Nordic Council
of Ministers

NORDCRAWL 2

Nordcrawl 2

Kasper Mogensen and Troels Fjordbak

Co-author Lovisa Blomqvist

Nordcrawl 2

Kasper Mogensen and Troels Fjordbak

Co-author Lovisa Blomqvist

ISBN 978-92-893-5974-0 (PRINT)

ISBN 978-92-893-5975-7 (PDF)

ISBN 978-92-893-5976-4 (EPUB)

<http://dx.doi.org/10.6027/TN2019-503>

TemaNord 2019:503

ISSN 0908-6692

Standard: PDF/UA-1

ISO 14289-1

© Nordic Council of Ministers 2019

Cover photo: Unsplash.com

Disclaimer

This publication was funded by the Nordic Council of Ministers. However, the content does not necessarily reflect the Nordic Council of Ministers' views, opinions, attitudes or recommendations.

Rights and permissions

This work is made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0)
<https://creativecommons.org/licenses/by/4.0>

Translations: If you translate this work, please include the following disclaimer: *This translation was not produced by the Nordic Council of Ministers and should not be construed as official. The Nordic Council of Ministers cannot be held responsible for the translation or any errors in it.*

Adaptations: If you adapt this work, please include the following disclaimer along with the attribution: *This is an adaptation of an original work by the Nordic Council of Ministers. Responsibility for the views and opinions expressed in the adaptation rests solely with its author(s). The views and opinions in this adaptation have not been approved by the Nordic Council of Ministers.*

Third-party content: The Nordic Council of Ministers does not necessarily own every single part of this work. The Nordic Council of Ministers cannot, therefore, guarantee that the reuse of third-party content does not infringe the copyright of the third party. If you wish to reuse any third-party content, you bear the risks associated with any such rights violations. You are responsible for determining whether there is a need to obtain permission for the use of third-party content, and if so, for obtaining the relevant permission from the copyright holder. Examples of third-party content may include, but are not limited to, tables, figures or images.

Photo rights (further permission required for reuse):

Any queries regarding rights and licences should be addressed to:

Nordic Council of Ministers/Publication Unit
Ved Stranden 18
DK-1061 Copenhagen K
Denmark
Phone +45 3396 0200
pub@norden.org

Nordic co-operation

Nordic co-operation is one of the world's most extensive forms of regional collaboration, involving Denmark, Finland, Iceland, Norway, Sweden, and the Faroe Islands, Greenland and Åland.

Nordic co-operation has firm traditions in politics, economics and culture and plays an important role in European and international forums. The Nordic community strives for a strong Nordic Region in a strong Europe.

Nordic co-operation promotes regional interests and values in a global world. The values shared by the Nordic countries help make the region one of the most innovative and competitive in the world.

The Nordic Council of Ministers

Nordens Hus
Ved Stranden 18
DK-1061 Copenhagen K, Denmark
Tel.: +45 3396 0200
www.norden.org

Download Nordic publications at www.norden.org/nordpub

Contents

Foreword	7
Executive summary	9
1. Intro	11
1.1 Content of report	11
1.2 Glossary	12
1.3 Other improvements	12
2. NordCrawl for market surveillance	13
2.1 Market surveillance analysis – missing labels	13
2.2 Market surveillance analysis – banned energy classes due to ecodesign limits	20
3. Policy Evaluation – market share calculation	25
3.1 Market share calculation	25
3.2 Market share approximation methods	27
Conclusions	35
Recommendations and ideas for future work	37
Sammanfattning	39

Foreword

The project presented in this report has been performed on behalf of the Swedish Energy Agency within the Nordic cooperation project Nordsyn sponsored by the Nordic Council of Ministers. The project was performed by Kasper Mogensen and Troels Fjordbak at Big2Great Aps, in close cooperation with Nordsyn project leader Lovisa Blomqvist at Swedish Energy Agency.

The aim of the NordCrawl2 project was to use and further develop the market surveillance and policy capability of the earlier developed Nordcrawl tool.

Within this project the Nordcrawl-tool was developed for market surveillance – to find missing labels and products not meeting ecodesign limits – and for policy – where a method to estimate market share was developed.

