

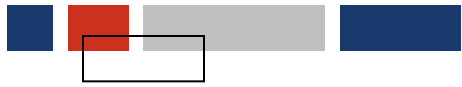
What price? How to combine scanner data with traditional price collection?



Vitor Mendonça
National Accounts Department

Scanner Data Workshop
26|27 September 2013
Lisbon, Portugal



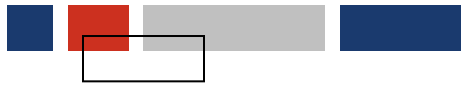


Index



- Background: Scanner Data
- Theoretical and conceptual aspects
- Study with actual data
- Conclusions / further work





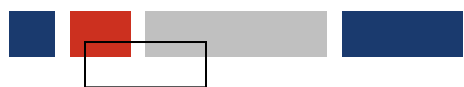
Background: Scanner Data



- Scanner Data: electronic data on sales provided by retailers
- Statistics Portugal study: two providers
 - **SONAE** group (33.5%)
 - **Jerónimo Martins** group (32.6%)

(turnover, food products, 2011)



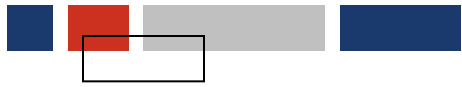


Background: Scanner Data



- Data provided
 - EAN/SKU – numeric code, theoretically unique for each product
 - Product description – text field, not standardized
 - Shelf price (daily) – “marked” price, price paid by any customer
 - Quantity (daily) – number of items sold each day



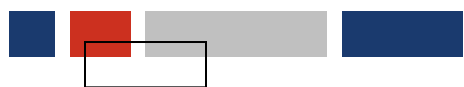


Theoretical & conceptual aspects



- Main requirements for HICP
 - Timeliness of data delivery
 - Linking of EAN/SKU to COICOP classification
 - Compliance with Legal Framework
 - Consistency, comparability and quality



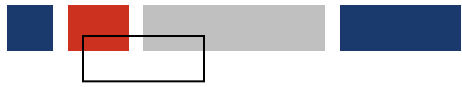


Theoretical & conceptual aspects



- Potential uses of Scanner Data
 - Quality control
 - Quality adjustments
 - Item selection
 - Weights compilation
 - Price collection



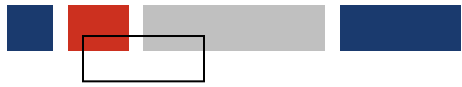


Quality control



- Main goal
 - Check if price collectors are registering the correct prices





Quality control

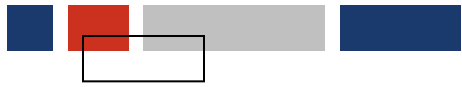


- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- Timeliness requirements are less important
- If EAN/SKU of items which are collected are known, there is no need for COICOP classification of Scanner Data
- No problems related to Legal Framework
- Consistency and comparability are unaffected; can lead to quality improvements



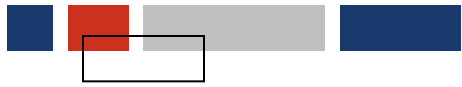


Quality control



- Potential drawbacks
 - Price collectors may sense lack of confidence in their work
 - Unless EAN/SKU is known, this can be time-consuming, as items must be “found” in the database and quality control cannot be exhaustive





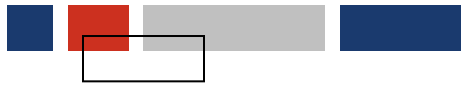
Quality control



■ Suggestions

- Price collectors should collect EAN/SKU
 - Manually – has potential for errors but is easily implementable (low cost)
 - Automatically – needs specific hardware to read barcodes
- Price collectors should see this as a form of quality improvement, not as a potential problem for them





Quality adjustments



- Main goals

- Use Scanner Data from $t-1$ for replacement products, in order to ensure that only pure price changes are considered
- Additionally, replacement products may be discarded if Scanner Data proves that either
 - replaced product is still available and representative or
 - replacement product is clearly not representative – suggestion of another replacement product for $t+1$





Quality adjustments

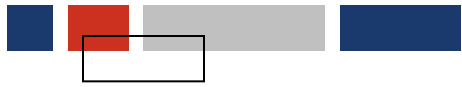


- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- Timeliness requirements must be met, allowing for enough time to process the data before the deadline; in case of delays, methods that don't rely on Scanner Data can still be used
- If EAN/SKU of items which are collected are known, there is no need for COICOP classification of Scanner Data
- No problems related to Legal Framework – pure price changes are considered – no judgments, just use the data





Quality adjustments

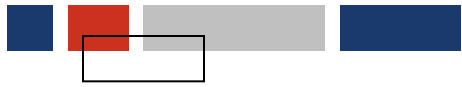


- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- Consistency and comparability can be affected – data is not available from all the stores, different quality adjustment methods will be used – but current practice already has this drawback;
- Can lead to quality improvements



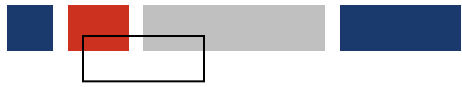


Quality adjustments



- Potential drawbacks
 - Unless EAN/SKU is known, this can be time-consuming, as items must be “found” in the database
 - Replacement product can be *new* – no data for *t-1* – another quality adjustment method must be used





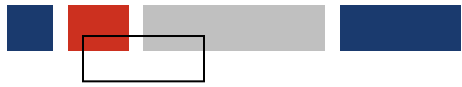
Quality adjustments



■ Suggestions

- Price collectors should collect EAN/SKU
- Quality adjustments can be mostly automatic (must still require validation of the replacement product)
- Use scanner data as a source for a list of possible replacement products (requires linking of EAN/SKU to COICOP)



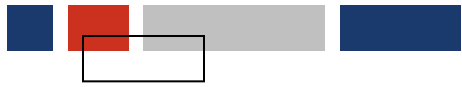


Item selection



- Main goal
 - Use Scanner Data from the previous 12 months for product selection, in order to choose the most representative items in each store for inclusion in the HICP basket





Item selection

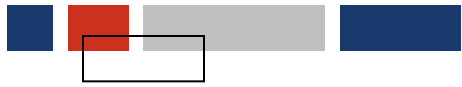


- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- Timeliness requirements must be met, allowing for enough time to process the data for the chain-linking process; on the other hand, one can always choose to consider the latest 12 months available
- Linking of EAN/SKU to COICOP must be ensured, but
 - This process can be conducted throughout the year, not only on the chain-linking process
 - This process does not need to be exhaustive, as long as the most representative items are considered [at store level]





Item selection



Example:

- Only 3 products to be classified for this store: Milk, Butter, Rice
- Each has several different varieties
- Fixed number of items: collect only the most sold variety of each product
- In the end, classification can stop after line 5

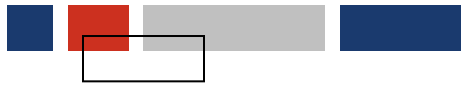
This process does not to be exhaustive, as long as the most representative items are considered

- As long as there is a fixed number of items to be considered, sorting by expenditure will probably result in earlier end of the process

Preliminary testing with real data: 90% of expenditure covered by top 30% of all varieties

Line	Product	Expenditure	COICOP
1	Milk (1)	1000	01.1.4.1.
2	Milk (2)	945	01.1.4.1.
3	Butter (1)	900	01.1.5.1.
4	Milk (3)	894	01.1.4.1.
5	Rice (1)	853	01.1.1.1.
6	Butter (2)	785	-
7	Rice (2)	356	-





Item selection



- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- Potential problem with the Legal Framework
- **Household final monetary consumption expenditure**
- Scanner Data does not separate household expenditure; business like restaurants can account for a large amount of recorded expenditure





Item selection

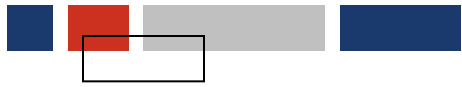


- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- Consistency and comparability can be affected – data is not available from all the stores, non-Scanner Data varieties will still be chosen based on the price collectors judgement – but still an improvement over the current practice
- Can lead to quality improvements



