

New ways of measuring price development on consumer electronics

Kjersti Nyborg Hov, Ragnhild Nygaard
Division for price statistics
Statistics Norway

17th Ottawa group meeting
Rome, 7 – 10 June 2022

Summary

Today Statistics Norway makes regular use of scanner data for Consumer Price Index (CPI) and harmonised CPI (HICP) measurements for main areas like groceries, clothing, pharmaceutical products and fuels. Now the aim is to expand the use of scanner data to consumer electronics. Consumer electronics poses additional challenges in index compilation compared to the other consumer areas; they are notoriously prone to high item churn and rapid technological changes, often resulting in large and rapid quality improvements. This necessitates some form of quality assessment and adjustment in price index compilation, a task complicated by the large amount of data.

Scanner data is collected from the two leading retailers of consumer electronics in Norway. In order to calculate quality adjusted price indices using e.g. hedonic regression models, there is a need for item information. Since scanner data seldom contains very detailed item specifications, metadata has been scraped from the retailer's own web sites. The web scraped metadata has then been combined with scanner data to obtain more extensive item information.

We believe that hedonic price indices have the potential to improve practices of calculating indices for consumer electronics in the Norwegian CPI. A primary target of this study has been to find practical solutions for exploiting the full potential of the scanner data. Consumer electronics is a heterogeneous group of products, therefore separate hedonic models need to be set up and maintained over time. In addition, web scraping metadata which entails dealing with incomplete data as well as structuring and cleaning, is resource demanding and comprehensive, placing a large burden on the CPI staff. One possible strategy without compromising the results is to combine different methods for calculating quality adjusted price indices, depending on data properties of the product groups. Our study concentrating on product groups such as mobile phones, computers, flat screens and computer tablets indicates that satisfactory results may be achieved by relying on a limited set of explanatory variables, either in hedonic regression modelling or by applying and exploiting a broader item definition by calculating indices of homogenous products (HPs). Our study also shows that different calculation methods provide very different results. We believe that longer time series are needed before any valid conclusions can be drawn, nevertheless this study has given useful insight for compiling future indices on consumer electronics.

Keywords:

Price indices, CPI, hedonics, consumer electronics, scanner data, web scraping, TDH, Hedonic double imputation method

1. Introduction

Statistics Norway already makes regular use of scanner data¹ for Consumer Price Index (CPI) and Harmonised Index of Consumer Prices (HICP) measurements for consumer areas like groceries, clothing, sport equipment, pharmaceutical products and fuels. For many years there has been a strong focus on increasing the use of new data sources in price statistics, and in the Norwegian CPI in particular, in order to reduce the burden² on data providers and increase price index quality.

Since the early start of scanner data implementation in the Norwegian CPI in the mid-90s the calculation methods have gradually changed over time in light of new methodology being developed and presented internationally. Statistics Norway started out by mimicking traditional methodology by replacing price information on representative items from supermarket retailers with the corresponding transaction price³. In 2005 a bilateral superlative index formula based on all products for a sample of representative stores was used at detailed level in the index of food and non-alcoholic beverages. In 2013 the method was changed to an unweighted formula in order to minimize chances of indices drifting, and in 2021 a multilateral method for the index of food and non-alcoholic beverages was introduced (Johansen 2013, and Johansen, Nygaard 2021). The use of multilateral method was further extended in 2022 by applying it also for indices of non-food products from grocery stores and goods and services from petrol stations and kiosks.

In order to make price comparisons for products of equal quality, the scanner data method mostly applied at elementary level in the Norwegian CPI, is matching GTINs or other product IDs^{4,5} over time. Defining the product at item code level works well in case of relatively stable codes like for the majority of supermarket items. For items like consumer electronics however, such a detailed definition of a product is less appropriate as item (codes) are frequently changed. New calculation

¹ Scanner data is defined as transaction data that provides information on turnover and quantity sold of products identified by certain product IDs, for instance the GTINs, aggregated over a defined period. Currently, Statistics Norway is in the starting phase of receiving non-aggregated transaction data daily streamed from supermarket retailers, firstly to be used in the Household Budget Survey, but also planning for future implementation in the CPI. Statistics Norway is also in the process of moving data collection from physical servers to cloud databases and cloud computing platform.

² Traditionally data providers need to manually fill out web questionnaires.

³ The so-called *static approach* for processing supermarket scanner data (Eurostat, 2017)

⁴ In this paper product ID, article or item codes are treated as equal terms.

⁵ The product ID may also be a more generic code used for instance on fuels, take away services and others where there is a high degree of code stability.

methods like multilateral price indices presented internationally, do not automatically solve the problem related to product definition and replacements of product IDs. These issues must therefore be addressed separately.

During 2021 regular transmission of scanner data on consumer electronics was established, requiring new ways of dealing with high churn scanner data. Statistics Norway receives data from two leading retailers on consumer electronics covering most of the market⁶. As part of a Eurostat grant agreement, work has been done in order to explore the possibilities of using the full potential of scanner data on consumer electronics in the Norwegian CPI. This paper concentrates on presenting the challenges encountered such as incomplete metadata and combining data sources as well as some results of hedonic quality adjustment derived from hedonic regression models. The Norwegian CPI has no previous history of making use of hedonics for quality adjustment, an important part of this study has therefore been to gain experience. Questions we have looked more closely into have been: Are hedonic quality adjustments feasible given our data availability and resources? And, are there cost-effective solutions to be used that exploit the full potential of scanner data? The aim of this study has been to look for practical solutions for implementing scanner data on consumer electronics in official price indices.

The paper is structured in the following way: Section 2 of this paper provides some information of the Norwegian consumer electronics market while section 3 describes the specific challenges this type of data provides. Section 4 gives an overview of the available data, how data from different sources are combined, and challenges related to incomplete information. Section 5 compares different price index methods and presents some empirical results, while in section 6 some concluding remarks are given. Information on the selected product groups, methods applied, and results gained are found in annexes.

2. The consumer electronics market

The Norwegian market of consumer electronics is dominated by some retailers with large market shares. The two main retailers providing scanner data, alone stand for at least 2/3 of the total Norwegian market. In light of the pandemic, the industry has seen changes in shopping habits and

⁶ One of the retailers has already provided data for the online store for many years.

new needs for technology during the last two years, both due to working from home and new digital leisure activities. This has led to an increased consumer demand and increase in turnover for certain product categories. Working from home has in particular increased the turnover of web cameras, computer screens, keyboards and other typical offices equipment. In addition, we see that sales of gaming accessories have increased strongly. More time spent at home both during working hours and leisure time has also led to increased expenditure on kitchen appliances. Another effect reinforced by the pandemic is increased turnover share for online purchases during 2020 and 2021. Empirically, we see that about 15 per cent of total turnover is currently from online sales. There is however diversity among the different product categories where gaming accessories, videogame software and lighting have higher shares while for instance loud speakers, soundbars and hi-fi systems have lower shares. In recent years retailers selling consumer electronics offer various “click and collect” solutions where the consumers order the product online and pick it up themselves in the physical store. These purchases are likely to be registered as either online or offline purchases depending on where the actual payment transaction takes place.