25 April 2019

Lovisa Blomqvist, Swedish Energy Agency

Executive summary

The NordCrawl2 project was a project under the Nordic Nordsyn cooperation, sponsored by the Nordic Council of Ministers. It was performed by consultants at Big2Great. The aim of the Nordcrawl2 project was to show how the Nordcrawl-tool could be used for market surveillance and for policy work in the Nordic countries. The project consists of two parts focusing on market surveillance and policy, respectively. The first part describes how to use the NordCrawl system to do market surveillance with two examples, identifying webpage that are missing energy labels and identifying products with an energy class below to minimum requirement. The result of the missing energy label analysis is presented in a table with percentages for missing labels for each country, including the retailers with highest and lowest missing percentage. Quite many pages with missing labels were found, and this can be a good input to the market surveillance authorities. The analysis results on products with an energy class below the requirement is present in a table showing the non-compliance rate for each product type for each country. Here not so many faulty products were found, but still give some interesting indications for the market surveillance authorities.

In the second policy part of the project, methods to estimate market share is described, tested and evaluated. The most promising method is the sales/popularity rank method or in the cases where some cases a hybrid of that method. The results show that when this method is used, the estimated energy class distribution is close the actual distribution. This method can be helpful to find out the market share of non-compliant products or the market share of A+++ products, which is good input to policy discussions.

Finally, recommendations and ideas for future NordCrawl activities are described.

1. Intro

The initial NordCrawl project was a project sponsored by the Nordic Council of Ministers, running from August 2015 to June 2017. The aim of the project was to develop a system to support market surveillance authorities (MSAs) doing online market surveillance and policy evaluation. The NordCrawl system is an analysis platform, that takes product data from web crawling. The data is provided by a web crawling provider. The system can clean and normalize the data, perform quality assurance and store the data. Furthermore, a long range of analysis and data browsing views are built into the system. For a more detailed description read the NordCrawl/Maple report.¹ The aim of the NordCrawl2 project was to use and further develop the market surveillance and policy capability of the system, for instance by adding a method to identify retailer pages that does not show the required energy label. Another extension of the system is the development of a market share estimation method to support policy evaluation. Finally, this report shows results from non-compliance analysis for missing energy labels and appliances below the Ecodesign minimum-performance standard. The NordCrawl2 project also included some other updates (like website speed improvements) that will only be described briefly in this report.

1.1 Content of report

This report is a mix of method descriptions, analysis results and “how to use the system” guides. The report consists of two parts. The first part describes market surveillance approaches to identify retailer webpages with missing energy labels and products that does not comply with the ecodesign regulations demands. The second policy part describes and evaluate four approaches to calculate the market share of a model.

Retailer names, brand names, product names, and other sensitive information is anonymized, blurred or removed. The relevant energy authorities have the real names.

¹ Delivering Successful Product Policy Outcomes by Harnessing Realtime Market Data, Energimyndigheten Sweden (Peter Bennich), Big2Great Denmark (Kasper Mogensen), 2018.

1.2 Glossary

- *Web crawling*: The process in which a software program on a server download a website, store the website and extract information from that website and in some cases continues to visit link on the downloaded page;
- *Energy label*: An image of the energy label in the format required by the energy label framework regulation 2010/30/EU and its product specific regulations;
- *Energy class*: The energy class of the product on the A+++ – G scale;
- *Product page*: A retailer web page presenting information about the product like price, size, energy use etc.;
- *Market share*: A decimal number or a percentage. The number tells how much of the market a model or group of models represent;
- *Ecodesign*: EU directive 2009/125/EC and its product specific regulations that, among other things, set minimum performance standards;
- *Product type*: Group of products like washing machines, dishwashers etc.

1.3 Other improvements

A number of technical improvements on the NordCrawl system has been made during the project. The most important improvement is moving the database to MySQL 8 and increase the database cache size to hold a copy of all data in memory (RAM). This have improved the data selection speed significantly (in some cases 10,000X). Another speed improvement that have been developed, is caching of standard analysis results. Every night a script that recalculate all standard analysis results and save them run. When a user now wants to view a standard analysis, the systems does not have to calculate the result, it just load the saved cached result. This is also a significant speed improvement.

2. NordCrawl for market surveillance

2.1 Market surveillance analysis – missing labels

2.1.1 *Purpose*

The energy labeling framework regulation 2010/30/EU and its product specific regulations demand that the energy label is showed on the website for all covered products. The purpose of this NordCrawl task is to look for the energy label image and report if it is found or not. A new tool in the NordCrawl system was developed to provide an overview for the latest week. Other weeks are possible, but not relevant for an inspection point of view. However, the information is stored because it can be used to learn about how fast the market adapts to regulation or how fast the retailer corrects a mistake.

2.1.2 *Method*

In the crawling and extraction process, the link/element of the energy label image or the energy label element was added and extracted like any other attribute. The link/element attribute is used as a proof, that the energy label in showed correct at the retailer website. If the link/element is missing, then the specific product webpage is marked as non-compliant. We did many manually controls of both pages missing and having the label link/element, to ensure that this approach worked. In the NordCrawl system, the report showing the webpages that are missing the energy label, has hyperlinks directly to those pages, so it is easy to manually double check the results.