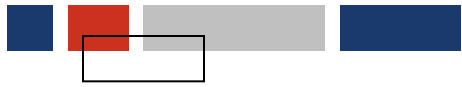


Item selection



- Potential drawbacks
 - Linking EAN/SKU to COICOP is time-consuming, even without considering all the varieties
 - Since new products may not have been on sale for all 12 previous months, they may be discarded based on reported expenditure, so expert judgment is still essential





Item selection



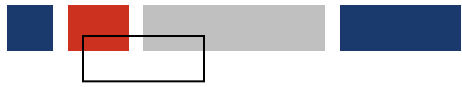
▪ Preliminary testing

- Most varieties chosen by price collectors as being the most representative are validated by Scanner Data, but some are not the most representative
- Product replacements are not always required – price collector does not find the expected variety and chooses another, even though Scanner Data shows that the variety was available and bought by customers

▪ Suggestions

- Request data without business expenditure
- Select a fixed number of items to lessen the burden of classification
- Pre-select substitute varieties to maintain representativity
- Compare with selection based on the last available month to identify newly introduced products
- Avoid selection of items that were unavailable in the last month



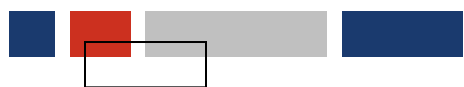


Weights compilation



- Main goal
 - Use Scanner Data from the previous 12 months for weights calculation
 - Varieties weights
 - Product weights
 - Store weights





Weights compilation

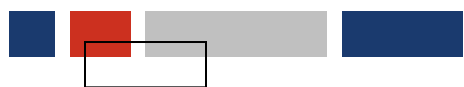


- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- Timeliness requirements must be met, allowing for enough time to process the data for the chain-linking process; on the other hand, one can always choose to consider the latest 12 months available
- **Linking of EAN/SKU to COICOP must be exhaustive in order to calculate product weights**





Weights compilation

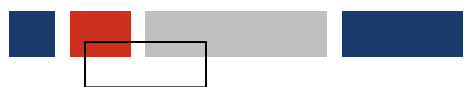


- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- Potential problem with the Legal Framework
- **Household final monetary consumption expenditure**
- Scanner Data does not separate household expenditure; business like restaurants can account for a large amount of recorded expenditure
- Scanner Data should only be used as a complement of National Accounts and Household Budget Surveys





Weights compilation

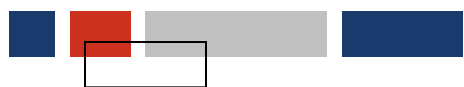


- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- Consistency and comparability can be affected – data is not available from all the stores – thus Scanner Data should only be used to compile weights between varieties in each store
- Product weights – use of Scanner Data instead of HBS can only be reliable if data is available for *all* kinds of stores, not only big chains that provide Scanner Data – but even so, business expenditure must be excluded
- Unclear whether this will lead to quality improvements





Weights compilation



- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- Store weights – overall expenditure (all transactions) can be used to compile these weights, but use of administrative data may be more adequate, as there should be data for all stores
- Unclear whether the use of Scanner Data for weights compilation will lead to quality improvements



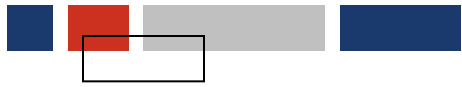


Weights compilation



- Potential drawbacks
 - Exhaustive linking of EAN/SKU to COICOP
 - Availability of Scanner Data for all stores is highly unlikely





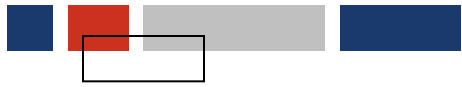
Weights compilation



■ Suggestions

- Request data without business expenditure
- Use scanner data from the last 12 months to calculate relative weights of varieties in each store (if more than one variety is selected)



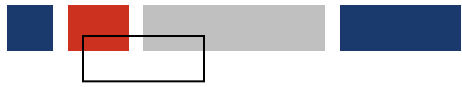


Price collection



- Main goal
 - Use Scanner Data as source for prices





Price collection

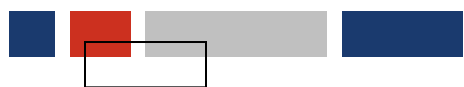


- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- **Timeliness requirements must be met, allowing for enough time to process the data for publication (including replacements and quality adjustments); any delay will compromise the HICP**
- Linking of EAN/SKU to COICOP depends of item selection method





Price collection

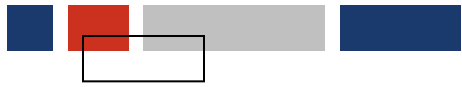


- Timeliness of data delivery
- Linking of EAN/SKU to COICOP
- Compliance with Legal Framework
- Consistency, comparability, quality

▪ Main requirements for HICP

- No problems related to current Legal Framework
- Consistency and comparability can be affected – data is not available from all the stores
- Use of Scanner Data should not decrease quality – but timeliness is essential



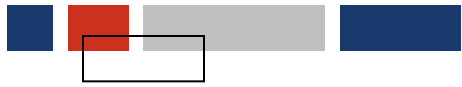


Price collection



- One product, one shop = one variety
- Should this *rule* be changed because of Scanner Data?
- If this *rule* is changed, how about other stores that don't provide Scanner Data?
- Is there a need for a backup plan if there is some failure on the data transmission process?
- How can this backup plan be implemented in a time-effective way if more than one price is considered on Scanner Data?



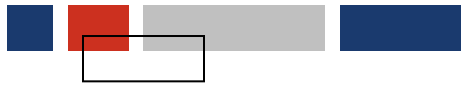


Price collection



- Potential drawbacks
 - NSIs become highly dependent on data providers – potential risks of missed deadlines
 - More data = more time
 - Product replacements
 - Quality adjustments
 - Classification





Price collection



■ Suggestions

- Methodological changes in the HICP because of Scanner Data can lead to problems
- Keep it simple, one step at a time, to avoid impacts on the HICP



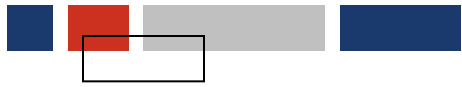


Study with actual data



- Goal: compare results with and without using Scanner Data
- COICOP 01
- 24 months – 01.2011 to 12.2012
- Monthly data [daily data for some months]
 - Big volume of data
 - Average monthly price [p^*q/Q]
- 2011 – sampling; 2012 – prices



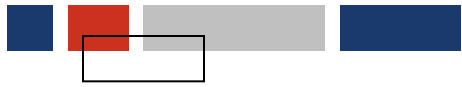


Study with actual data



- One region: Norte
- Data from **SONAE** group
 - 19% of CPI prices collected in these stores
 - 20 stores in the HICP; 60 total
- Only 4 stores considered so far
 - Classification: 5% of varieties for each store
 - Varieties with largest expenditure



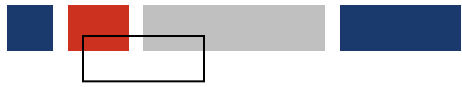


Study with actual data



- CPI practice:
 - For each store, price collectors are instructed to select the variety that they think is more representative
 - Each variety is considered until it is unavailable or not representative
- About 2% of prices replaced
- 3 methods of replacement
 - For each product, CPI price is replaced with
 - Most sold variety, if priced for all months
 - Most sold variety with prices for all months
 - Geometric average of all available prices





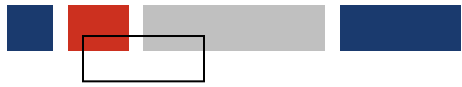
Study with actual data



- COICOP 01 – Norte
- 265 products
- Up to 43 stores per product
- On average, about 11 000 varieties are considered each month

