Statistics Finland

Chain Error as a function of Seasonal Variation

Yrjö Vartia, Antti Suoperä, Kristiina Nieminen, Hannele Markkanen

Presenter: Kristiina Nieminen

Ottawa Group Meeting 8-10th May 2019

Rio de Janeiro

Content of the presentation

1. Introduction
2. Current calculation method for the scanner-datasets
3. Measuring seasonal variation
4. Measuring ChainError CE
5. Quadratic means of CE and seasonal components
6. Empirical results
7. Conclusion

Introduction

- Scanner-data implemented to the production of CPI since 2017
  - Pharmaceutical products
  - Alcoholic beverages
  - Food
- Method: base-based strategy and Törnqvist formula
- Challenge:
  - How to identify seasonal products in scanner-data?
  - Is there need to differentiate seasonal products according to the seasonal variation?
  - Is our current strategy applicable also for seasonal products?
- Main focus in the research

  Do the seasonal variation of values cause chain error to the index series constructed by chain strategy?

Current calculation method for the scanner-datasets

- The formula is Törnqvist

- Strategy is based on the following links

$$year(t-1) \rightarrow Year(t).m$$

  - The base period is the previous year normalized as a average month
  - The base is updated annually.
- On annual level our strategy is pure chain, but on monthly level it is pure base strategy.

- The price ratio is calculated by each item and aggregated directly to the EA level

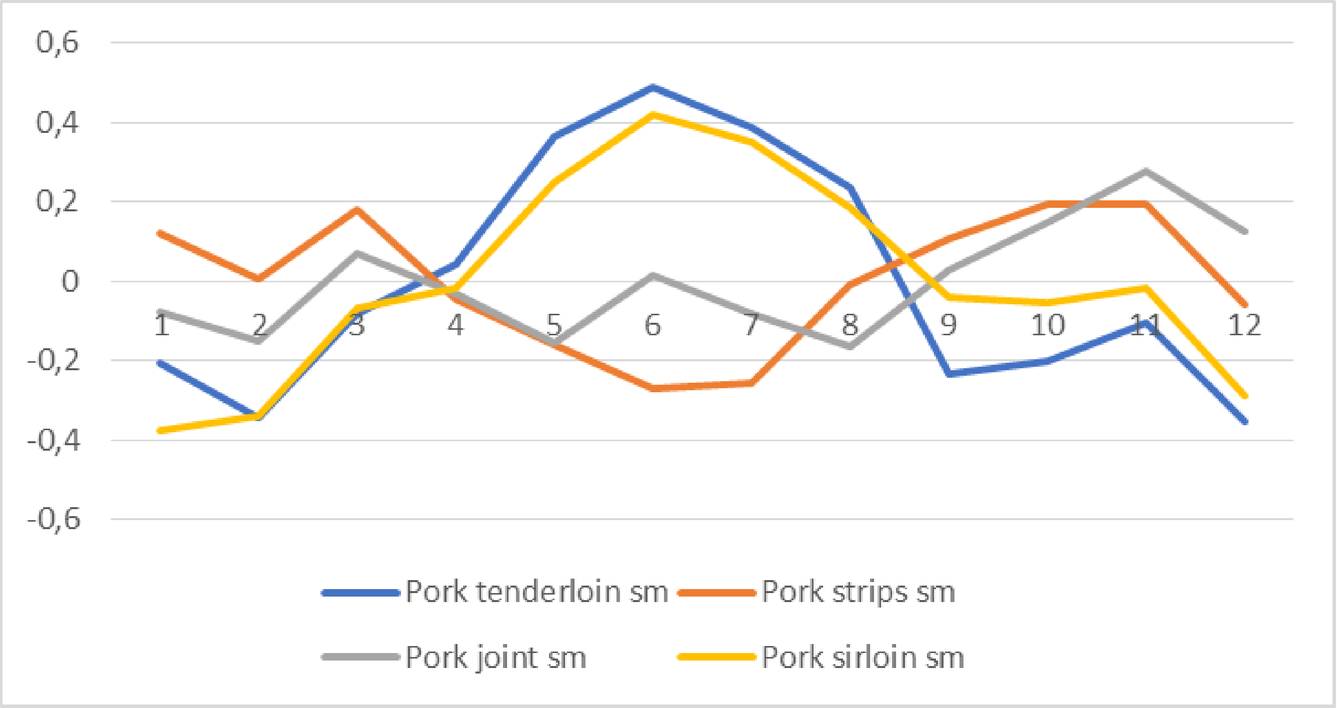
Measuring seasonal variation (>> seasonal index)

- Seasonal variation is calculated using logartihmic values ($y^{t.m} = logv^{t.m}$) and regression analysis by each sub-group (Ak)
- Variable to be explained is $y^{t.m}$ and eplanatory variables are time $(t = t.m)$ and its square and monthly dummies