Example of energy label links

Figure 1: Energy label links: What the customer sees

Figure 2: Energy label links: What the crawler sees

```
<div class="Pricebox-energyInfo u-nbfc">
  <a href="/api/ProductAsset/EnergyLabel?code=925501033" target="_blank" class="u-pullleft u-marginTsm u-marginRxsm">
    
  </a>
  <a href="/api/ProductAsset/EnergyFiche?code=925501033" target="_blank" class="Pricebox-energyFolder u-linkBorder u-marginRxsm">Produktblad</a>
</div>
```

2.1.3 NordCrawl user interface – missing labels

The user interface showed in the following user guide is a new tool developed in this project.

How to use the missing energy label tool

Figure 3: Step 1 – Choose the “missing labels” page in the menu

Figure 4: Step 2 – Expand the product type

Note: In this example we choose "All dishwashers" by clicking the "+" that will run the analysis. Only product for your country is showed in the list.

Figure 5: Step 3 – List of retailers

- All dishwashers		Has	Missing	Seen	More
runid	URL example				
79878	http://www.elfor.com/usuarios/defectos/defectos-dishwasher-454386	116	0	2018-10-01T10:13:07+00:00	View URLs
79975	http://www.elfor.com/usuarios/defectos/defectos-dishwasher-454386	116	0	2018-10-01T10:45:35+00:00	View URLs
80027	http://www.elfor.com/usuarios/defectos/defectos-dishwasher-454386	81	0	2018-10-01T11:06:28+00:00	View URLs
80136	https://fleming.se/ask/diskmaskiner/80136-55A346C3C25	137	33	2018-10-05T07:28:37+00:00	View URLs

Note: This list shows all the retailers for the selected product type. For each retailer, the number of models that have the energy label and the number of models where the energy label is missing is showed. Click on the "View URLs" button to view the list of links for the chosen retailer.

Figure 6: Step 4 – List of links to product pages for the chosen retailer

All Refrigerator-Freezer - x

Missing:		Has:	
[redacted]	X	[redacted]	X
[redacted]	X	[redacted]	X
[redacted]	X	[redacted]	X
		[redacted]	X

Note: Left: On the left side are list to pages that are missing the energy label for the chosen retailer. The user can click on the link to view the actual product webpage in a new browser tab. That way the user can confirm that it is still missing.

Right: On the right side is all the product pages that has an energy label for the chosen retailer. The user can click on the link to view the product page on a new browser tab, in order see how it looks.

The “X” button removes the link from the analysis. I can be used if the retailer has placed an accessory to a product as a product. If that is the case, the product is not under the regulation and should not be part of the analysis.

2.1.4 Results

The results are a snapshot from the most recent weeks (week 37–40 2018) at the time of analysis. The results may have changed afterwards. The result list below shows the average percentage of product pages that are missing the energy label per country. It also shows percentages of missing energy labels in the best and the worst web shop for each country. The name of the retailers is masked as “ShopX”, relevant energy agencies have the real names. All the online retailers on Iceland that we looked at were missing the labels. Therefore, they are left out of this analysis.

Table 1: Refrigerator-Freezers

	Average missing labels all retailers	Best retailer	Worst retailer*
Denmark	3.66%	Shop1: 0%	Shop5: 18.5%
Sweden	7.68%	Shop8: 0%	Shop14: 27.4%
Norway	2.22%	Shop19: 0%	Shop22: 3.0%
Finland	34.8%	Shop25: 0%	Shop27: 100%
Iceland	-	-	-

Table 2: Refrigerators

	Average missing labels all retailers	Best retailer	Worst retailer*
Denmark	9.50%	Shop1: 0%	Shop6: 28.9%
Sweden	11.64%	Shop11: 0%	Shop14: 29.0%
Norway	5.44%	Shop19: 0%	Shop23: 7.3%
Finland	18.6%	Shop26: 0%	Shop27: 100%
Iceland	-	-	-

Table 3: Freezers

	Average missing labels all retailers	Best retailer	Worst retailer*
Denmark	5.98%	Shop2: 0%	Shop6: 21.6%
Sweden	13.70%	Shop11: 0%	Shop14: 43.5%
Norway	1.36%	Shop19: 0%	Shop21: 1.7%
Finland	21.3%	Shop26: 0%	Shop27: 100%
Iceland	-	-	-