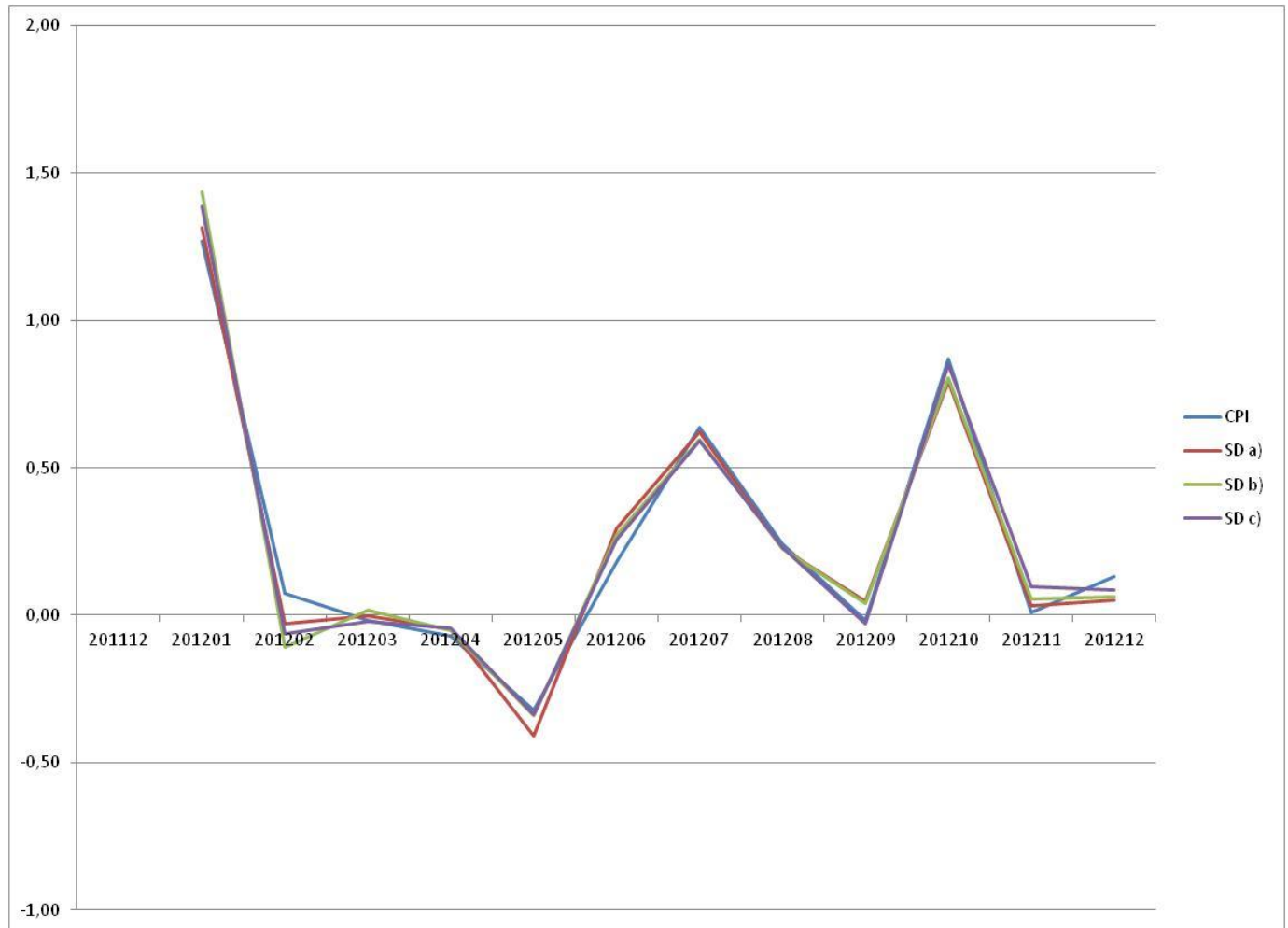
- Most sold variety, if priced for all months
 - 122 products with at least 1 SD price
 - 185 prices replaced each month
- Most sold variety with prices for all months
 - 141 products with at least 1 SD price
 - 222 prices replaced each month
- Geometric average of all available prices
 - 146 products with at least 1 SD price
 - 233 prices replaced each month

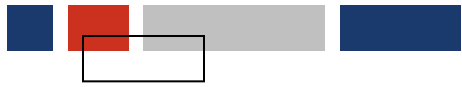




Study with actual data

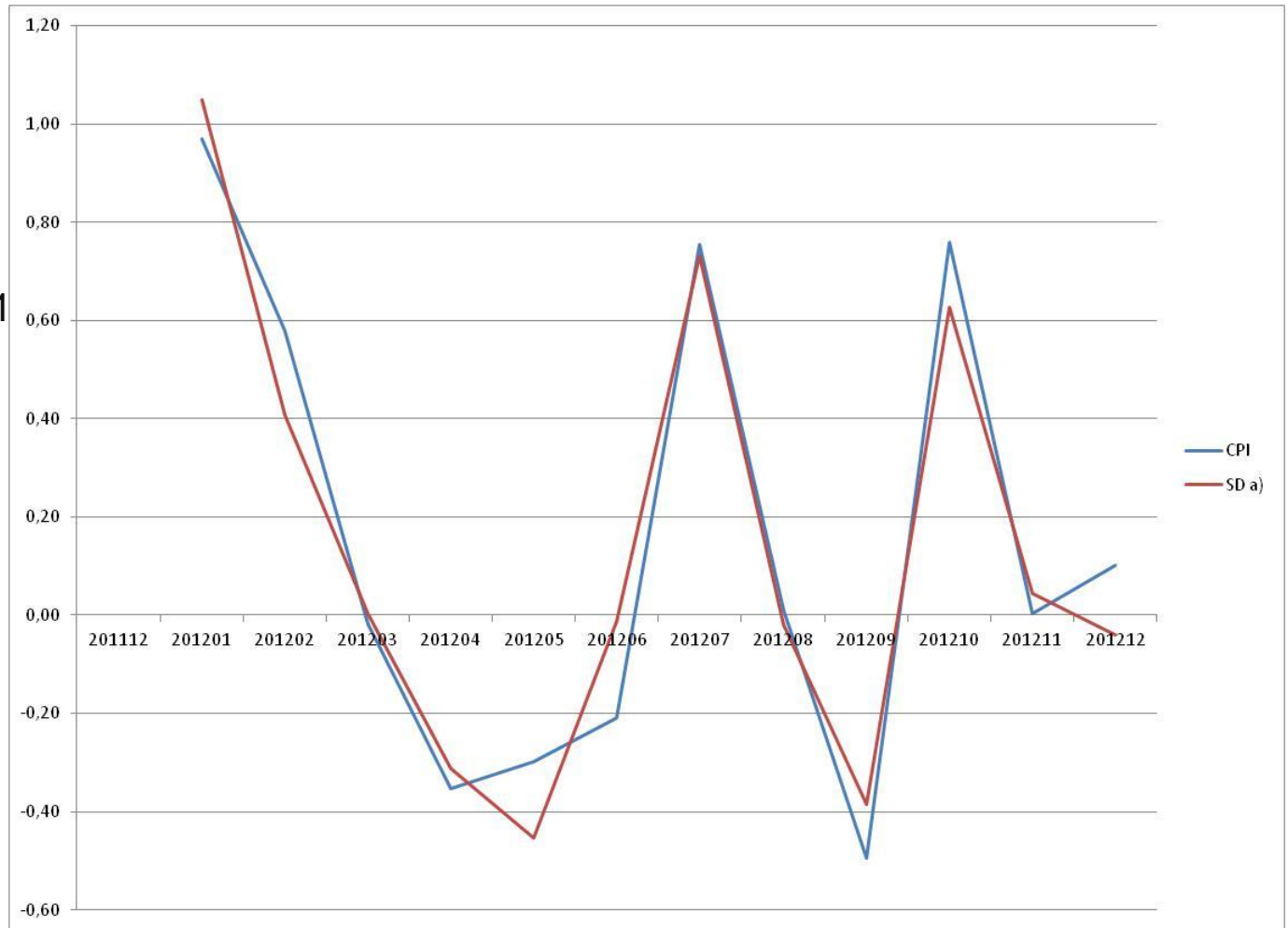
- Monthly rates
- Full COICOP 01
- a) Most sold variety, if priced for all months
- b) Most sold variety with prices for all months
- c) Geometric average of all available prices