3. Dealing with high churn data

3.1 Data properties

The item codes are the most detailed level of homogeneity in the scanner data. For price measurement it is very important to handle a dynamic item universe where new and old items continuously enter and exit the market. Ideally, we want methods that include both completely new items at the time of introduction, as well as re-launches or item replacements, i.e. new item codes, but essentially similar items. Defining the item at GTIN- or other item code level may in many cases be too detailed, especially for items of high churn, such as consumer electronics where the life span of item codes is short. In traditional methodology, items that are no longer available, are identified and replaced by comparable items with the same quality characteristics. If a comparable replacement is not available, a non-comparable replacement is chosen with a (suitable) quality adjustment method applied. In general, the traditional methodology, where items leaving the market are replaced by new ones of similar quality one-to-one, is difficult to apply aiming at maximizing the potential of scanner data and using the data “full-scale” as scanner data instead implies “many-to-many” replacements and a variety of possible matches.

Our findings in this study show that item codes for consumer electronics stay on the market for less than six consecutive months on average⁷. Items within some product categories like for instance personal care appliances, tend to be more long-lived compared to others such as for instance computers. Bestselling items appear to stay longer on the market, empirically we see that item codes with the largest quantity sold stay on the market for more than nine months on average. We also find that there is some distinction for items sold online and items sold in physical stores where larger persistence seems to prevail in physical stores.

3.2 Homogenous products (HPs)

Different methods have been suggested for dealing with high churn data. One approach is to apply a broader item definition, by combining different item (codes) of similar attributes/quality in order to create *homogenous products* (HPs). By calculating a unit value across comparable item codes, we allow for comparisons of new codes entering and old ones disappearing from the market. In 2018 Statistics Norway implemented the use of scanner data from one of the largest sport clothing and equipment retailers on the market in the CPI (Hov, Johannessen, 2018). This was the first time scanner data of high churn items was included in the index and the chosen solution was to establish and measure unit value prices of HPs over time. Since then, matching and measuring HPs over time, has also partly been implemented for general clothing.

Consumer electronics however poses additional challenges. Like for clothing, the item codes are too detailed for price index purposes, but unlike clothing there are often a larger range of price-determining characteristics that make the use of HPs more complicated. The consumer electronics market is characterized by a high degree of innovation and new launches. Several product categories face rapid rate of technological progress and new items are often of higher quality than the outgoing items. While clothing is characterized by price-determining factors that are relatively stable over time, such as brand and fabric and to a certain degree even design, this may not be the case with consumer electronics.

At the same time, consumer electronics consists of a large range of product categories. There seems to be a large diversity in the degree of gradual advancement and the number of price-determining characteristics are different across the various categories. There is great difference

⁷ Our data shows that item codes that are not sold for a couple of months seldom reappear on the market.

in the number of characteristics describing an electric toothbrush or a hairdryer compared to a computer or a smart phone. All these product categories are admittedly based on advanced technology, but the number and the stability of specifications or technology may differ.

3.3 Replacement and quality adjustment

Another approach to deal with high churn data is to rely on hedonic modelling techniques. Hedonic techniques control for quality differences across items and over time in order to build price indices of constant quality. The hedonic approach estimates a price as a function of the characteristics of the item. As a quality adjustment measurement, hedonic methods adjust the measured price changes for changes in the quality, i.e. the specifications of the items – it can, through modelling and prediction of prices, solve the issue of item replacements by systematically measuring and correcting for quality difference between the outgoing and incoming items. Using such methods estimates the importance of characteristics of the items and allows us to include unmatched items over time.

When faced with measuring prices for product categories that undergo rapid quality change, international best practice is to develop hedonic price indices. The overall aim in using consumer electronics scanner data is to calculate price indices where we try to control for quality changes. It can be used to estimate prices for the period prior to appearance on the market and prices for the period after disappearing. Such methods however, require comprehensive information about price-determining characteristics. Scanner data itself provides limited metadata, therefore alternative data sources have been exploited.

4. Data collection

4.1 Scanner data

Since the beginning of 2021 Statistics Norway has received weekly scanner data from the two leading retailers on consumer electronics. The weekly⁸ scanner data sets contain price and turnover information on all items sold, both from physical and online stores, in addition to some specifications of each item sold. The specifications are however limited, specifying product group

⁸ The three first full weeks of the month are used in the price indices.

and a short item description, and the level of detail at which the items are grouped by the retailers varies largely. The scanner data content and how it is presented rely on the retailer's own cash register systems, purchasing routines and logistics system in-store, amongst other things. Also, the cash register system, important for the price and quantity information, tends to be separated from those for web site content which is important for item specifications⁹. The data received from the two consumer electronics retailers covers the entire range of electric products directed towards consumers.

4.2 Web scraped metadata

In this study we have applied web scraping techniques to enrich the data with metadata from the retailer's own web site. Unlike scraping price data, scraping only the item specification is less sensitive to what time and day the data is scraped, and there is less need for frequent scraping. In addition, it's also sufficient to only scrape metadata for new items, thus we do not have to scrape the entire web page each time and therefore cause fewer constraints on the web site itself.

The automated web scraper for consumer electronics has been set up for one of the retailers, with a weekly frequency of scraping metadata for new items, for a chosen selection of product groups relevant for our study. The data is stored in a data set that is continuously updated with information on new items. The information scraped is item IDs in addition to all available information associated with the item¹⁰. The web scraper is set up in a Python programming language environment, using the web automated tool Selenium to navigate the web site with a remote driver to collect the specified item information. The web scraper has worked satisfactory for longer time periods, but not without some troubles regarding changing web site set up. These issues need to be immediately addressed to make sure no data is lost on the way. The scraper is set up and maintained by the CPI staff itself.

4.3 Combining data

The metadata and the scanner data are combined using unique item IDs. For the retailer that is also the owner of the scraped web site, the retailer specific article code is used to combine the

⁹ A desired solution is to have the retailers making a direct file delivery, but that has not been achieved.

¹⁰ The internal retailer specific article code is used as item ID for the retailer we scrape, while EAN (European article number) is used for the retailer we do not scrape.

data. To be able to include and combine data from the retailer not scraped, the EAN, scraped explicitly for that purpose, is used. The assumption was that setting up a scraper for one of the retailers would provide metadata also for the second retailer, however this has proved only partly correct; for some of the product groups we find merely a 30 per cent match of unique items sold by both retailers, mainly reflecting that the two retailers advertise and carry different brands and items. For other product groups such as smart phones we find a higher match, closer to 80 per cent between the two retailers. For the retailer from which we scrape information we find that some items are exclusively sold in-store, thus no metadata available online. This seems however to be a minor problem so far.

