STATEC
Institut national de la statistique
et des études économiques

Luxembourg

Short-term movements of the RY-GEKS price index: is the failure at the identity test really a problem?

Claude Lamboray

Price statistics unit, STATEC
13, rue Erasme, B.P.304, L-2013 Luxembourg
claude.lamboray@statec.etat.lu
www.statec.lu

Paper presented at the Ottawa Group meeting in Copenhagen, Denmark
May 2013

Short-term movements of the RY-GEKS price index: is the failure at the identity test really a problem?

Claude Lamboray *

Abstract

The RY-GEKS price index is sometimes criticized because it fails the identity test. Indeed, if prices of all commodities are identical in the two comparison periods, then the corresponding price index should show no movement, regardless of the quantities observed. In this study we analyze to what extent the RY-GEKS price index violates this test. In particular we show that if prices remain constant in two consecutive periods, then the deviation from the identity depends on a weighted covariance between price relatives and quantity relatives. This means for instance that the problem can be attenuated if the variance of price relatives is low. Simulations on a real scanner data set show that this is typically the case in practice.

1 Introduction

The rolling year Gini, Eltetö & Köves, Szulc (in short RY-GEKS) price index has been proposed as a pragmatic solution for compiling price indices based on scanner data (see Ivancic et al. (2011) [6]). It is an extension of the GEKS method, using a 13-month moving window. By making multiple pairwise month-on-month comparisons, this method has the advantage of using a large amount of information contained in the scanner data set. Moreover, unlike more traditional chaining, empirical studies (for instance J. de Haan (2011) [4]) show that the so-called chain drift is much less apparent. Nevertheless, many NSIs are still reluctant to use this method in their production process. One of the reasons is that the RY-GEKS price index may exhibit short-term movements that can be difficult to explain to users.

In this paper we investigate the month-on-month change of the RY-GEKS price index from the perspective of the identity test: if in two adjacent periods all the prices are identical, then the price index should also be constant, regardless of the quantities observed in the comparison periods. The identity test is sometimes questioned because, according to economic theory, quantities are fixed given a set of prices. It is thus unexpected to observe identical prices but different quantities. Nevertheless, it is still useful to understand how the index behaves in such a stylized situation. Moreover, in the practical context of scanner data, there can be a significant proportion of prices that do not change, although quantities inevitably fluctuate from one month to another.

It is known that the RY-GEKS price index fails the identity test. That is why Ribe (2012) [7] sees the need to *assess the risk of possibly too unnatural behaviour of the index, due to the failure on the identity test.* Hence the goal of this study is to evaluate the extent of the problem, taking into account the particular structure of scanner data.

*claude.lamboray@statec.etat.lu, STATEC, 13, rue Erasme, B.P.304, L-2013 Luxembourg

2 Notations

A bilateral price index $P_{t',t}$ compares prices $p_s = (p_s^1, ..., p_s^N)$ and quantities $x_s = (x_s^1, ..., x_s^N)$ in two comparison periods ($s = t', t$):

$$P_{t',t} = P(p_{t'}, x_{t'}, p_t, x_t)$$

To simplify the analysis, we assume here that the set of N commodities stays constant over time. In practice this is not the case because there are in general new and disappearing items.

Let us assume that the price index satisfies the time reversal test, which means that $P_{t',t} = 1/P_{t,t'}$. In practice, the RY-GEKS price index is based either on the Fisher price index or on the Tornqvist price index that both satisfy the time reversal test. We will concentrate on the former. The Fisher price index is defined as the square root of the Laspeyres price index and the Paasche price index:

$$P_{t',t} = \sqrt{\frac{p_t \cdot x_{t'}}{p_{t'} \cdot x_{t'}} \frac{p_t \cdot x_t}{p_{t'} \cdot x_t}}$$