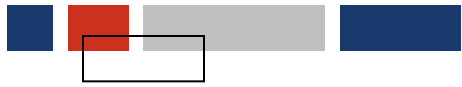




Study with actual data

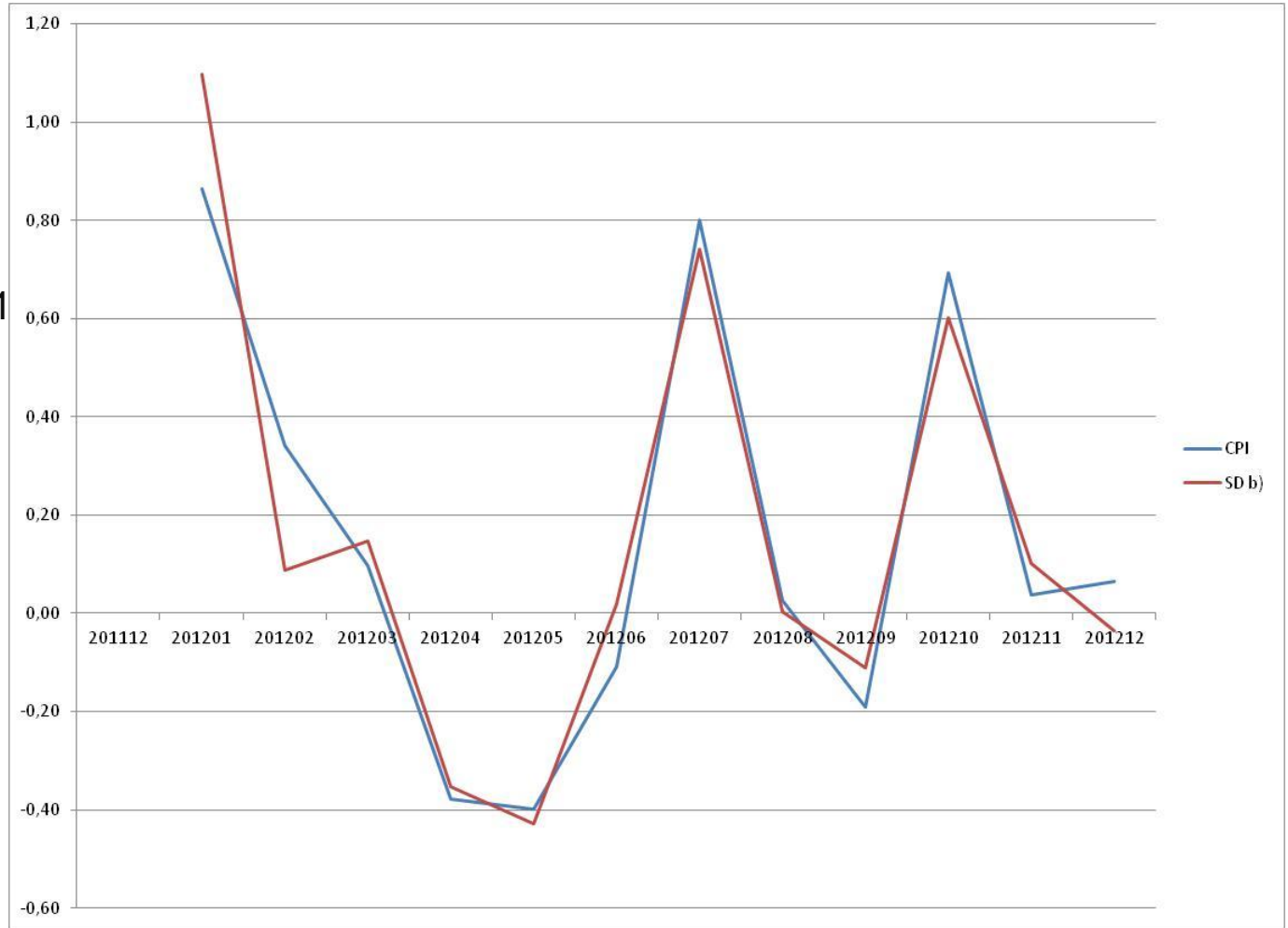
- Monthly rates
- Restricted COICOP 01
- a) Most sold variety, if priced for all months
- b) Most sold variety with prices for all months
- c) Geometric average of all available prices

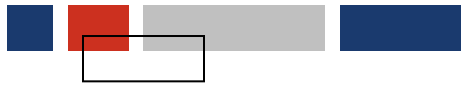




Study with actual data

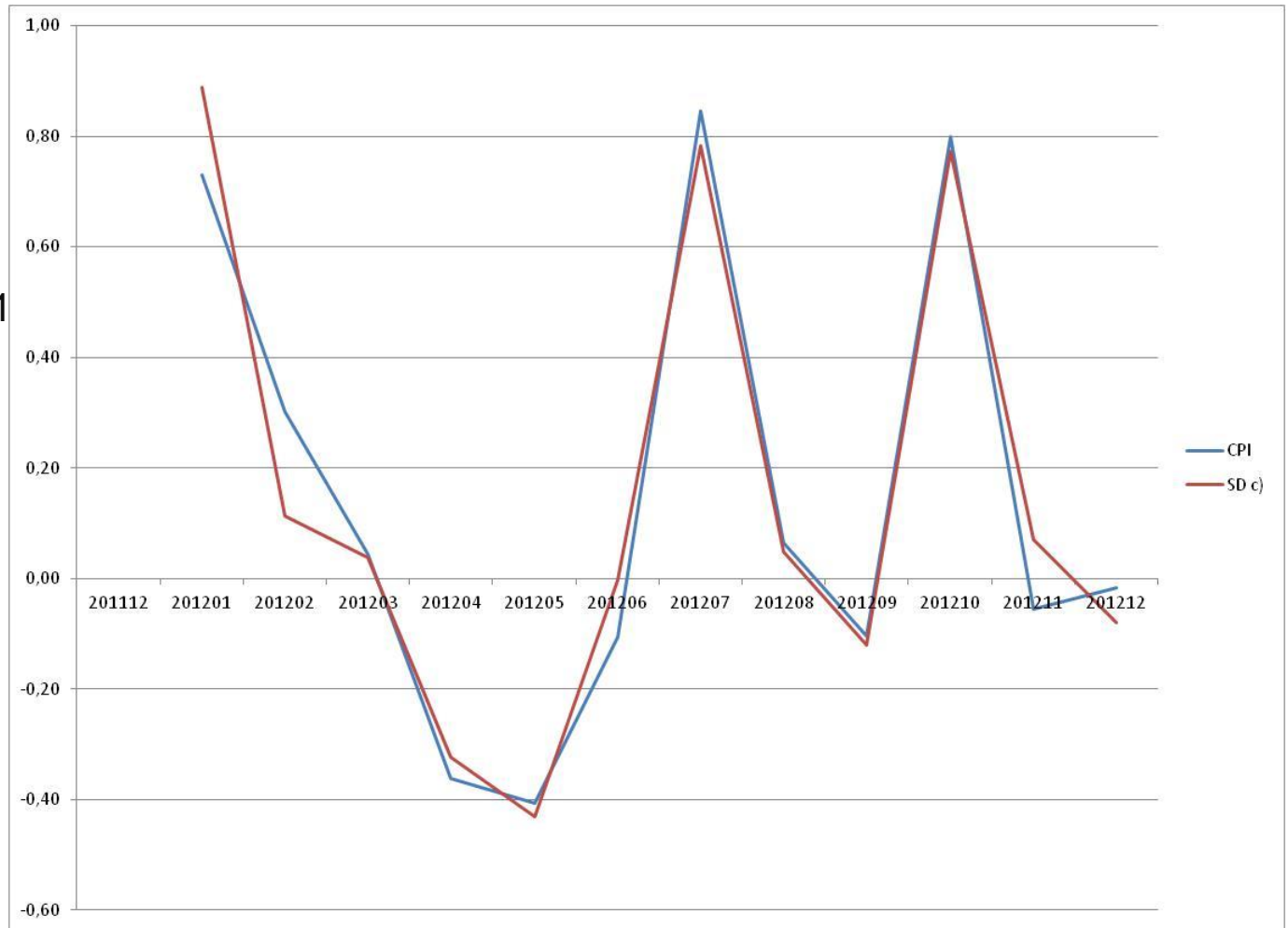
- Monthly rates
- Restricted COICOP 01
- a) Most sold variety, if priced for all months
- b) Most sold variety with prices for all months
- c) Geometric average of all available prices





Study with actual data

- Monthly rates
- Restricted COICOP 01
- a) Most sold variety, if priced for all months
- b) Most sold variety with prices for all months
- c) Geometric average of all available prices





Study with actual data

- Monthly rates differences
- COICOP 01
 - a) Most sold variety, if priced for all months
 - b) Most sold variety with prices for all months
 - c) Geometric average of all available prices