Table 4: Chest freezers

	Average missing labels all retailers	Best retailer	Worst retailer*
Denmark	9.60%	Shop4: 0%	Shop6: 33.3%
Sweden	6.92%	Shop10: 0%	Shop16: 18.8%
Norway	3.28%	Shop20: 0%	Shop23: 7.1%
Finland	24.0%	Shop26: 0%	Shop27: 100%
Iceland	-	-	-

Table 5: Washing machines

	Average missing labels all retailers	Best retailer	Worst retailer*
Denmark	5.31%	Shop3: 0%	Shop6: 17.6%
Sweden	6.18%	Shop11: 0%	Shop14: 11.1%
Norway	3.21%	Shop19: 1.1%	Shop23: 7.3%
Finland	4.9%	Shop25: 0%	Shop28: 50%
Iceland	-	-	-

Table 6: Tumble driers

	Average missing labels all retailers	Best retailer	Worst retailer*
Denmark	12.73%	Shop2: 0%	Shop7: 36.5%
Sweden	15.37%	Shop12: 0%	Shop16: 42.7%
Norway	3.14%	Shop19: 1.4%	Shop22: 7.7%
Finland	19.1%	Shop25: 0%	Shop27: 100%
Iceland	-	-	-

Table 7: Dishwashers

	Average missing labels all retailers	Best retailer	Worst retailer*
Denmark	4.48%	Shop1: 0%	Shop5: 28.6%
Sweden	4.28%	Shop9: 0%	Shop15: 71.4%
Norway	2.41%	Shop19: 0%	Shop22: 3.7%
Finland	16.0%	Shop26: 0.7%	Shop27: 100%
Iceland	-	-	-

Table 8: Vacuum cleaners

	Average missing labels all retailers	Best retailer	Worst retailer*
Denmark	9.85%	Shop1: 0%	Shop6: 25.5%
Sweden	13.58%	Shop13: 0%	Shop10: 16.2%
Norway	40.72%	Shop21: 16.9%	Shop23: 56%
Finland	0.4%	Shop26: 0%	Shop29: 12.5%
Iceland	-	-	-

Table 9: TV

	Average missing labels all retailers	Best retailer	Worst retailer*
Denmark	7.51%	Shop1: 0%	Shop7: 18.8%
Sweden	15.05%	Shop13: 0%	Shop18: 30.9%
Norway	12.84%	Shop19: 0%	Shop24: 27.1%
Finland	8.1%	Shop26: 0%	Shop28: 100%
Iceland	-	-	-

Note to tables above

*It should be noted, that a problem rarely occurs alone. A retailer with a low compliance rate in this analysis often has other compliance issues like missing important product information, product fiche etc.

A general outcome is that if a retailer is good in one product category, it is often good at all product categories and likewise if a retailer has a low compliance it applies to most of the products categories.

2.1.5 Challenges

One interesting thing we noticed in this analysis is that some retailers have some rather alternative interpretations of the requirements, when they add the energy label image. They do not have a uniform way of presenting the image. In some cases, they put the images at the end of a product images slides, so the customer has to slide through a number of images before the label is showed. In other cases, they add the energy class arrow, but it does not link to anything.

2.2 Market surveillance analysis – banned energy classes due to ecodesign limits

2.2.1 Purpose

The ecodesign directive 2009/125/EC and its product specific regulations set minimum performance standards for a range of products. It is not legal to introduce a new a product on the market in EU and EFTA countries, that has an efficiency below the minimum requirement. In this NordCrawl analysis we look at a method to identify models that does not comply with the ecodesign regulation.

2.2.2 Method

The ecodesign product regulations does not define a minimum energy class but a minimum Energy Efficiency Index (EEI) or minimum consumption per year in kWh. The first step in the method is to identify the corresponding minimum allowed energy class. In the NordCrawl system, these minimum energy classes are saved in the database as compliance rules for each product group. Then the compliance rule tool searches for models in the database, that has an energy class below the minimum requirement. See description in figures below.

2.2.3 User interface – banned energy class

How to use the find banned energy class tool

Figure 7: Step 1 – Choose the “Compliance” page in the menu

Figure 8: Step 2 – Choose the “Energy label”

Figure 9: Step 3 – Table of minimum energy label compliance rules

Compliance Rules - Min Energy label

Rule type	Product	Product name	In power	Min value	Active	Find non-compliant	Regulation
Energy class		All Washingmachines	12-2013		✓	Find non-compliant	1015/2010
Energy class		All Vacuum cleaners	1-2014		✓	Find non-compliant	666/2013
Energy class		All range hoods	1-2016		✓	Find non-compliant	66/2014
Energy class		All Ovens	1-2016		✓	Find non-compliant	66/2014
Energy class		All Refrigerator-Freezer	7-2014		✓	Find non-compliant	643/2009
Energy class		Refrigerator only	7-2014		✓	Find non-compliant	643/2009
Energy class		All freezers	7-2014		✓	Find non-compliant	643/2009
Energy class		All Chest Freezers	7-2014		✓	Find non-compliant	643/2009
Rule type	Product	Product name	In power	Min value	Active	Find non-compliant	Directive