The RY-GEKS price index in t with respect to $t-1$ can be expressed as follows (see De Haan (2011) [3]):

$$RYGEKS_{t-1,t} = \left(\frac{P_{t-12,t}}{P_{t-12,t-1}} \cdot \frac{P_{t-11,t}}{P_{t-11,t-1}} \cdot \ldots \cdot \frac{P_{t-2,t}}{P_{t-2,t-1}} \cdot (P_{t-1,t})^2 \right)^{\frac{1}{13}} \qquad (\alpha)$$

Let us assume that prices in t and in $t-1$ are identical. According to the identity test, we would like the price index to be equal to 1. The difference in percentage points from the identity is thus equivalent to the month-on-month change:

$$RYGEKS_{t-1,t} - 1 \quad if \quad p_t = p_{t-1}$$

Note that this is a weaker form of the identity test. We only consider two consecutive periods, whereas usually this test should hold for any period compared to the base period.

3 An axiomatic perspective

The RY-GEKS price index is based on an underlying bilateral price index. We assume that this bilateral price index satisfies the following tests (see Chapter 16 in ILO et al. (2004) [5]):

- Identity test (T3):

$$P(p, x_{t'}, p, x_t) = 1$$

- Proportionality in current prices (T5):

$$P(p_{t'}, x_{t'}, \lambda \cdot p_t, x_t) = \lambda \cdot P(p_{t'}, x_{t'}, p_t, x_t) \quad \forall \lambda > 0$$

- Invariance to proportional changes in current quantities (T7):

$$P(p_{t'}, x_{t'}, p_t, \lambda \cdot x_t) = P(p_{t'}, x_{t'}, p_t, x_t) \quad \forall \lambda > 0$$

The Fisher price index satisfies these three basic tests. It is now possible to highlight two special cases where the identity test holds for the RY-GEKS.

1. *Constant price relatives*
   If $\frac{p^i_t}{p^i_{t-m}} = \lambda_m, (\ \forall i = 1, \ldots, N, \forall m = 2, \ldots, 12)$ and if the bilateral price index satisfies T3 and T5, then:
   $$RYGEKS_{t-1,t} = 1 \quad if \quad p_t = p_{t-1}$$
2. *Constant quantity relatives*
   If $\frac{x^i_t}{x^i_{t-1}} = \lambda, (\ \forall i = 1, \ldots N)$ and if the bilateral price index satisfies T3 and T7, then:
   $$RYGEKS_{t-1,t} = 1 \quad if \quad p_t = p_{t-1}$$

The first case means that if the price for each commodity moves up or down by the same rate between $t - m$ and t, then the identity test will hold for the RY-GEKS price index between $t - 1$ and t. In practice, this assumption is fairly reasonable although of course in a real scanner data set, not all prices move at exactly the same rate. This also covers the particular situation in which all prices stay constant during the entire period of 13 months ($p^i_{t-12} = p^i_{t-11} = \ldots = p^i_t, \forall i$).

In the second case, we assume that quantities for each commodity fluctuate by the same rate between $t - 1$ and t, although prices are identical in these two periods. Such a situation may happen if for instance there is a sudden increase (or decrease) in demand, although the demand structure stays the same. The second case also includes the trivial situation where quantities are constant ($x^i_{t-1} = x^i_t, \forall i = 1, \ldots, N$).

4 An analytical perspective

In order to get a deeper understanding of how the RY-GEKS price index behaves, we derive an analytical expression of the change observed with the RY-GEKS if prices stay constant. To do so, we first introduce the notation of a weighted relative covariance. Let a_i and b_i be two evaluations of observation i and let w_i be the weight associated with that observation so that $\sum_i w_i = 1$. The weighted relative covariance is then defined as the weighted covariance between a and b, divided by the product of the weighted average of a and the weighted average of b:

$$RelCov(a_i, b_i, w_i) = \frac{\sum_i w_i (a_i - \sum_j w_j a_j)(b_i - \sum_j w_j b_j)}{\left(\sum_i w_i a_i\right)\left(\sum_i w_i b_i\right)}$$

The relative variance is obtained by simply taking the relative covariance between two same variables: $RelVar(a_i, w_i) = RelCov(a_i, a_i, w_i)$.