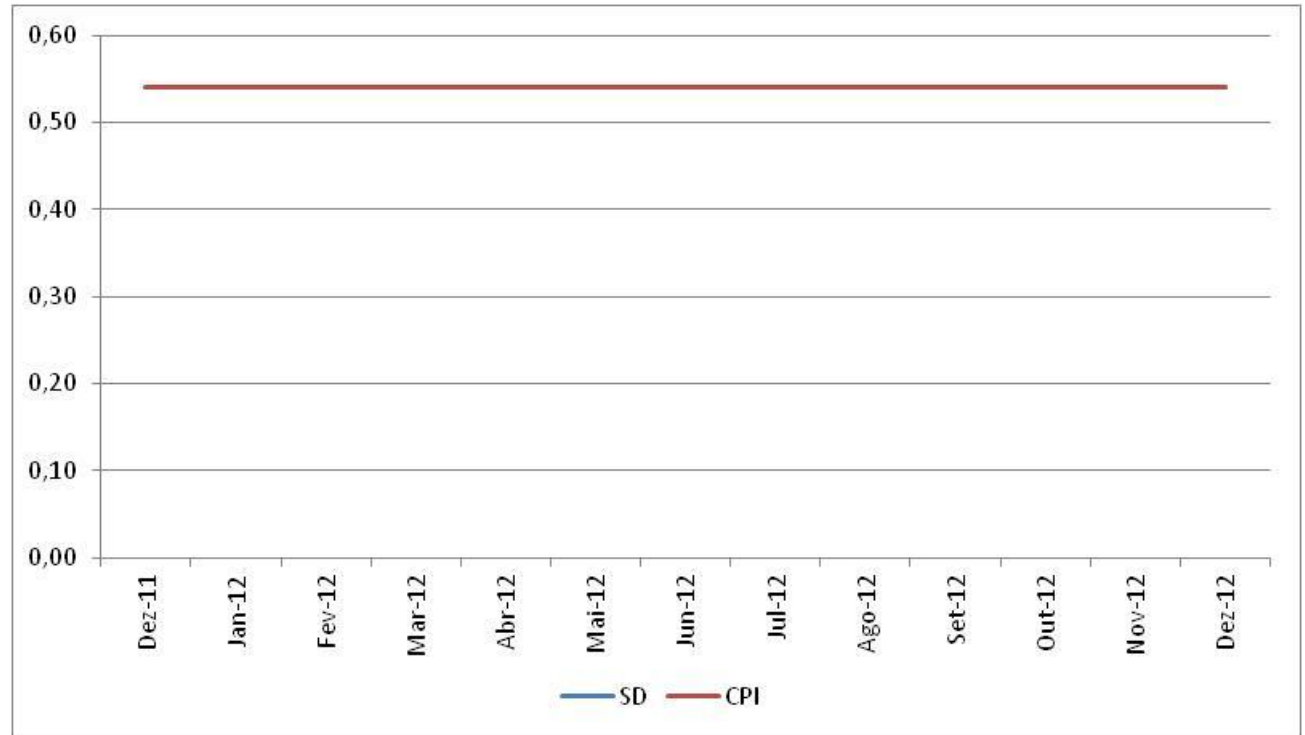
COICOP 01	Method	Max	Min	Average
Full	a)	0.12	-0.10	-0.01
Full	b)	0.17	-0.18	0.00
Full	c)	0.12	-0.14	0.00
Restricted	a)	0.20	-0.17	-0.01
Restricted	b)	0.23	-0.25	0.00
Restricted	c)	0.16	-0.19	0.00

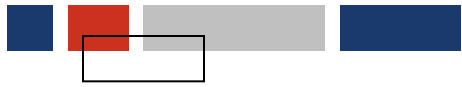




Study with actual data

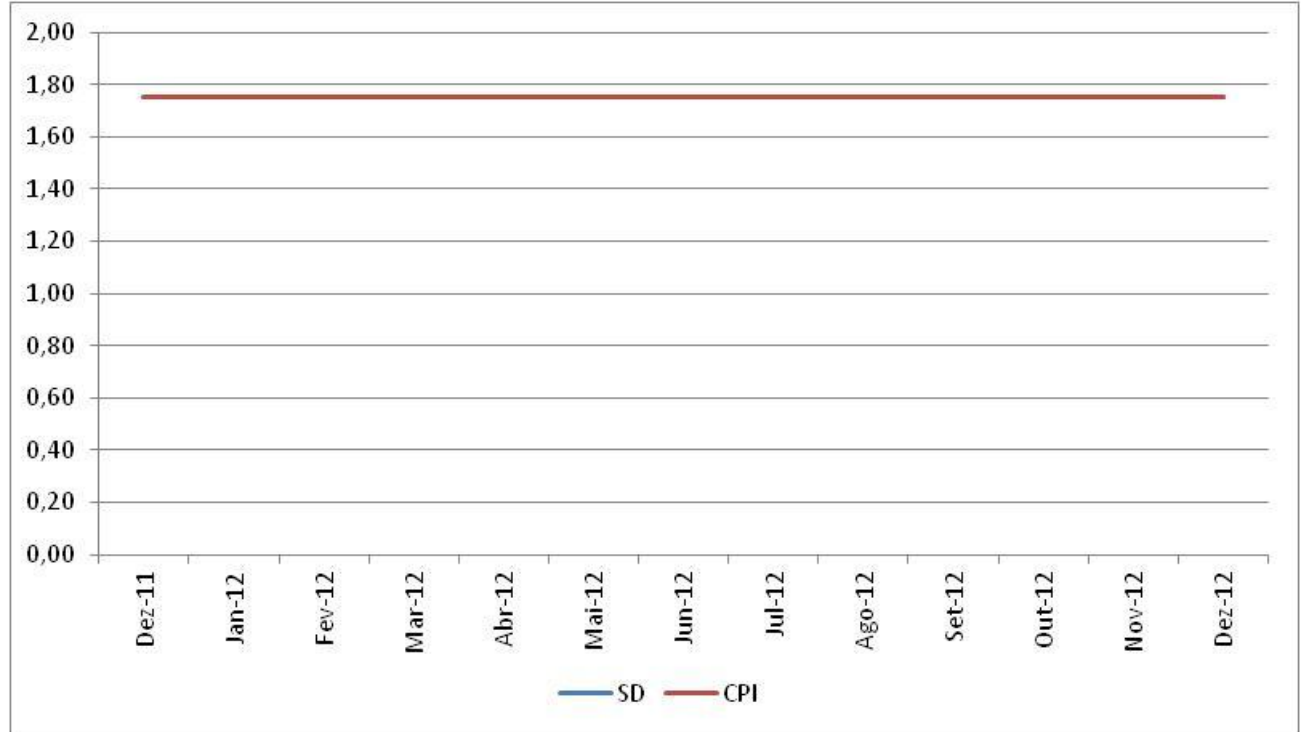
- Monthly prices
- Spaghetti
- Same variety
- Same price
- No problem