Note: Each product type that has a compliance rule is showed in the compliance rule table. A compliance rule have an “in power” date that shows when the rules enters into power. A rule also has a minimum energy class, a link to the ecodesign regulation and a “Find non-compliant” button. In this example, we click on the “Find non-compliant” button for refrigerator-freezers.

Figure 10: Step 4 – Compliance analysis results and non-compliant models

Note: The compliance analysis results page shows the non-compliance rate, the number of non-compliant models and the total number of models that was evaluated for non-compliance.

Below that is a table with the non-compliant models containing the energy class of the model and a link the model page. The user can investigate the model further by clicking on the model menu and look at the sources of the model.

2.2.4 Results

Table 10: Non-compliance minimum energy class results

	Minimum	DK	SE	NO	FI	ICE
Washing machines	A+	0.75% (4)	1.36% (7)	0.44% (2)	0.67% (1)	0% (0)
Vacuum cleaners	D	1.86% (6)	0.6% (1)	0% (0)	5.22% (7)	0% (0)
Range hoods	F	0% (0)	0% (0)	0% (0)	0% (0)	-
Ovens	C	0% (0)	0% (0)	0% (0)	0% (0)	-
Refrigerator-Freezer	A+	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Refrigerators	A+	1.19% (9)	0% (0)	0.83% (4)	0.36% (1)	0% (0)
Freezers	A+	0.22% (1)	0% (0)	0% (0)	0% (0)	0% (0)
Chest Freezers	A+	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)

In general, the results of this analysis showed good results. It was not investigated whether the non-compliant models were introduced to the market before the regulation was in power. That could be the case for the vacuum cleaners. We have identified and removed some false positive from the washing machine and refrigerator categories. Some retailers had categorized combined washing/dryers as a washing machine.

3. Policy Evaluation – market share calculation

3.1 Market share calculation

3.1.1 *Intro*

The market share of a model or group of models is an important input, in order to use the crawled data to do police evaluation. Currently, the data in the system is product weighted, meaning that every model has the same weight. An example, if 10 out of 100 models have a “A+++” label, 10% is “A+++”. The challenge with this approach is, that in the real word not all models are equally popular, some models are sold more frequently than others. If we want to know what the consumers are actually buying, then we need to know how many of each model are sold within a given timeframe, or how many models within a specific group, like energy class A+++, are sold.

One of the reasons why it is relevant to know what is sold and not just what is for sale, is to use that knowledge to evaluate as given policy. One example is to evaluate how much electricity that has been saved by the ecodesign directive, the energy labelling framework regulation, and their product specific regulations. This is something that has been calculated in the past in Denmark and Sweden using real sales data from Applia.

In the following, we will look at various approaches to come up with a market share estimate, one advanced based on a popularity ranks from a from a price comparison site, and some more simple methods based on available attributes.

3.1.2 *Data*

The following data used (besides crawled retailer data):

- Swedish GfK data from three month in 2013–2014 for dishwashers, washing machines, TVs, refrigerators and refrigerator-freezers. The sales number for each model is included. GfK claims to have a market coverage between 83%–93%;
- Sales data from The Danish Association for Suppliers of Electrical Domestic Appliances (Applia) for whitegoods on a monthly basis. The data is aggregated on group levels like energy classes;
- Popularity ranks from crawling large price comparison sites. The data is a rank of each model in a product category at the time of crawling.

3.1.3 *Method*

In this project we looked at four market share approximation methods and compared the results to the real market data (from Applia) and a baseline.

As a baseline, the product weighted is selected. The product weight is where each model weight equally. In other words, the evaluated method needs to perform better than not using a weighting method. We know from experience that non-weighting or model weighted results are relatively close to the real sales weighted results when looking at energy class distribution, therefore this was chosen as a baseline. It can also be observed in the calculation of correlations coefficients, that the baseline is highly correlated.

We evaluate the results by comparing the calculated energy class distributions to the baseline distributions (product weighted) and the real energy label distribution from Denmark. We assume that the Danish distribution is representative for the other Nordic countries. This assumption is based on the similarities between the countries, and that the major retailers are represented in most of the Nordic countries.

We use the sum of squared deviations to measure how close the result is to the real distribution. An example, if the result shows that the 44% of the market is energy class A+++, and the real A+++ share is 40% then the deviation is 4 and the squared deviation is $4^2 = 16$. Then we take the sum of all squared deviations and compare the results. In the following we call the sum of squared deviations R_2 . The smaller R_2 the better. Furthermore, Pearson correlation coefficient is used to calculate the degree

of the correlation between result and the real distribution. The closer the Pearson correlation coefficient is to 1, the stronger the positive correlation is.

3.2 Market share approximation methods

3.2.1 Market share from sales rank method

Robert Van Buskirk (Touzani, S., & Buskirk, R. 2015)² presented a method that relates the sales rank of a model to the sales volume of that model. The paper shows that the log-normal distribution and specifically the truncated version are well suited for that purpose. The method requires a sales rank. In our case we use a popularity rank from a price comparison site as the sales rank. The equation is showed below. We will not explain the details of the equation here, but just refer to paper for more information.

The idea is that the distribution function has a shape very close to the shape of the real market share distribution, where the most popular models account for the largest market share and market share falls when the popularity falls and finally flattens out. This can be seen in the chart below, where the actual market share for each model is marked with a blue dot, and the estimated market share curve is showed with an orange line. We used GfK data from 2014 to calibrate the equation, and the example below shows that the result confirms that the estimates are close to the real data.

Figure 11: Sales rank method equation

$$\hat{s}_i^{TLN} = \exp \left(\sigma \Phi^{-1} \left[\alpha \left(1 - \frac{r_i}{N} \right) + \beta \right] + \mu \right) \quad (1.7)$$

with

$$\alpha = \Phi \left(\frac{\ln b - \mu}{\sigma} \right) - \Phi \left(\frac{\ln a - \mu}{\sigma} \right) \quad (1.8)$$

and

$$\beta = \Phi \left(\frac{\ln a - \mu}{\sigma} \right) \quad (1.9)$$

² Touzani, S., & Buskirk, R. (2015). Estimating Sales and Sales Market Share from Sales Rank Data for Consumer Appliances. Lawrence Berkeley National Laboratory. LBNL Report #: LBNL-6989E. Retrieved from <https://escholarship.org/uc/item/9699p060>

Figure 12: Estimated market share and popularity rank

One challenge with this approach is, that the equation is fitted to data from 2014. To study if it will affect the results today four years after, we applied the distribution equation to Swedish data and compared the estimated energy label distribution with real energy label distribution. An example for dishwasher data from September 2018 is showed in the table below. It shows that the pretty close to the real distribution and a lot better than the baseline.

Table 11: Market share from sales rank method – evaluation result

	Product weight (baseline)	Rank method	Real distribution (From DK)	Deviations
A+++	32%	40%	42%	Product weight (baseline) to real R2 167.2619 (Pearson: 0.9640)
A++	35%	32%	34%	
A+	25%	25%	22%	
A	8%	2%	1%	Rank method to real R2 17.33014 (Pearson: 0.9537)
B	0%	0%	0%	
C	0%	0%	0%	
D	0%	0%	0%	

The first problem with this approach is, that it requires a popularity rank for all models, though it can be argued the if the models is not at the price comparison site, it is not very popular and does not account for a very significant market share. The problem is, however, that we currently do not have a popularity rank for all products in all countries.

The second problem is that we only have data to fit the equation for a few products: dishwasher, washing machine, TV, refrigerator and refrigerator-freezers. Furthermore, this approach has not been tested on product types that are very different from TVs or whitegoods.

3.2.2 Market reach method

The popularity rank is not always available for a certain product or an entire market. There is therefore a need for another approach to estimate the market share with the data available. We developed the market reach method to address this challenge.

In the market reach method, our thesis is that market share correlates with the number of potential customers for a model. The reach is a function of the number of stores selling the model. Popular models are sold by more online retailers than the less popular ones. Our equation for market share using market reach is showed below.

Figure 13: Market reach equation

$$\text{Market Share} = \frac{\text{number of models} * \text{Number of shops}^{\text{shop factor}}}{\sum \text{number of models} * \text{Number of shops}^{\text{shop factor}}}$$

The method was tested by fitting the shop-factor variable, so that the result was as close to the real energy label distribution as possible. The results for all cases are, that it is as good as the baseline. Hence, this method is not better than the baseline and should not be used.

3.2.3 *Shop size and normal pricing method*

This method is developed by our self. It is based on two input variables, the size of the shop selling the model and the price. The size is measure by how many other models the shop is selling in the specific product category. The thesis is, that the more models a shop has for sale, the fewer is selling of each model. Likewise, if a shop only has a few models for sale then the sale per model is higher. We calculate this by giving each shop 1,000 points to split between the models. An example, if shop have 20 models for sale, then the points given to each model is $1,000/20 = 50$.

The other input variable is the price. The thesis is, that models having a price close to or below the average price are more popular than more expensive models.

The equation for this method is showed below:

Figure 14: Shop size and normal pricing method equation

$$\begin{aligned} & \text{Sales approximation (model } x) \\ &= \left(a * \frac{1,000}{\text{number of models in shop}} \right) \\ &+ \left(b * \sqrt{\text{price above average}} \right) \end{aligned}$$

The variables a and b are fitted to match expected results using a solver function.

The results of this method are not convincing. In some cases, it is even worse than using the baseline, while in other cases it is slightly better. The example below, shows the result for tumble dryers from Denmark in August 2018. As we can see, the result is only slightly better than the baseline.

Table 12: Market reach method – evaluation result

	Product weight (baseline)	Method	Real distribution (From DK)	Deviations
A+++	16%	18%	12%	Product weight (baseline) to real R2 238.88877 (Pearson: 0.9385)
A++	50%	48%	55%	
A+	11%	12%	6%	
A	1%	0%	8%	Method to real R2 187.97795 (Pearson: 0.9537)
B	6%	14%	13%	
C	16%	7%	7%	
D	0%	0%	0%	

3.2.4 Conclusion method 2 and 3

Method 2 and 3 is not an exhaustive list of methods that can be used but exemplifies attempts to design a method that can be used to estimate the market share of the appliances. Although the methods do not show convincing results, it cannot be excluded that they can be used in other markets or for other types of products with other characteristics. One of the reasons why the results are not convincing may be that the baseline is quite good, which is not a bad thing.

3.2.5 Average sales rank method

Now we have looked at some different approaches to calculate the market share. So far, the best method is the sales rank method, but it is limited to the five product groups where we have data to fit the equation. When looking at the results' input variables, Mean (μ), Standard deviation (σ), and number of products sold (N) for each product type, they appear to be very close to each other. Mean (μ) does not change because we want the relative market share for each product and not the sales volumes. The Standard deviation (σ) defines the shape of the slope and number of products (N) defines the tail of the slope.

Table 13: Average sales rank method equation input variables

	Dishwashers	Washing machines	TV	Refrigerators	Refrigerator-freezers	Average
Mean (μ)	0.4668499	0.4668499	0.4668499	0.4668499	0.4668499	0.4668499
Std dev (σ)	1.5908267	1.4293448	1.2566234	1.3456263	1.4550102	1.4154863
N	533	436	913	351	617	570

This indicates that the shape/slope of the various sales distributions are not very different, so an average method may be used for other product types where we do not have a data fitted equation. Our thesis is, that if the shapes does not vary too much, then an average model can be "good enough" to predict the market share based on the popularity rank. Another way to explain this method is that if you have two different types of products that are alike and a sale that is the same, then the two most popular products from each type will have the same market share, the two second most popular products also have the same market share and so on. The market shares will of course never be exactly the same, but the thesis is that the small differences does not matter on an aggregated level like distribution of energy classes.

The results are quite convincing. An example is showed the table below, where the average sales rank method is applied to upright freezers in Sweden in august 2018. The results are showing, that this method can be used for all the product types where we have a popularity rank.

Table 14: Average sales rank method – evaluation result

	Product weight (baseline)	Rank method	Real distribution (From DK)	Deviations
A+++	7%	4%	5%	Product weight (baseline) to real R2 36.550813 (Pearson: 0.9942)
A++	56%	53%	53%	
A+	36%	42%	41%	
A	1%	1%	0%	Rank method to real R2 0.9142601 (Pearson: 0.9999)
B	0%	0%	0%	
C	0%	0%	0%	
D	0%	0%	0%	

3.2.6 Market share calculation applied

In the table below, the market calculation result from the sales rank method is applied to the result of the Ecodesign banned product analysis. The table shows the calculated market share for the non-compliant models. This analysis shows that the banned products do not represent a large share of the sold products. If any, vacuum cleaners should be followed up.

Table 15: Ecodesign – market share of non-compliant models (Ecodesign banned products)

	Minimum	DK*	SE	NO	FI	ICE*
Washing machines	A+	0.75% (4)	0.95% (7)	0.54% (2)	0.04% (1)	0% (0)
Vacuum cleaners	D	1.86% (6)**	0.6% (1)**	0% (0)	5.22% (7)**	0% (0)
Range hoods	F	0% (0)	0% (0)	0% (0)	0% (0)	-
Ovens	C	0% (0)	0% (0)	0% (0)	0% (0)	-
Refrigerator-Freezers	A+	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Refrigerators	A+	1.19% (9)	0% (0)	0.0% (4)	0.04% (1)	0% (0)
Freezers	A+	0.22% (1)	0% (0)	0% (0)	0% (0)	0% (0)
Chest Freezers	A+	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)

Note: *We do not have any popularity rank for Denmark or Iceland at the moment. Therefore, the market share is the product weighted.

** We do not have a popularity rank method for vacuum cleaners at the moment. Therefore, the market share is the product weighted.

This “average sales rank method” could also be used to calculate the market share of A+++ or A++ products, which is a good input for policy discussions regarding whether or not the ecodesign limits and energy labelling classes are correctly set. This analysis show that we have methods to calculation market shares on a model level. This could be developed further as a feature in the NordCrawl system.

Conclusions

Within this project the NordCrawl-tool was used for market surveillance and policy applications. For market surveillance two new features were added; to identify webpages with missing labels and to find products with energy classes below the requirements. The missing energy label analysis found quite many pages with missing labels, and this can be a good input to the market surveillance authorities. The analysis results on products with an energy class below the requirement did not find so many faulty products, but still some interesting indications for the market surveillance authorities.

The market share calculations for policy evaluation showed that the sales rank method is very good and can also be used for product groups where you do not have sales numbers as long as you have a popularity rank. This can lead to helpful analyses on how many products on the market are in the A+++-class or, if using the previous NordCrawl2 result, can show how large market share are below ecodesign limits.

This project showed that the NordCrawl-tool can indeed already be a useful tool for both market surveillance and policy evaluation for ecodesign and energy labelling. Still it has the potential to be further developed.

Recommendations and ideas for future work

The analyses within this project lead to ideas where further work and new analyses can be done:

- Missing product fiche – A functionality similar the missing energy label could easily be implemented to make market surveillance even more efficient;
- Ecodesign other minimum standards – In this project we have looked at minimum energy efficiency performance standards. Likewise, it would not be particularly demanding to look at similar standards like requirements for noise;
- Less false positive – in the “banned” energy classes analysis we removed some false positives. The reason was that the models were put into the wrong category by the retailer. In the future, it will be an advantage to have a tool to identify and recategorize this type of error;
- New energy label – When the energy label is rescaled, the NordCrawl system may need some adjustment to handle the rescaling. One example is the time series analysis. It might also be interesting to analyze how fast the market adjust to the news scale;
- Connection to EPREL database – Connection of NordCrawl to the EPREL database would be useful for a number of reasons. One of the opportunities is to use the database as list of “true” attributes, and then use it to validate the data from the retailer’s website and identify errors in the retailer’s data. Another option is to catch “freeriders”, manufacturers that does not register in the database;
- Risky product –the NordCrawl system could be further developed to identify risky products (likely to be non-compliant);
- Calculate what market surveillance has saved (using sales + kWh faults of the products tested at market surveillance);

- Consumer interface – One opportunity that have been discussed for some time is to make a consumer interface, where data in the NordCrawl system can be used to guide consumer to buy more energy efficient products;
- Add more product categories and develop the ones already covered.

Sammanfattning

Projektet Nordcrawl2 har utförts för det nordiska samarbetet Nordsyn, finansierat av Nordiska Ministerrådet. Inom detta projekt har webb-verktyget Nordcrawl utvecklats för marknadskontroll – för att hitta saknade energimärkningsetiketter och produkter som inte klarar ekodesignkraven; samt för policy – där en metod för att uppskatta marknadsandel utvecklats. Projektet visar att Nordcrawl-verktyget redan kan vara användbart för både marknadskontroll och policyutvärdering för ekodesign och energimärkning. Samtidigt har Nordcrawl-verktyget stor potential att utvecklas vidare och några idéer och rekommendationer ges i denna rapport.

Nordic Council of Ministers
Nordens Hus
Ved Stranden 18
DK-1061 Copenhagen
www.norden.org

NORDCRAWL 2

The NordCrawl2 project was done for the Nordic Nordsyn cooperation, sponsored by the Nordic Council of Ministers. Within this project the webcrawling Nordcrawl-tool was developed for market surveillance – to find missing labels and products not meeting ecodesign limits – and for policy – where a method to estimate market share was developed.

The project shows that the Nordcrawl-tool can indeed already be a useful tool for both market surveillance and policy evaluation for ecodesign and energy labelling. Also the Nordcrawl tool still has great potential for further developments, and some ideas and recommendations are given in this report.

9 789289 359740