In an effort to increase the scanner data and metadata match, an approximate string-matching procedure *fuzzy matching* on item text has been carried out for seemingly similar items¹¹. The idea was to exploit and adapt metadata information for similar items for items without a match. The fuzzy matching procedure has proved only partly successful. For an optimal result the item text should be written on the same form for all items, this is however not the case. In the examples shown below, we compare a Samsung Galaxy tab, written as “Galaxy tab...”, with three other Samsung Galaxy tab items. The item which is the most similar is however written as “*Samsung* Galaxy tab ...”. The penalty of having “Samsung” in the item description for the item that otherwise would be the most similar, was severe, and the item ranks lower than the two other less similar items.

Table 1: Fuzzy matching example

Item to match	Item suggested	Compged score	Rank
GALAXY TAB A 10.1 2019 4G (32GB) BLACK	GALAXY TAB S7 128 GB 4G MYSTIC BLACK	1260	1
GALAXY TAB A 10.1 2019 4G (32GB) BLACK	GALAXY TAB S7+ 128 GB 5G MYSTIC BLACK	1370	2
GALAXY TAB A 10.1 2019 4G (32GB) BLACK	SAMSUNG GALAXY TAB A 10,1" 2019 4G 32GB	1540	3

¹¹ A procedure that identifies elements of text strings that are similar and penalizes elements that are not similar. The procedure gives a score according to the sum of penalty values, also known as the COMPGED score.

The fuzzy matching procedure shows that data cleaning is essential, both for possible item matching across retailers, but as explained below also for the metadata itself.

4.4 Data cleaning and imputation

Web scraping techniques give a unique opportunity to collect metadata for items, but the information online is also rather overwhelming. For the consumer electronic product groups, we see that the metadata for most items contains 70+ variables/characteristics. The metadata is not only abundant, but also rather incomplete with missing information and inconsistencies, making the task of navigating the most suitable and comprehensive metadata for the items very time-consuming. Items that look seemingly similar might be provided with quite different metadata, both regarding what type of information is present, but also how it is spelled out. Examples shown below on three similar items, showcasing various information.

Table 2: Item specifications for computer tablets

	<i>Item 1</i>	<i>Item 2</i>	<i>Item 3</i>
<i>Operating system</i>	Android	Android 11	NA
<i>Number of cameras</i>	1	NA	1+3
<i>Storage capacity (GB)</i>	64GB	129	1TB
<i>Brand</i>	Samsung	Samsung	Apple

In general, a large share of the metadata is sparsely filled out and therefore not very useful in identifying price-determining variables. Approximately 10-20 variables per product group show coverage to some extent, however missing information is still a challenge. The variables that show the most coverage in item description are (naturally) also the variables that seem to be the most price-determining. On the other hand, some of the variables that we would expect metadata from are showing remarkably low degrees of coverage. An example being battery capacity on standby or battery capacity during phone calls for computer tablets. Less than three and six per cent of the computer tablets has information on battery capacity on standby or during phone calls respectively. Battery capacity in general has poor coverage. The reason might be that there is a total of six different variables all measuring various battery characteristics, thus not one covering them all.

In an effort to complete the information fuzzy matching has been tried also to impute missing information in the variables by matching and using information from similar items. This has proved difficult as the text strings are written in various forms.

Other ways to fill in the missing information is by searching for item information on other web sites such as price comparison sites and/or sites meant for item comparisons. The scanner data specifications itself, though limited, may also provide some information in filling in the missing information. Item specifications for consumer electronics are rather technical which makes it challenging to navigate and detect the precise information for imputation. An example being a seemingly simple question as the number of cameras on a mobile phone, and the number of pixels associated with the camera(s). As the technical features of a mobile phone has progressed, an increasing number of smart phones have more than one camera; at a minimum a front (selfie) camera and a back-end camera. In recent years also with multiple cameras in the back, and varying number of pixels and other technical specifications. In the metadata this can be expressed as 1+3, meaning one front camera and three in the back, or it can be expressed by only the most advanced camera specification. In the perspective of item matching and comparing items, the lack of standardization together with the vast amount of technological specification make data cleaning a vital, but also a severely time consuming, and rather manual task.

5. Price index methods and empirical results

5.1 Current index calculations

Measured by the CPI weights, consumer electronics in total makes out close to 5 per cent of the Norwegian CPI basket. The current price indices of consumer electronics are based on traditional matched-model methodology where representative items are defined, selected and matched over time. When an item disappears, it's replaced by another. Hedonic price indices have not been used in the Norwegian CPI. If need of quality adjustment (manual validation) imputed values based on class mean imputation are used for correcting the price in the price reference period. Thus, no explicit quality adjustment is carried out. Furthermore, eight elementary indices stratified by region are calculated by Jevons unweighted geometric formula, and then weighted into a national index of the representative item. A future practical calculation method should ideally be able to make use of the potential scanner data provides related to both scope and weight information (quantity sold/turnover).

5.2 Hedonic function specifications

Given the metadata available, hedonic methods have been tested. Thus, the first stage is to specify hedonic price functions relating the transaction prices to the relevant characteristics of the items. In order to find optimal fit our starting point is the Ordinary Least Square (OLS) methodology. The regression model which includes data over several periods can be expressed in the following way:

$$\ln p_i^t = a + \sum_{t=1}^T \delta^t D_i^t + \sum_{k=1}^K \beta_k z_{ik} + \varepsilon_i^t$$

Where $\ln p_i^t$ denotes the logarithmic price of the item i in period t , z_{ik} is the characteristics, β_k the corresponding parameter and ε_i^t the random error term where the expected average value is zero. Furthermore, a is the intercept, and the time dummy variable D_i^t takes the form as 1 for period t (across the window $0 \dots T$) or 0 otherwise, while the δ^t is the corresponding parameter. Many different functional forms have been tested, both double-logarithmic form of the regression model, i.e. that the model uses natural logarithms on both side of the function as well as categorization of the explanatory variables. We see that grouping of specifications into fine categories may be a way to reduce some noise in the metadata. There is a risk that we lose information, but comparisons made on categorization of continuous variables or not show similar index results.

The regression analyses are done mainly in SAS programming language and partly in R. SAS has several procedures for doing regression analysis, and we have been using mostly the *PROC GLM* (general linear models) procedure. In specifying the hedonic functions, we have aimed at feasible practical solutions and simplicity as well as high explanatory factors of the model and the parameters. The variable selection processes have been a mix of theory and practical knowledge. We have tested automatic model selection procedures¹² in addition to more manual combinations of different variables. In practice, we see that there is a very high number of variables in the web scraped metadata, but many of them are overlapping with other variables and/or incomplete with limited value for analysis. One important aspect has been to see whether it is possible to reduce the number of explanatory variables in the models without reducing too much of the explanatory power of the model in order to find the simplest specifications possible.