		Month											
coicop7		1	2	3	4	5	6	7	8	9	10	11	12
Sweet pastry	s^m	-0,104	0,071	0,115	0,082	0,318	-0,008	-0,091	-0,060	-0,173	-0,052	0,074	-0,173
	t	-4,967	3,393	5,509	3,899	15,164	-0,376	-4,325	-2,875	-8,250	-2,459	3,540	-8,253
Crisp bread	s^m	-0,299	-0,219	-0,029	-0,138	0,072	0,104	0,095	0,056	-0,073	0,015	0,221	0,195
	t	-12,986	-9,519	-1,268	-6,014	3,128	4,531	4,135	2,431	-3,179	0,648	9,604	8,489
Filet of beef	s^m	-0,180	-0,296	-0,004	-0,044	0,095	0,224	0,134	0,113	-0,083	-0,092	0,124	0,009
	t	-4,618	-7,615	-0,099	-1,138	2,448	5,761	3,453	2,896	-2,137	-2,353	3,174	0,227
Beef strips	s^m	0,193	0,065	0,207	0,001	-0,155	-0,316	-0,341	-0,145	0,020	0,129	0,211	0,131
	t	6,812	2,303	7,294	0,039	-5,472	-11,145	-12,035	-5,122	0,708	4,538	7,444	4,636
Pork strips	s^m	0,121	0,005	0,179	-0,046	-0,160	-0,272	-0,257	-0,007	0,107	0,196	0,195	-0,061
	t	4,935	0,204	7,330	-1,903	-6,534	-11,118	-10,528	-0,287	4,382	8,042	7,986	-2,508
Pork sirloin	s^m	-0,375	-0,341	-0,070	-0,017	0,251	0,419	0,351	0,183	-0,040	-0,054	-0,017	-0,290
	t	-9,664	-8,783	-1,807	-0,430	6,456	10,794	9,041	4,715	-1,036	-1,394	-0,427	-7,465
Cucumber	s^m	0,109	0,042	0,117	-0,046	-0,052	-0,061	-0,109	0,122	-0,123	-0,109	0,136	-0,026
	t	4,727	1,816	5,085	-1,984	-2,267	-2,676	-4,758	5,306	-5,335	-4,726	5,941	-1,129

Measuring seasonal variation…

- Profile of the seasonal variation within year
  - Downward or upward increasing
  - Up- or downward concave curving
  - Shape of saw blade or some mix of them

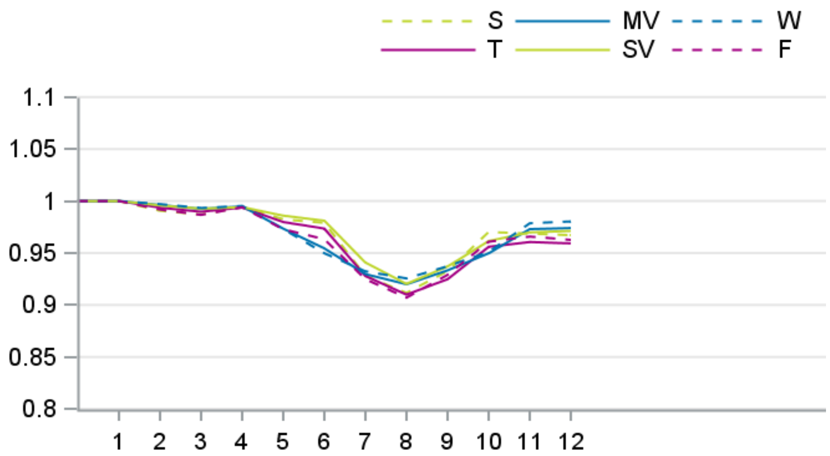


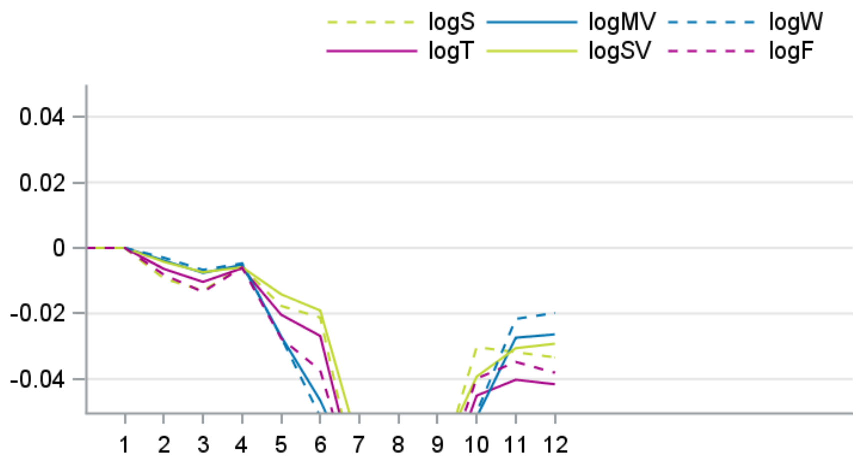
Measuring ChainError (CE)