Let us suppose that $p_{t-1} = p_t$. In the expression (α) provided in section 2, the last term $(P_{t-1,t})^2$ equals 1 because the Fisher price index satisfies the identity test. For the 11 first terms, following the decomposition introduced by Bortkiewicz [2], it can be shown that:

$$\frac{P_{t-m,t}}{P_{t-m,t-1}} = \left(\frac{\frac{p_t \cdot x_t}{p_{t-m} \cdot x_t}}{\frac{p_t \cdot x_{t-1}}{p_{t-m} \cdot x_{t-1}}} \right)^{\frac{1}{2}} = \left(1 + RelCov\left(\frac{p^i_t}{p^i_{t-m}}, \frac{x^i_t}{x^i_{t-1}}, \frac{p^i_{t-m} x^i_{t-1}}{\sum_i p^i_{t-m} x^i_{t-1}} \right) \right)^{\frac{1}{2}}$$

Consequently, the deviation from the identity in percentage points of the RY-GEKS price index will be:

$$RYGEKS_{t-1,t} - 1 = \prod_{m=2}^{12} \left(1 + RelCov\left(\frac{p_t^i}{p_{t-m}^i}, \frac{x_t^i}{x_{t-1}^i}, \frac{p_{t-m}^i x_{t-1}^i}{\sum_i p_{t-m}^i x_{t-1}^i}\right)\right)^{\frac{1}{26}} - 1$$

In other words, the monthly change depends on the (weighted) covariance between price relatives (period t compared to period $t-m$) and quantity relatives (period t compared to period $t-1$). It is also interesting to notice that the quantities of $t-2,..,t-12$ do not matter. However, the prices from all the 13 periods influence the result. The weights are expenditure shares where prices of $t-m$ are mixed with quantities of $t-1$.

Note that the relative covariance between two variables is, in absolute terms, less than the square root of the relative variance of the two variables taken separately (Cauchy inequality). Using our notations, we thus have:

$$\left(RelCov\left(\frac{p_t^i}{p_{t-m}^i}, \frac{x_t^i}{x_{t-1}^i}, w_{t-m}^i\right)\right)^2 \leq RelVar\left(\frac{p_t^i}{p_{t-m}^i}, w_{t-m}^i\right) RelVar\left(\frac{x_t^i}{x_{t-1}^i}, w_{t-m}^i\right) \quad (\beta)$$

The two particular cases highlighted in the previous section can be easily understood now. Constant price relatives imply that the variance of the price relatives is zero and consequently the covariance is zero as well. The same holds if quantity relatives between $t-1$ and t are identical for all commodities.

We will see in the next section that under normal circumstances, the relative variance of price relatives is low, and consequently the monthly change of the RY-GEKS price index is in practice also rather small. Moreover, if we assume that constant prices $p_t = p_{t-1}$ imply almost identical quantities $x_t \approx x_{t-1}$, then the relative variance of quantity relatives will be close to zero and consequently the failure at the identity test will also be very limited.

The unit value price index does not satisfy the identity test either. Using for instance the formulation presented by B. Balk (1998) [1], it can be shown that , then:

$$UV_{t-1,t} - 1 = RelCov\left(p_t^i, \frac{x_t^i}{x_{t-1}^i}, \frac{x_{t-1}^i}{\sum_i x_{t-1}^i}\right), \quad if \;\; p_t = p_{t-1}$$

Hence, for the unit value price index, the monthly change under constant prices depends on the (weighted) covariance between price levels and quantity relatives. It is known that the unit value should preferably be used over "homogeneous" products. In such a case, price levels tend to be similar, and consequently the monthly change should be not too large because of a low variance of price levels.

With this analytical insight, it is easy to construct a theoretical example where the failure at the identity test is particularly obvious (see Table 1). There are two products with stable prices over several periods. In period $t-1$, item 1 gets significantly cheaper whereas item 2 gets significantly more expensive. The corresponding shift in quantities is only observed with a time lag in period t. The prices between $t-1$ and t are identical. However, the RY-GEKS price index decreases by 75%. This is due to the large variance of the price and quantity ratios. Note that in this example, the error made by the unit value index is even larger, with a decrease of 80% between periods $t-1$ and t.