¹² The *PROC GLMSELECT* procedure in SAS provides different methods for automatic model selection.

We find that the weighting of the functions seems to impact the regression model results, and both expenditure share- and quantity share-weighted functions have been explored as well as no weighting. We also see that the effects of the weighting differ across product categories. This indicates that the items measured are quite heterogenous in price and/or in characteristics. It can even be an indication of different degree of completeness in the metadata across items. We see that quantity share-weighting seems to provide best results. It can be an advantage to standardize the choice of weights across product groups and going forward we will however most likely use expenditure share-weighted functions (a type of WLS) since this is more in line with general weighing procedures in price indices, and it also seems to be mostly used in other studies. Using expenditure-weighted hedonic models means that the items with highest expenditure are the ones that are emphasized in the models. We also assume that the metadata is the most complete for these items.

A high R-squared is important for the overall explanatory power of the regression models, but it's not a sufficient condition. Different tests of the model performances have been performed, related to multicollinearity, residual behavior, heteroscedasticity and possible outliers.

5.3 Comparisons and empirical results

In this project we have mainly concentrated on four different product groups: mobile phones, computers, computer tablets and flat screens. These product groups have properties that make them suitable for hedonic regression analysis, but they also differ in some respects which make them interesting and the results applicable for other similar product groups. The rate and size of technological improvement differ, however they are all high-tech product groups that are subject to rather frequent quality improvement. Different methods, both hedonic indices, HPs - and more standard formula price indices have been tested for these product groups.

The calculations made in this study show that different methods provide very different results, and some indices calculated serve more as references rather than candidates for actual implementation. We see that hedonic indices may be computed in many different ways. *"A hedonic price index is any price index that makes use of a hedonic function"* (Triplett, 2004). We have tested a weighted Time dummy hedonic method (TDH) and a Hedonic double imputation method (DI), both using the regression function differently. The data period is short, hence it's difficult to draw any valid conclusions from the results of these series. It is of more interest how the index series

relate to each other than what price development they show. We see that in order to set up indices for regular production, further work is needed.

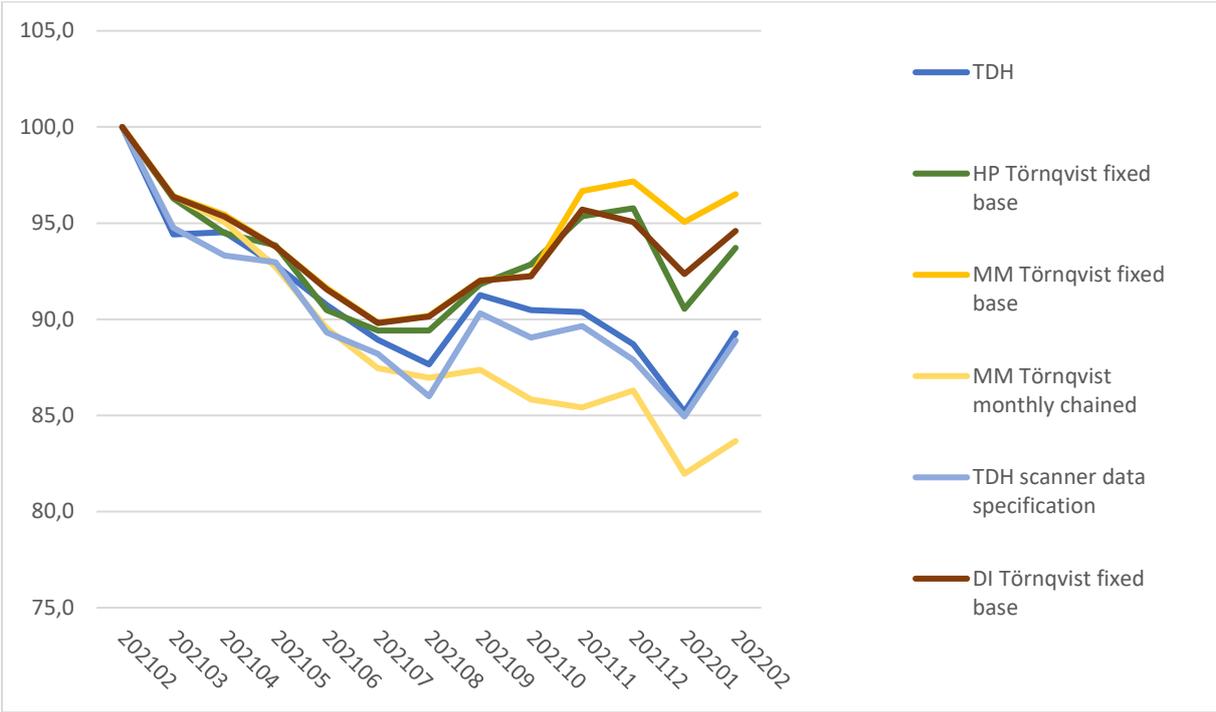
The figures 1-4 below illustrate comparisons of different indices of the four product groups that we have been focusing on in the study. We see from these figures that the various methods provide very different results.

Mobile phones

In figure 1 we compare the TDH method, the DI method with Törnqvist fixed base formation and matched-model (MM) indices for mobile phones. In addition, we include a HP Törnqvist price index using fixed base. We have included two versions of the TDH, one with full model specification (“TDH”) and another based on fewer specifications limited to information that we normally get from the scanner data set (“TDH scanner data specification”). We see that there are only minor differences between these series in the analyzed time period. All series seem to coincide quite well in the first half of the time period, but during the second half the TDH indices lie below the fixed base counterparts. Unlike the other indices, the TDH versions are using data from the entire window in the index calculations. There are smaller differences between the DI Törnqvist fixed base price index and the HP Törnqvist fixed base price index throughout the time period. The DI index is increasingly regression model-based and does not include possible matches throughout the time window. The HP version is based on classification variables that is information normally provided in the scanner data specification. Other HP versions show that defining the HPs on full model specification (including camera, display technology etc.) does not seem to affect the index development much. Also, somewhat unexpectedly the MM Törnqvist fixed base price index is quite similar to HP and DI fixed base versions. We see however that the MM Törnqvist fixed base price index soon becomes unrepresentative given that approximately 85 per cent of the matches have disappeared at the end of the time period. Furthermore, the MM monthly chained version shows signs of chain drift, as expected. Compared to official Norwegian CPI series of mobile telephone equipment (European COICOP 08.2.0.2)¹³ the price decrease is stronger in these experimental series.