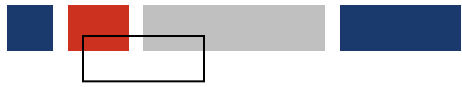




Study with actual data

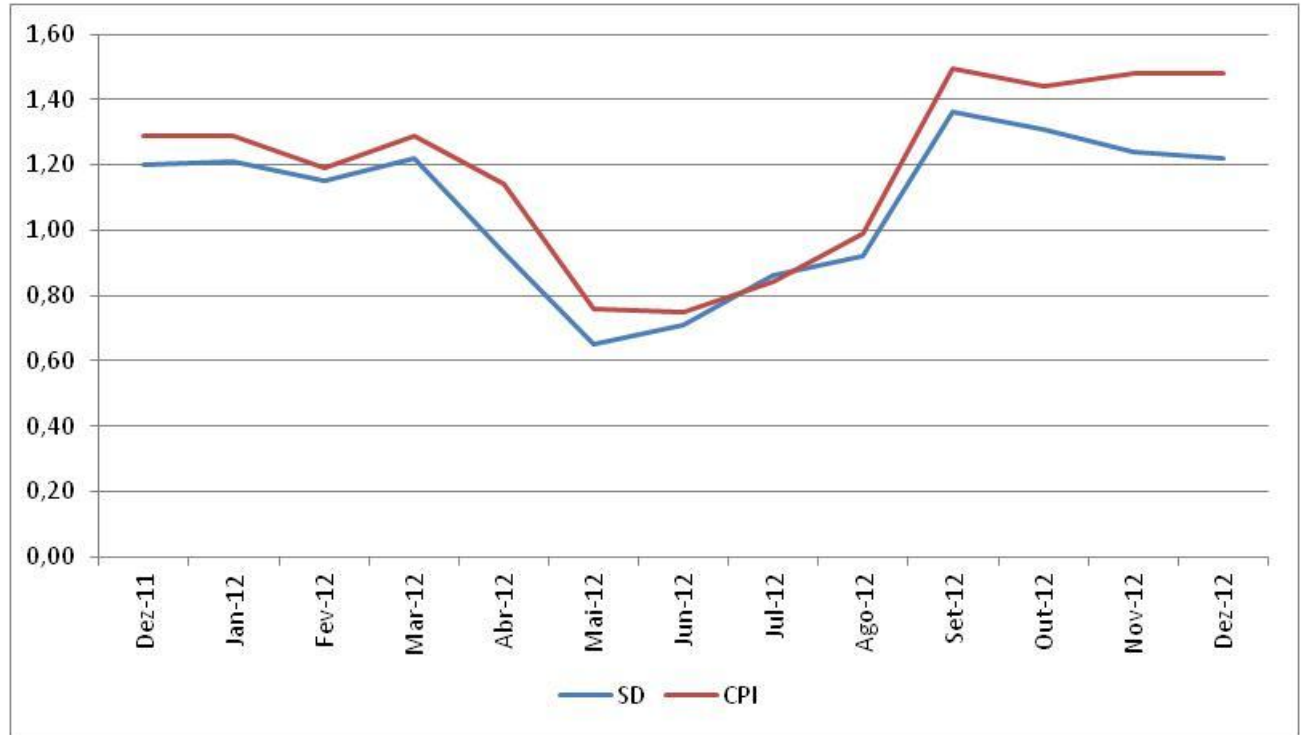
- Monthly prices
- Butter
- Same variety
- Same price
- No problem

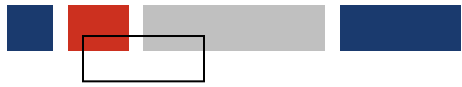




Study with actual data

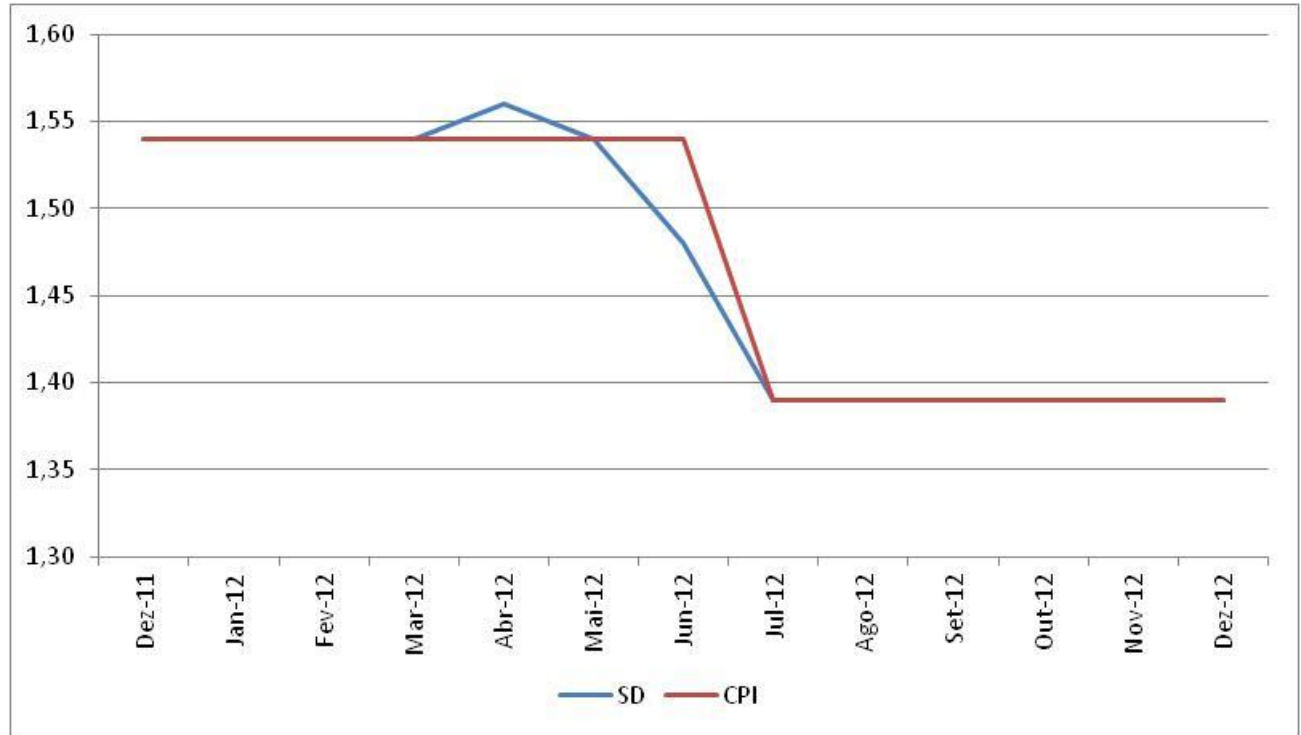
- Monthly prices
- Lemons
- Same variety
- Similar evolution
- Differences due to monthly SD prices
- No problem

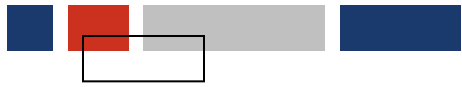




Study with actual data

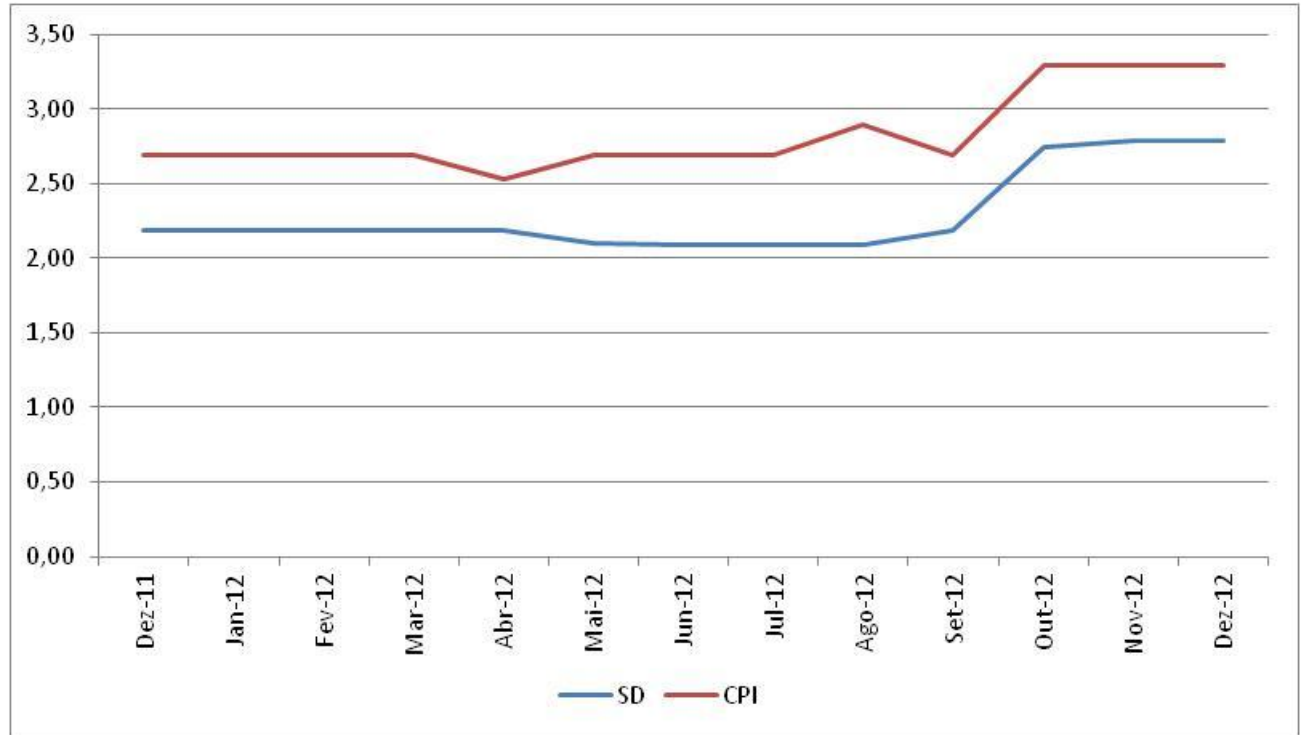
- Monthly prices
- Mayonnaise
- Same variety
- Similar evolution
- Differences due to monthly SD prices
- No problem