Period	p^1	p^2	x^1	x^2
1	100	10	10	100
...				
$t-2$	100	10	10	100
$t-1$	10	100	10	100
t	10	100	100	10

Table 1: A theoretical example where the identity test fails.

5 An empirical perspective

We have performed various simulations using a scanner data set available at STATEC with information from a single shop of a local supermarket chain [1]. In order to quantify the failure at the identity test, we create an artificial dataset by simply copying for each EAN code prices from month $t-1$ (in this case September 2012) to month t (in this case October 2012). Alternatively, prices from month t (October 2012) can also be assigned to month $t-1$ (September 2012) in order to have identical prices in $t-1$ and t. In both cases, quantities are left unchanged, which is a rather strong assumption. Nevertheless, this can be seen as a stress test on the data with respect to the identity test.

Price indices have been estimated for different elementary levels of the Luxembourg CPI. We concentrated on those sections of the first COICOP division with a sufficient number of EAN codes. Only those EAN codes have been kept that are available in every month of the period covered in the simulations. There are on average 158 EAN codes per section.

Results can be found in Tables 2 and 3. The last column depicts the month-on-month change obtained with the RY-GESK price index if prices in $t-1$ and in t are identical. If the identity test is satisfied, then the month-on-month change would be 0%. As we have seen in the previous section, the change for the RY-GEKS depends on the weighted covariance between price relatives and quantity relatives. The third column presents the weighted relative variance of the quantity ratios (average value of $RelVar\left(\frac{q_t^i}{q_{t-1}^i}, w_{t-m}^i\right)$, $m = 2\ldots12$) and the fourth column presents the weighted relative variance of the price ratios (average value of $RelVar\left(\frac{p_t^i}{p_{t-m}^i}, w_{t-m}^i\right)$, $m = 2\ldots12$).

The following conclusions can be drawn from these simulations:

- In general, the movements made by the RY-GEKS price index are rather small, but in some cases they can nevertheless be significant. On average between two consecutive months, the RY-GEKS stands at 0.43% if prices in $t-1$ are copied to t and it stands at -1.12% if prices in t are copied to $t-1$.
- The relative variance of price relatives is rather low (0.67% on average in the first case and 0.79% in the second case). It is precisely this feature that allows the covariance between price relatives and quantity relatives, and consequently the monthly change to be low as well (see Cauchy inequality (β)).

[1] Scanner data is available from January 2012 onwards. In order to have 13 months of data, October 2011, November 2011 and December 2011 have been replaced by January 2012 data.

Section	Nbr EAN	RelVar quant.	RelVar prices	**Change**
Rice	61	42.29%	0.44%	**0.55%**
Bread	96	10.17%	0.65%	**0.31%**
Biscuits	310	64.01%	0.52%	**0.44%**
Pasta	291	65.75%	0.76%	**0.58%**
Pastry	52	26.58%	1.06%	**1.00%**
Cakes and pies	87	80.42%	0.39%	**-0.55%**
Pizzas	55	313.12%	1.50%	**1.50%**
Cereals	209	72.49%	0.32%	**0.36%**
Sausages	68	87.81%	2.12%	**2.49%**
Ham	52	8.54%	0.36%	**0.14%**
Yogurt	287	15.13%	0.43%	**0.17%**
Cheese	368	41.92%	0.29%	**0.27%**
Other dairy products	157	8.83%	0.16%	**0.00%**
Dried fruits and nuts	123	45.64%	0.48%	**0.34%**
Canned vegetables	137	246.91%	0.26%	**0.67%**
Jam	117	31.63%	0.33%	**0.42%**
Chocolate	199	31.33%	0.35%	**0.33%**
Candies	258	53.78%	0.25%	**0.12%**
Ice cream	95	60.82%	0.35%	**0.29%**
Salt, pepper and spices	136	40.91%	0.83%	**-0.02%**
Coffee	165	82.88%	0.86%	**1.02%**
Tea	82	63.94%	0.50%	**0.50%**
Mineral water	103	38.88%	2.67%	**-0.80%**
Sodas	225	25.32%	0.26%	**0.10%**
Fruit juices	225	100.07%	0.68%	**0.49%**
AVERAGE	158	66.37%	0.67%	**0.43%**