¹³ Mobile phones make out more than 90 per cent of the weight.

Figure 1: Price indices of mobile phones, February 2021-February 2022

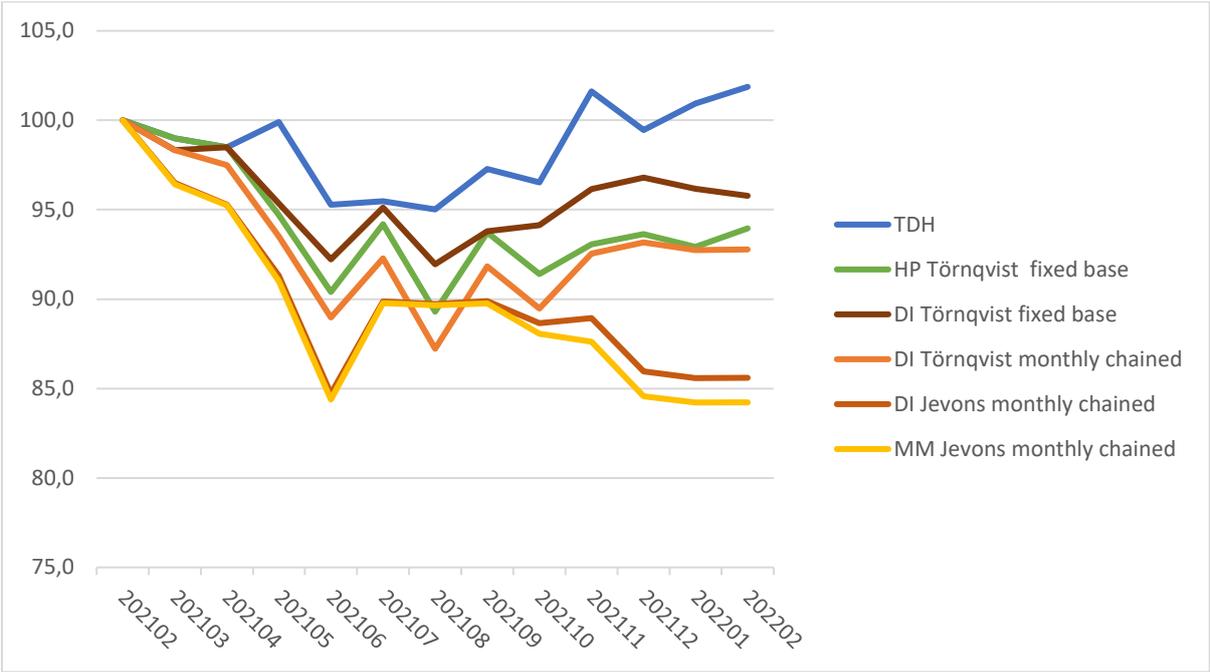


Computers

In figure 2 below we compare the price index series for laptop computers¹⁴ displaying the TDH method, the DI method using both period-on-period chaining and fixed base. In addition, a MM Jevons monthly chained index and a HP Törnqvist price index with a fixed base period is presented.

¹⁴ Both laptop and desktop computers were analyzed in this study, however due to the limited number of observations for desktop computers and the heterogeneity of laptop and desktop computers, only laptop computers are described in this paper.

Figure 2: Price indices of laptop computers, February 2021- February 2022



The TDH method which uses information from the whole time period in the hedonic estimation, lies above the other indices for most of the period analyzed. The MM monthly chained index however shows clear signs of downward drift and lies below the other indices throughout the period. This is as expected for product groups where the items are characterized by shorter life span and where the price tend to fall before the item exits the market. The DI Jevons monthly chained index lies close to the MM Jevons monthly chained index throughout the period, which indicates that the number of matched items between two consecutive periods is high, and therefore less quality adjustment has been applied and the two indices coincide. The DI Törnqvist monthly chained index however is less subject to downward drift, indicating that weighting is of importance. A monthly chained bilateral Törnqvist index is usually subject to chain drift caused by discount effects, however this does not seem to appear for laptop computers, maybe because consumers are more interested in the newer items of higher quality than possible discounts on older items.

While the monthly chained indices tend to entail downward drift, the fixed base indices might be too much influenced by the lack of item matching at the end of the period. The DI Törnqvist fixed base and the HP Törnqvist fixed base both lie above the monthly chained indices through most of the period, the drawback however for the fixed base indices is that we find a large degree of discontinuation of items throughout the period. A large share of the index for the latter part of the

13-month time period is based on estimated values; the item match at the end of the period is merely around 10 per cent. The HP version lies below the DI Törnqvist fixed base which can be a result of a too tight HP definition, thus not allowing new items to be included in the HP.