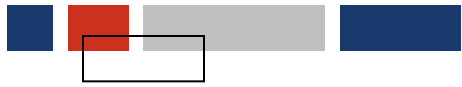




Study with actual data

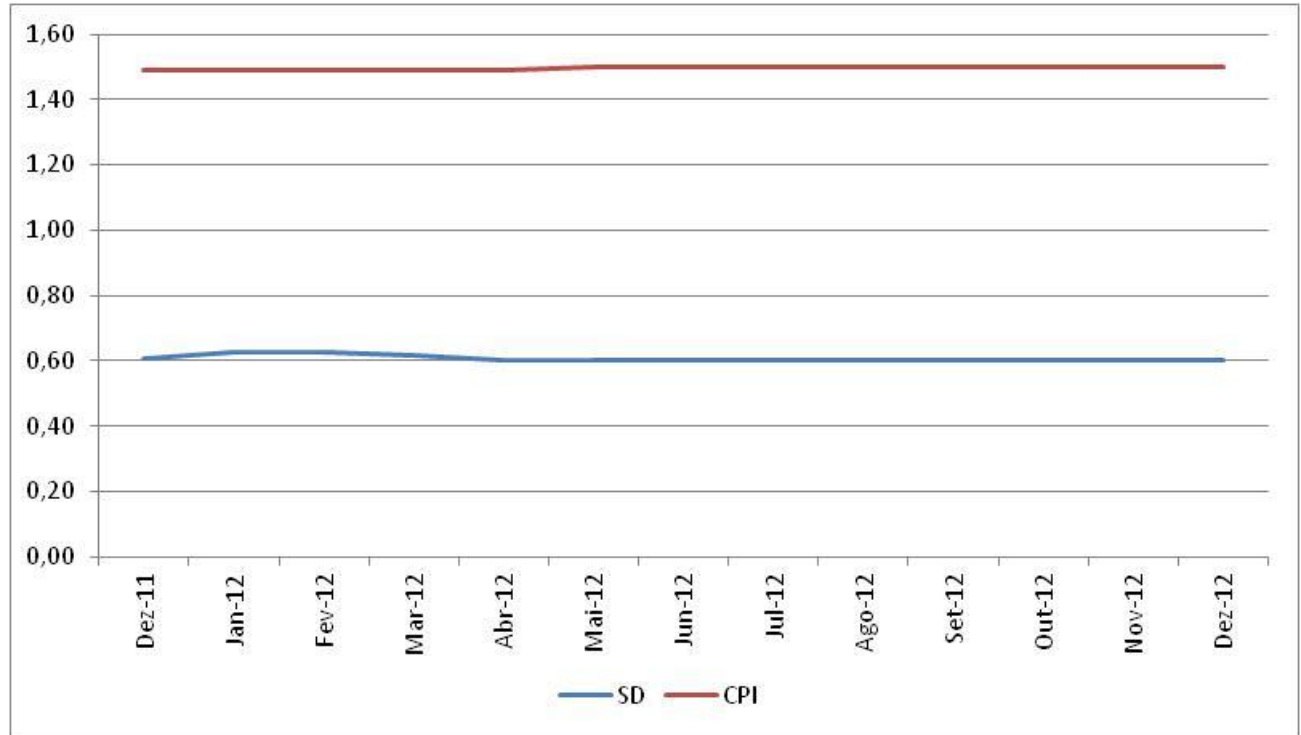
- Monthly prices
- Olive oil
- Different variety
- Similar evolution
- No problem

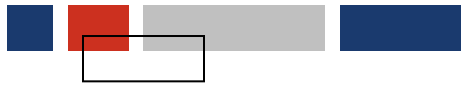




Study with actual data

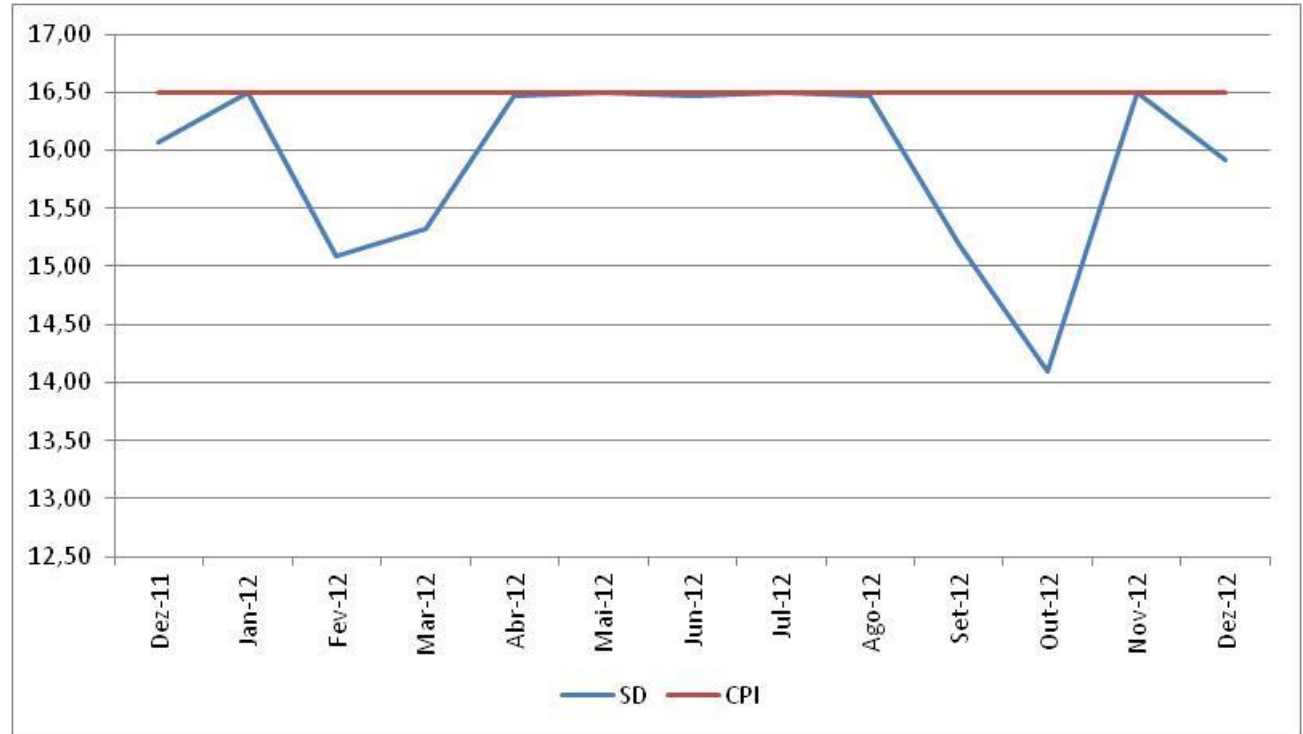
- Monthly prices
- Cookies
- Different variety
- Similar evolution
- No problem