Multi Period Identity Test MPIT

>> here its logarithmic difference

$$ChainError(P, Period) = (z^{t.m}) = \left(logP^{t.m}_{Base} - logP^{t.m}_{Chain}\right)$$





Quadratic means

- Seasonal Components and Chain Error varies around zero
  - How much do they deviate from zero ?
  - Is there mutual dependence between these two?
- Define Quadratic Means of them
  - Quadratic mean = root mean square
  - QM of Seasonal Components and
    QMof CE
  - for all subsets Ak

Quadratic Means: QM for ChainError CE

Use three simply steps to calculate QM of CE:

1. Square all signed log-differences over the whole period of 48 observations.
2. Calculate means of squared Chain Errors (=MS)
3. Finally take square root of MS to get QM of ChainError(P,Period)

$$QM\ of\ ChainError(P, Period) = QM(z^{t.m}) = \sqrt{\frac{1}{T}\sum_{t.m\ \epsilon\ Period}(z^{t.m})^2}.$$

$$QM\ of\ ChainError(P, Period) = QM\left(z^{t.m}\right) = \sqrt{\frac{1}{48}\left[\sum_{2015.m=1}^{12}(z^{2015.m})^2 + \cdots + \sum_{2018.m=1}^{12}(z^{2018.m})^2\right]}$$

Quadratic means: QM for Seasonal Components

- Quadratic means for all sub-groups (Ak) are calculated of the monthly indicators

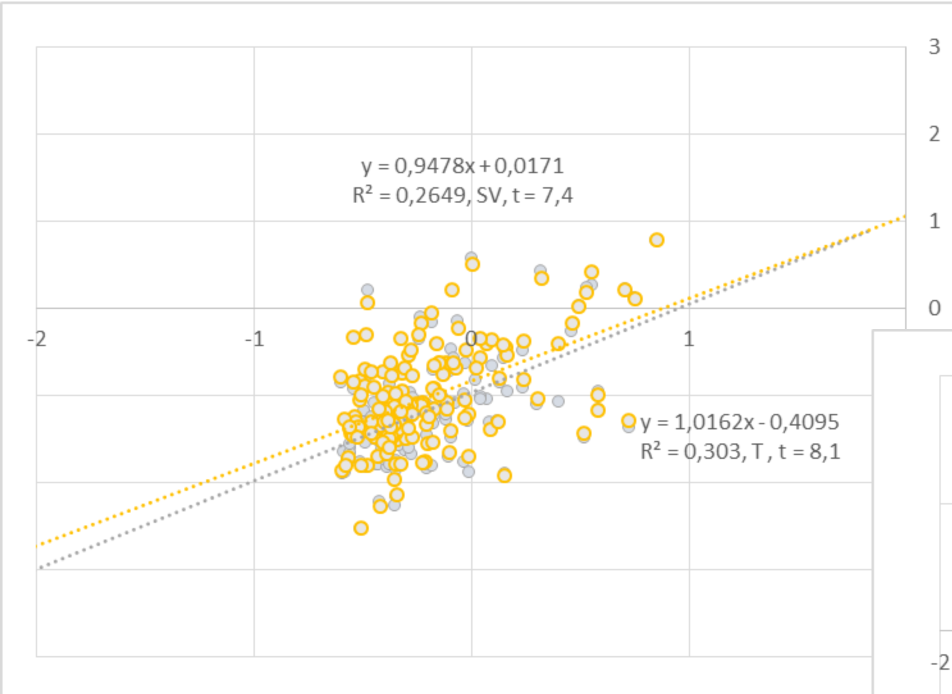
$$QM: (s^{t.m} = seasonal\ index\ \ expressed\ in\ log - values) =$$

$$QM: (Season, Ak, Period) = QM\left(s^{2015.m}\right) = \sqrt{\frac{1}{12}\left[\sum_{2015.m=1}^{12}(s^{2015.m})^{2}\right]}$$

Statistics Finland

Empirical results

Figures: logQM of Chain error according to logQM of Seasonal component



LINES:
Sato-Vartia=yellow, Törnqvist=grey,
Stuvel=blue

AXIS: log10

Conclusion

- According to our tests:
  - *logarithmic differences in the largeness of seasonal variation in the log-values of time series, clearly have average positive effects on the log-values of largeness of Chain Errors*
  - Construction strategies of index series based on chaining should be avoided
- Our strategy
  - removes all problems caused by chaining because it is drift-free
  - treats all months of every year *equally*
  - treats weakly, strongly seasonal and non-seasonal commodities totally symmetrically

Thank you all!

Statistics Finland

Kristiina Nieminen kristiina.nieminen@stat.fi
OttawaGroup meeting 8th – 10th May 2019