Flat screens

Figure 3: Price indices of flat screens, February 2021- February 2022

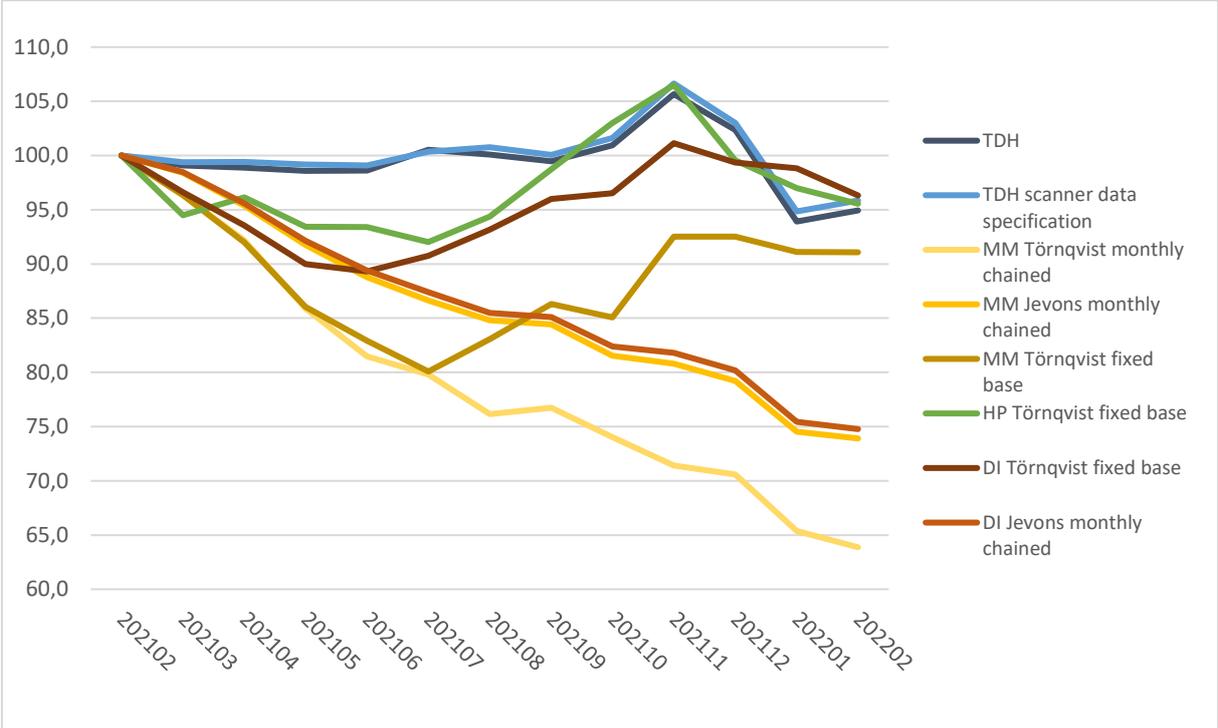


Figure 3 shows very different price movement across the various price indices for flat screens. The TDH indices lie above the other price indices in most of the period. We see that there are only marginal differences between the TDH index based on full specification and a TDH based on fewer specifications that can be extracted from scanner data descriptions. The MM monthly chained indices and especially the Törnqvist price index clearly suffer from chain drift. The fixed base version on the other hand is not very representative, given that the majority of the data disappears through the period. We find that after one year only about 10 per cent are still matching. The HP fixed base price index which incorporates many of the items over time, seems to work pretty well and shows quite similar development with the DI price index with fixed base. Both the HP and the DI price indices are more volatile compared to the TDH indices that uses data from the entire window in its estimation. A possible complication is that by using a single month as a base period we put too much emphasis on the items in this particular period. Comparing the HP and the DI

fixed base price indices with aggregated price indices of TVs in the official CPI show a similar development during the time period.

Computer tablets

In figure 4 below we find the indices calculated for computer tablets. Unlike the treatment for mobile phones, flat screens and computers, the indices for computer tablets are based solely on the metadata drawn from the scanner data item text. Two variants of model specifications have been tested for both the HP Törnqvist fixed base and the TDH method. While both variants are based on screen size, storage capacity, retailer and retailer type, there are two alternative versions of brand definition: Model 1 contains a definition of brand where the mother brand and computer tablet family are combined (e.g. Apple Ipad Pro, Samsung Galaxy tab S4 etc), and Model 2 where the brand variable contains only the mother brand such as Apple, Samsung, Huawei etc.

Figure 4: Price indices of computer tablets, February 2021- February 2022



As shown above, while the two TDH indices seem to trend downwards throughout the time period, both HP indices are showing upward trends at the second part of the time period. This can be a sign of new items being included in the HP throughout the period, but also that the indices are subject to unit value bias. The model 2 for HP Törnqvist fixed base in particular, where brand is

defined by the mother brand, seems to show unit value bias. This is as expected as it does not take into account the wide range of products within the mother brand, going from rather simple to advanced, thus the unit value can shift substantially upwards (down) if advanced (simple) models are introduced.

An interesting point is that the HP indices for both models are showing similar development throughout the period, though some larger deviations at the end of the period, the TDH indices for both models are showing larger deviations in both size and direction on several occasions. This indicates that the model specifications in the TDH have large impact on the quality adjustment and therefore also the development in the indices.

6. Concluding remarks

The aim of this paper is to build knowledge on the use of hedonic regressions in the Norwegian CPI and search for practical solutions for exploiting full potential of scanner data on consumer electronics. This Eurostat funded study has given us useful insights for compiling future indices, and as expected, calculating hedonic price indices is not a quick fix. We see that we need longer time series and more complete metadata in order to implement hedonic price indices for consumer electronics in the CPI. Furthermore, highly technical product categories are often associated with life cycle effects where the price tends to fall before the item exits the market, and care must be taken even for explicit quality adjustment methods like hedonic modelling. Up until now no systematic strategy to deal with possible life cycle effects, e.g. removing discount prices from the regression estimates, has been taken. Dumped prices have however been removed. This is something we will analyze further with more data.

The calculations made in this study have showed that different methods seem to provide very different results and some of the indices calculated are not candidates for actual implementation. We believe that hedonic price indices clearly have the potential to improve existing practices of calculating indices for consumer electronics, but we also recognize that hedonic indices have their flaws; if the regression models are not specified correctly they can provide skewed result. Hedonic modelling also puts strong assumptions on consumers' valuation of the different characteristics of the items, however this study has not pursued that issue any further.

Using regression analysis in the official indices is likely to increase the index accuracy, but the costs of index compilation are high with large burden on the CPI staff. While for instance clothing items to a large extent may rely on the same regression model across different product categories, consumer electronics is not a homogenous group and will require separate regression models including frequent updates and assessments, to be defined for all product categories. We are facing a trade-off with quality gains on one hand and production costs on the other. One possible strategy is to combine different calculation methods, given that consumer electronics differs in the degree of technical progress. Using HP based indices for some product groups may provide acceptable results, the same for regression models based on limited number of explanatory variables. The relative importance of the product groups in terms of weights in the CPI, combined with a technological progress validation may be a way of diversifying across product groups. Our testing shows that we seem to come a long way for the analyzed product groups making use of information that is incorporated in the item description of the scanner data. Extracting information directly from the item description however requires extensive routines for data cleaning and adjustments as the item text is not standardized in any way. Nevertheless, the item texts can help us identify what metadata/variables we, at least as a minimum, need to web scrape and utilize. Another solution that will make this process much easier is naturally to receive metadata directly from the retailers themselves, but attempts so far have been unsuccessful. We see that a reduced number of price-determining variables may still provide satisfactory results for different product categories.

No conclusions are yet drawn in this study. The aim is to implement practical solutions for quality adjusted price indices on consumer electronics using scanner data, hence the work will continue. Relevant price index methods, like multilateral variants, have not yet been tested due to the short period of data.