Study with actual data

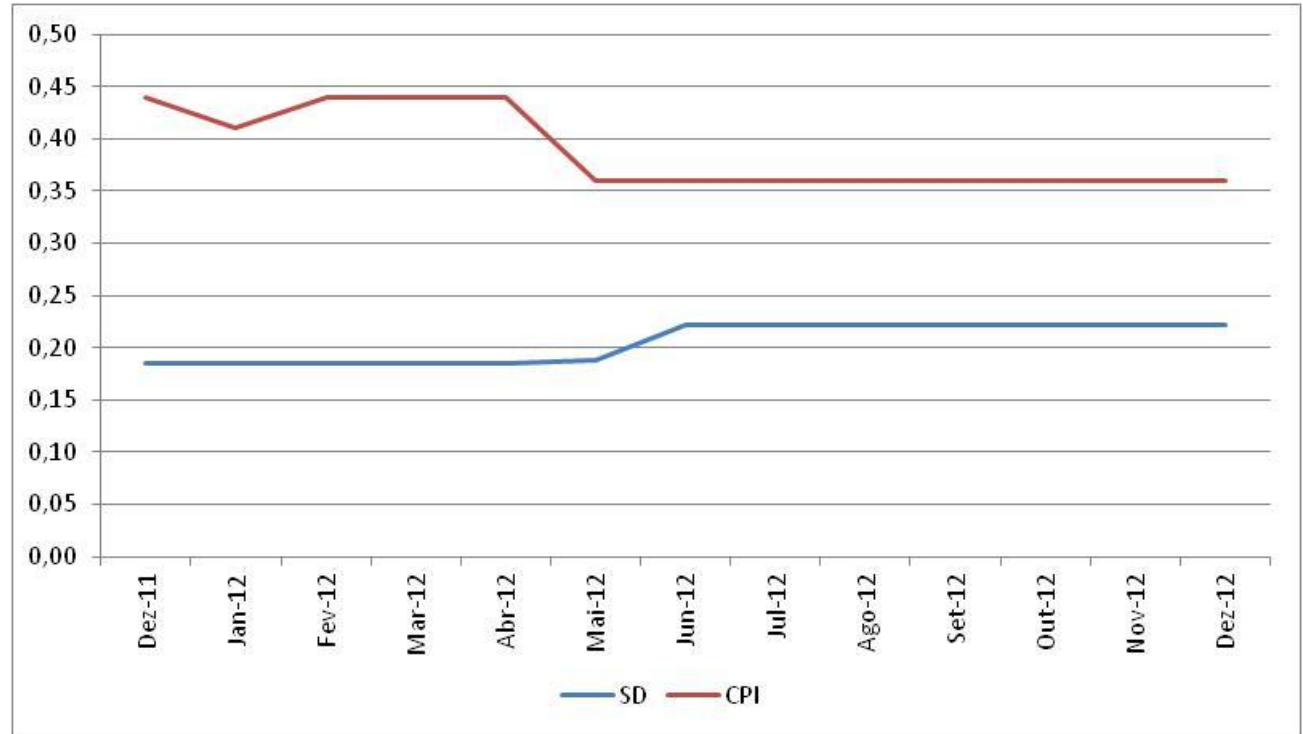
- Monthly prices
- Cheese
- Different variety
- Different evolution
- Annual rates would have been contaminated if scanner data was integrated in 12/2011
- Potential problem

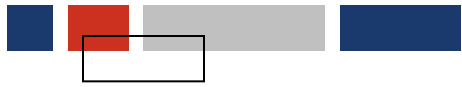




Study with actual data

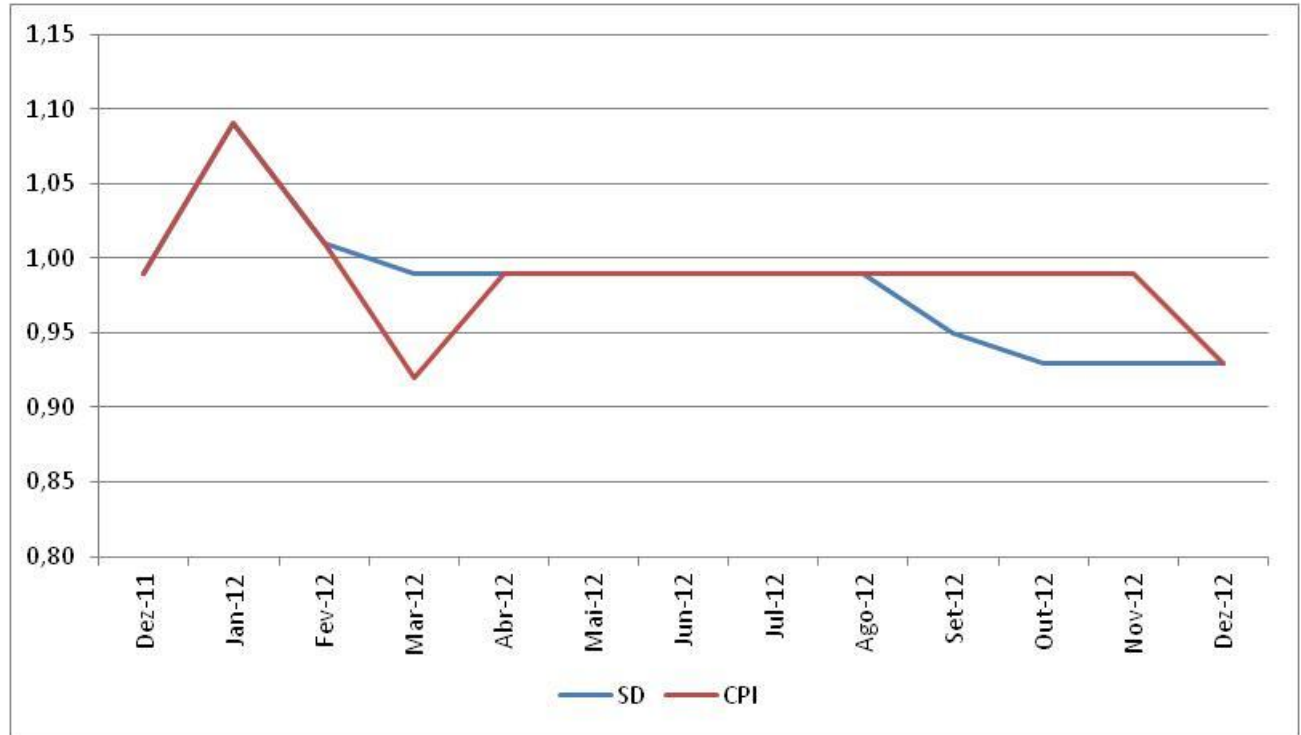
- Monthly prices
- Cookies
- Different variety
- Different evolution
- Annual rates would have been contaminated if scanner data was integrated in 12/2011
- Potential problem





Study with actual data

- Monthly prices
- Crab sticks
- Same variety
- Mostly same prices
- Problem in CPI price for 10/2012 & 11/2012?





Conclusions / further work



- Main conclusions:
 - Scanner Data can lead to quality improvements in the HICP
 - Integration of Scanner Data must be carefully implemented in order to avoid loss of quality and comparability
 - Chain-linking does not prevent potential impacts on annual rates





Conclusions / further work



- Specific topics for discussion:
 - What price?
 - What should be our “target” price?
 - Is there any evidence that considering a single price per month leads to bias?
 - Can we use SD as the only source for prices?
 - Big differences between countries
 - Some products need price collectors





Conclusions / further work



- Specific topics for discussion:
 - Should legislation be changed, leading to potentially lower comparability, because of SD?
 - Originally, legislation was produced in order to have a high level of comparability between countries, thus leading to restrictions that may produce “less than optimal” national HICPs
 - Giving more choices of alternate methods can decrease comparability



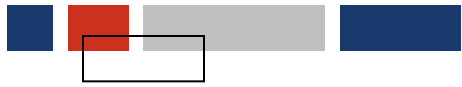


Conclusions / further work



- Specific topics for discussion:
 - Potential problems with proposed New Legal Framework: discriminatory discounts
 - Current data does not consider discriminatory discounts (only shelf price)
 - Need to request data in order to sample those discounts; also, any special discounts for business customers should be excluded
 - How will discriminatory discounts be accounted for stores that don't provide scanner data?
 - Question to NSIs that use unit values: why not use the shelf price, as unit values are currently not allowed?





Conclusions / further work



- Further work:
 - Complete study with more stores and providers
 - Test other pricing criteria:
 - Single price
 - Week's average
 - ...





Conclusions / further work



- Parallel work:
 - Use of administrative data for store weights
 - Statistics Portugal: IES (since 2007)
 - “Informação Empresarial Simplificada”
 - *Simplified Business Information*
 - Includes items such as turnover
 - Companies report directly *online*
 - Data is sent to Statistics Portugal regularly





SCANNERDATAPT
1 2 3 14 26 7 1 59 1 0 18 121 3



Thank you!



Vitor Mendonça
vitor.mendonca@ine.pt

Scanner Data Workshop
26|27 September 2013
Lisbon, Portugal



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL