.88 (0.218)	0.81 (0.019)	0.8 (0.013)
Expert Overall Rating					1.51 (< 0.001)	1.41 (< 0.001)	1.39 (0.001)	1.17 (0.081)	1.17 (0.07)
Avg Price × Expert Overall Rating					1.15 (0.244)	1.17 (0.18)	1.17 (0.203)	1.15 (0.139)	1.12 (0.236)
Expert S&P Rating × Expert Overall Rating					1.22 (0.098)	1.19 (0.076)	1.17 (0.116)	1.15 (0.11)	1.18 (0.049)
(Avg Price x Expert S&P Rating) x Expert Overall Rating					0.86 (0.225)	0.8 (0.062)	0.78 (0.045)	0.95 (0.598)	0.97 (0.688)
Winkel User Rating						2.04 (< 0.001)	2.02 (< 0.001)	1.41 (< 0.001)	1.42 (< 0.001)
Amazon User Rating							1.05 (0.447)	1.14 (0.043)	1.09 (0.15)
Winkel Total Reviews								2.46 (< 0.001)	2.18 (< 0.001)
Amazon Total Reviews									1.29 (0.002)
Random Effects									
σ2		1.00	0.97	0.96	0.93	0.87	0.87	0.68	0.66
τ00		0.04DeviceType	0.19DeviceType	0.16DeviceType	0.11DeviceType	0.00DeviceType	0.00DeviceType	0.12DeviceType	0.23DeviceType
ICC		0.04	0.17	0.14	0.11			0.15	0.26
N		4DeviceType	4DeviceType	4DeviceType	4DeviceType	4DeviceType	4DeviceType	4DeviceType	4 DeviceType
Observations	304	304	304	304	304	304	304	304	304
R2 conditional / R2 marginal	NA / 0.000	0.000 / 0.041	0.132 / 0.276	0.195 /0.310	0.251 /0.331	0.450 /NA	0.453 /NA	0.623 /0.679	0.594 /0.701
AIC	5290.8	5289.4	5272.9	5270.4	5257.2	5210.1	5211.6	5080.1	5067.9
BIC	5298.2	5300.5	5287.7	5292.6	5294.3	5251.0	5256.1	5128.4	5119.3
LogLik	-2643.4	-2641.7	-2632.4	-2629.2	-2618.6	-2594.1	-2593.8	-2527.1	-2520.0