References

- de Haan, J., Daalmans, J. (2019), "Scanner data in the CPI: The Imputation CCDI Index Revisited", <https://www.business.unsw.edu.au/research-site/centreforappliedeconomicresearch-site/Documents/emg2018-Imputation-CCDI-Index-Revisited-24-march-2019.pdf>
- de Haan, J., Diewert, E.W. (2017), "Quality Change, Hedonic Regression and Price Index Construction", <https://www.bundesbank.de/resource/blob/635108/8c57b268050f67992fe7cc2a0388024c/mL/2017-05-10-ottawa-group-01-2-paper-data.pdf>
- Eurostat (2017), "Practical Guide for Processing Supermarket Scanner Data", <https://circabc.europa.eu/sd/a/8e1333df-ca16-40fc-bc6a-1ce1be37247c/Practical-Guide-Supermarket-Scanner-Data-September-2017.pdf>
- Eurostat (2022), Report of the work related to Task Force Quality Improvement (TFQI) subgroup on mobile telephone equipment product descriptions
- GSMArena, webpage of mobile phone specifications, <https://gsmarena.com>
- Hov, K.N., Johannessen, R. (2018), "Using scanner data for sports equipment", https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Norway_session_1.pdf
- IMF et al. (2020), "Consumer Price Index Manual: Concepts and Methods", <https://www.imf.org/en/Data/Statistics/cpi-manual?msclkid=c5c3b0aca82211ec8ac1fb0e4654bab6>
- Johansen, I. (2013), "Bruk av strekkodedata i matvareindeksen" (only in Norwegian), https://www.ssb.no/a/publikasjoner/pdf/oa_201206/johansen.pdf
- Johansen, I., Nygaard, R. (2021), "Ny beregningsmetode for prisindeks for matvarer og alkoholfrie drikkevarer i KPI" (only in Norwegian), <https://www.ssb.no/priser-og-prisindekser/artikler-og-publikasjoner/attachment/445212?ts=17780d4a6e0>
- Listianingrum, T., Nefriana, R. (2020), "Utilizing a Price Comparison Website to Produce Hedonic Price Indices for Mobile Phone", https://www.unescap.org/sites/default/files/APS2020/82_Utilizing_A_Price_Comparison_Website_to_Produce_Hedonic_Price_Indicesfor_Mobile_Phone_Indonesia.pdf
- SAS Documentation, 9.4, The GLM Procedure, https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.3/statug/statug_glm_syntax01.htm
- Trewin, D. (2005), "The Introduction of Hedonic Price Indexes for Personal Computers", [https://www.ausstats.abs.gov.au/ausstats/free.nsf/0/6B5EF2E6C1935D18CA256FA1007856A2/\\$File/64580_2005.pdf](https://www.ausstats.abs.gov.au/ausstats/free.nsf/0/6B5EF2E6C1935D18CA256FA1007856A2/$File/64580_2005.pdf)
- Triplett, J. (2004), "Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special application to information technology products". STI working papers.

Appendix I

In appendix I we provide some information on the different product groups analyzed in this study. A brief overview of the data sets and explanatory variables are included.

Mobile phones

The data set of the entire time window (February 2021 - February 2022) on mobile phones with available metadata consists of approximately 10 500 price observations aggregated to type of retailer, sales channel (online/offline), article code and time period. The data set makes out 82 per cent of the total number of scanner data observations, i.e. 82 per cent of the data set has information on the characteristics of the items. Metadata is however not complete for all explanatory variables, after data cleaning and imputation of missing values roughly 9 200 price observations are used in the calculations. The hedonic function is based on nine different explanatory variables excluding the time dummy variable. The retailers' own classification into the categories "high- mid- low- end phone" has most explanatory power. Together with brand, the most important quality variables of the items are storage and memory capacity measured in GB (gigabyte) and RAM. Phone storage in GB is used to store data such as apps, videos etc. necessary for the mobile phone to run while RAM refers to Random Access Memory and is part of the phone that is used to store the operating system and where apps and data currently in use are kept. The following explanatory variables are defined in the model: retailer, sales channel (online/offline), the retailers' own group classification, brand, storage capacity in GB, memory in RAM, screen size, number of cameras, screen/display technology like AMOLED, OLED etc. The brand variable categorizes over "low-end" feature phones. Different groupings or categories of variables, in order to create robustness in the estimates, seem to fit the data best, for instance grouping within number of cameras, screen size etc. Outliers have been removed by comparing log price to their predicted values. The hedonic regression has a high R-squared value of almost 0.95 per cent meaning that the model explains close to 95 per cent of the diversity in the prices.

Computers

The data set for laptop computers consists of close to 12 000 price observations, aggregated by retailer, online/physical store, article code and time period. The metadata coverage related to the selected explanatory variables corresponds to only about 50 per cent of the entire data material, i.e. close to 5 600 observations, over the 13-month period. The low coverage can be explained by the diversity in products being sold by the two retailers and that we only collect metadata from the one retailer.

The hedonic model function for laptop computers consists of characteristics related to the retailer such as sales channel (online/offline) and retailer product grouping, and the quality characteristics brand, screen size, internal memory in RAM, storage capacity in GB, number of processor cores and net weight in kilograms, in addition to the time dummy variable. All explanatory variables are treated as categorical; those variables originally containing continuous values have been transformed and grouped into categories to reduce noise and better capture the various levels. The explanatory variables have been chosen based on various tests of robustness, p-values and explanatory power to the model. The overall explanatory power of the hedonic model is 0.94 measured by the R-squared.

Flat screens

The data set for TVs is based on over 6 100 price observations from the two retailers over a 13-month period. Eight different quality-variables including characteristics related to retailer are defined in the regression model. The most important quality specification of flat screens is the screen/display technology. Besides the time dummy variable, the model is explained by the following variables; retailer, sales channel (online/offline), brand, resolution/picture quality, screen size interval (inch), screen technology, smart TV (yes/no) and net weight. The characteristics are treated as categorical variables. For instance, screen size defined in narrow size intervals (in inches) performs better than the screen size of each item. The size intervals are part of the retailer's own classification. The hedonic regression has a R-squared value of about 0.87.

Computer tablets

The data set of computer tablets is rather small; only about 3 000 price observations during the 13-month period for the two retailers. This corresponds to about 500 unique items. Given the relatively small data set, an effort was made to test whether the information found in the price data (scanner data) itself was sufficient for creating a hedonic model. The price-determining item characteristics used in the hedonic model is drawn from the item text itself, covering the variables brand, screen size in inches, storage capacity in gigabyte, in addition to retailer including sales channel (online/offline). Screen size and storage capacity are log-transformed, while the others are treated as dummy variables. About 2 000 price observations had full coverage of item characteristics after the manual metadata extraction from the item text, which corresponds to about 60 per cent of the total data set. The overall explanatory power of the model turned out to be as high as 0.97 measured by the R-squared.

Appendix II

Price index methods

Appendix II provides an overview of the methods that we have used for the empirical testing.

Weighted time dummy hedonic method

Having scanner data that only covers a 13-month period makes it difficult to do meaningful testing of multilateral methods like for instance the imputation CCDI index (de Haan, Daalmans, 2019). That method would be a natural candidate in the Norwegian CPI given that the index of food and non-alcoholic beverages and others are calculated by the CCDI index. The only multilateral method that is tested in this study is the Time dummy hedonic method (TDH), a weighted version. The TDH method uses the hedonic regression directly in the price index by using a time dummy as one predictor in the model. Data is pooled together and used over the entire time period, the coefficient for each period is the index number by exponentiating the parameter estimate. A direct TDH method is keeping the characteristics fixed over the entire time window which may be a disadvantage when there are changes both in consumer preferences and changes in characteristics. The main advantage of the TDH method is its simplicity, but we do not regard the direct method as a possible solution for regular index production. When new periods are added earlier indices will be revised which is a solution that cannot be used in actual index production. This can be dealt with by using a rolling window of a certain length spliced together over time. This will be tested as soon as longer time series allow this.

