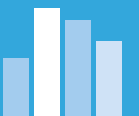


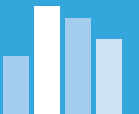
# Presentation of Thomas Jacobsen

- Graduated with a master in Economics from University of Aalborg in summer of 2013
- Working in Statistics Denmark since September 2013
- Involved in the project for the introduction of Scanner data in the Danish CPI/HICP
- Co-responsible for the production of the Danish CPI/HICP from 2013-2017
- Member of Eurostat Expert Group on scanner data and UN Global Working Group on scanner data
- Involved in a project to link data from data providers of commercial properties with Danish register of housing in cooperation with the Danish tax authorities
- In my free time I'm a keen racing biker and big movie fan



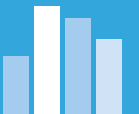
# Session 1.1

- **Overview of scanner data**



# Sources

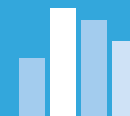
- In general scanner data can either be collected at the individual firm (Supermarkets, petrol firms and so on)
- Or by marketing bureaus- e.g. AC Nielsen



# Dimensions of data

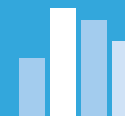
- We receive scanner data per week, pre-aggregated by:

***Week per shop per GTIN***



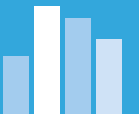
# The nature of scanner data?

Date	Shop	GTIN number	Store id	Turnover	Volume	Unit	Amount per unit	Product number	Product description
1104	Shop1	1234567891 011	1234	3402.70	211	Gram	300	1234567	Bacon 2x150 G.
1104	Shop2	1234567891 011	1235	2119.65	163	Gram	300	1234567	Bacon 2x150 G.
1104	Shop3	1234567891 011	1236	1516.05	108	Gram	300	1234567	Bacon 2x150 G.
1104	Shop4	1234567891 011	1237	1478.13	105	Gram	300	1234567	Bacon 2x150 G.
1104	Shop1	9876543210	1238	302.50	14	Gram	200	7654321	Chicken filet
1104	Shop2	9876543210	1239	102.50	5	Gram	200	7654321	Chicken filet



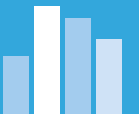
# This implies

- **Number of lines:** Approximately 11 million
- **Number of shops:** More than 2.000 shops
- **Number of distinct GTINs:** more than 150.000



# General application

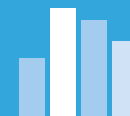
- In general the unit price is applicable. However the unit price followed over time needs to be for homogeneous goods
- Scanner data is often used mainly at COICOP 1 and 2, because of the low attrition rate
- This means that GTINs might be the best level of product identifier here
- The use in COICOP groups with higher attrition might imply the need for less static methods



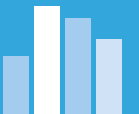


# Countries that use scanner data either in production or in testing

- In production for sure:
- Netherlands
- Belgium
- Norway
- Sweden
- Denmark
- Iceland
- Switzerland
- New Zealand
- Australia

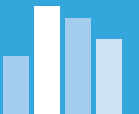


- Testing:
- Most other (western) European countries ( Germany, Austria, France etc.)
- Canada



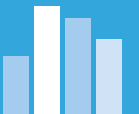
# Some examples of International experiences: Netherlands

- Started with scanner data in production around 10 years ago
- They have used several methods (bilateral methods), but have now switched to a GK-method in production
- Extensive research is carried out in the Netherlands
- Moving coverage outside COICOP 1 and 2



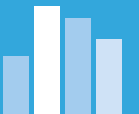
# Belgium

- The started with scanner data in production a couple of years ago.
- The use a dynamic method with dumping filters
- The have relatively good product identifiers (SKU)
- Moving coverage outside COICOP 1 and 2



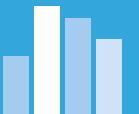
# Norway:

- Implemented scanner data more than a decade ago. They had relatively easy access to the data
- At first they used a method where the individual shop was the identifier- later they switched to a higher level of aggregation
- They use a dynamic method where they put in dumping filters
- Moving coverage outside COICOP 1 and 2



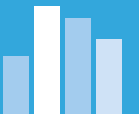
# Sweden

- Started with scanner data in production around 10 years ago
- They use a method of sequential Poisson sampling (semi-dynamic method)
- Moving coverage outside COICOP 1 and 2



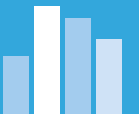
# Switzerland

- Started with scanner data in production around 10 years ago
- The use a static sample method where resampling is essential
- Is the method most similar to the Danish experience, which I will show you tomorrow



# France

- They had difficulty getting scanner data
- They have now passed a law where shops over a certain size (square metres) need to provide INSEE with scanner data
- They are now experimenting on scanner data





# Tentative conclusion

- A lot of methods are used
- A lot of applications are found ( both within COICOP and within providers)
- The aim of the workshop is to get an insight into methods that can be tested without having access to scanner data