Hedonic double imputation method

The hedonic double imputation method (DI) can be regarded as a “indirect” method as the regression function is used to predict prices and incorporated into standard price index formulas. The method is a combination of the matched-model (MM) approach and the TDH method:

$$I_{DI}^{t_0,t} = [I_{MM}^{t_0,t}]^{S_{MM}} [I_{TDH}^{t_0,t}]^{1-S_{MM}}$$

The DI method uses a MM to determine a price index over two periods for items that are matched and present in both periods, and a second price index that uses hedonic regression to implicitly measure price movements for new and disappearing items in the same periods (TDH). In this study both the Törnqvist and Jevons formula have been calculated for the MM index. Both indices (MM and TDH) are then combined to create a hedonic quality adjusted price index. The expenditure share of the MM index s is calculated as an average of the 2 periods expenditure shares:

$$S_{MM} = \frac{(S^0 + S^1)}{2}$$

The double imputation indicates that predicted prices are used in both periods, even if prices are available in one of the two periods. The benefit of using double imputation (in contrast to the single imputation strategy, i.e. making use of the actual prices collected) is that the predicted prices may be biased thus, any possible omitted variable bias of the price relative is cancelled out. The method has the advantage that the model is re-estimated every period and therefore the coefficients are not fixed over time. In this study, both period-on-period chaining and fixed base formation are tested. The fixed base version means that the index is using a regression function based on only the data of the base period and the period t .

Price indices based on HPs

For certain consumer electronics the technological progress is slower and more stable over time. For those product groups calculating unit values i.e. HPs, over similar items (but with different article codes) of the same limited number of price-determining variables is an option. Compared to regression models calculating HPs will most likely be a less resource demanding solution. New replacement items are directly captured through the HPs by calculating unit values across the item codes in the different stratum, while completely new products or new HPs will not (unless some imputation technique is used). As we have experienced from other commodity groups, when calculating HPs we are faced with a trade-off between heterogeneity in items on the one hand and matches of items over time on the other. If the HPs are too broadly defined (i.e. based on too few price-determining variables), unit value bias will be a problem. An important complicating factor is that consumer electronics is not a homogenous group and consists of a multitude of different product categories that will require separate HP specifications (the same goes for the regression functions).

Matched-model indices (MM)

The standard method of constructing price indices is to use a MM approach. We have calculated both fixed base and period-on-period (monthly chained) price indices. The fixed base price index is directly calculated by comparing price of a specific item i in period t to the same item in a fixed base period, while the period-on-period MM indices select matched items in two consecutive periods (t_{t-1} and t_t). Using fixed base MM approach for product groups that undergo rapid technological change and high item churn, such as mobile phone and computers, is problematic. The sample of the matched items will deteriorate over a short period of time making the index unrepresentative. Monthly chained or period-on-period indices may on the other hand, depending

on index formula, suffer from chain drift. For the sake of comparability, we have also calculated these indices in this study.

Appendix III

Annex III provides a short overview of some of the estimation results of the hedonic models we have used. The results are based on the TDH models.

Mobile phones						
Dependent variable	log(price)					
R-Square	0.95					
Number of observations read	10 514					
Number of observations used	9 182					
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
Retailer	1	802.64	802.64	568.67	<.0001	
Sales channel	1	298.44	298.44	211.44	<.0001	
Period	12	376.43	31.37	22.22	<.0001	
"Quality indicator"/Retailers' own classification	2	3068.98	1534.49	1087.19	<.0001	
Brand	10	13193.82	1319.38	934.79	<.0001	
Screen size	2	577.74	288.87	204.66	<.0001	
Storage capacity (GB)	4	2706.39	676.60	479.37	<.0001	
Internal memory (RAM)	5	3629.52	725.90	514.31	<.0001	
Display technology	1	322.51	322.51	228.5	<.0001	
Camera	2	1088.97	544.49	385.77	<.0001	
Computers (laptop)						
Dependent variable	log(price)					
R-Square	0.94					
Number of observations read	5 595					
Number of observations used	4 587					
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
Retailer incl. sales channel	2	8840.52	4420.26	1.38	0.2507	
Period	12	195193.17	16266.10	5.09	<.0001	
"Quality indicator"/Retailers' own classification	12	2024075.42	168672.95	52.80	<.0001	
Brand	11	4977105.48	452464.13	141.64	<.0001	
Screen size	3	588944.83	196314.94	61.46	<.0001	
Storage capacity (GB)	4	3078805.64	769701.41	240.95	<.0001	
Internal memory (RAM)	4	4757384.96	1189346.24	372.32	<.0001	
Processor cores	6	1054961.88	175826.98	55.04	<.0001	
Net weight (KG)	4	237310.60	59327.65	18.57	<.0001	

Flat screens						
Dependent variable	log(price)					
R-Square	0.87					
Number of observations read	6 168					
Number of observations used	6 166					
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
Retailer	1	944.59	944.59	814.84	<.0001	
Sales channel	1	76.38	76.38	65.89	<.0001	
Period	12	146.68	12.22	10.54	<.0001	
Screen size	9	2884.34	320.48	276.46	<.0001	
Brand	10	497.67	49.77	42.93	<.0001	
Screen technology	7	6372.79	910.40	785.34	<.0001	
SmartTV	1	25.81	25.81	22.26	<.0001	
Resolution	3	206.60	68.87	59.41	<.0001	
Net weight (KG)	3	986.96	328.99	283.80	<.0001	
Computer tablets (Model 1)						
Dependent variable	log(price)					
R-Square	0.97					
Number of observations read	4 254					
Number of observations used	2 623					
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
Retailer incl. sales channel	3	38.42	12.81	16.92	<.0001	
Periods	12	116.26	9.69	12.80	<.0001	
Brand incl. tablet family	44	25227.22	573.35	757.35	<.0001	
Screen size	1	12.09	12.09	15.97	<.0001	
Storage capacity (GB)	1	1750.85	1750.85	2312.74	<.0001	
Computer tablets (Model 2)						
Dependent variable	log(price)					
R-Square	0.86					
Number of observations read	4 254					
Number of observations used	2 623					
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
Retailer incl. sales channel	3	70.91	23.64	6.53	0.0002	
Period	12	312.66	26.06	7.20	<.0001	
Brand (main)	7	17764.58	2537.80	701.30	<.0001	
Screen size	1	90.95	90.95	25.13	<.0001	
Storage capacity (GB)	1	11580.10	11580.10	3200.06	<.0001	