Table 2: Change of the RY-GEKS price index in t compared to $t-1$, prices in t are replaced by prices in $t-1$

Section	Nbr EAN	RelVar quant.	RelVar prices	**Change**
Rice	61	42.29%	0.43%	**-0.51%**
Bread	96	10.17%	0.36%	**-0.30%**
Biscuits	310	64.01%	0.40%	**-0.77%**
Pasta	291	65.75%	0.72%	**-1.11%**
Pastry	52	26.58%	0.44%	**-0.36%**
Cakes and pies	87	80.42%	0.57%	**-0.83%**
Pizzas	55	313.12%	1.18%	**-4.93%**
Cereals	209	72.49%	0.22%	**-0.23%**
Sausages	68	87.81%	2.26%	**-3.82%**
Ham	52	8.54%	0.28%	**0.02%**
Yogurt	287	15.13%	0.46%	**-0.37%**
Cheese	368	41.92%	0.31%	**-0.78%**
Other dairy products	157	8.83%	0.26%	**-0.21%**
Dried fruits and nuts	123	45.64%	0.33%	**-0.06%**
Canned vegetables	137	246.91%	0.65%	**-3.12%**
Jam	117	31.63%	0.28%	**0.06%**
Chocolate	199	31.33%	0.20%	**-0.16%**
Candies	258	53.78%	0.34%	**-0.72%**
Ice cream	95	60.82%	0.50%	**-0.75%**
Salt, pepper and spices	136	40.91%	0.82%	**-0.07%**
Coffee	165	82.88%	1.53%	**-2.39%**
Tea	82	63.94%	0.46%	**-1.21%**
Mineral water	103	38.88%	5.76%	**-3.22%**
Sodas	225	25.32%	0.25%	**-0.02%**
Fruit juices	225	100.07%	0.87%	**-2.24%**
AVERAGE	158	66.37%	0.79%	**-1.12%**

Table 3: Change of the RY-GEKS price index in t compared to $t-1$, prices in $t-1$ are replaced by prices in t

- The relative variance of quantity relatives is rather high (66.37% on average). This confirms findings in other studies (see for instance Ribe (2012) [7]) that quantities can fluctuate a lot in scanner data. However, this is counterbalanced by the low variance of price relatives.

- The issue can become more serious if the variance of price relatives and/or of quantity relatives is for some reasons too large. From an empirical perspective, a relative variance of price relatives larger than 2% (sausages, mineral water) and/or a relative variance of quantity relatives larger than 200% (pizzas, canned vegetables) can lead to a significant violation of the identity test.

- In our simulations, the change is in general positive if prices in t are replaced by prices in $t-1$ and negative if prices in $t-1$ are replaced by prices in t. In fact price change and quantity change tend to be negatively correlated. In other words, if $\frac{x_t}{x_{t-1}}$ is low, then this suggests that p_{t-1} has been a low price, such as for instance a special offer. Consequently, the comparison between p_{t-1} and p_{t-m} will be be on average low as well. This explains the positive correlation if prices in $t-1$ are used both in $t-1$ and t. In a similar way, if $\frac{x_t}{x_{t-1}}$ is high, then probably p_t is low and consequently the comparison between p_t and p_{t-m} will be be on average low as well. This explains the negative correlation in the second scenario.

- It is not appropriate to apply a unit value price index on a whole product group. In any case, in these simulations, the unit value price index performs even worse with respect to the identity test. On average (in absolute terms), over the 25 product sections considered here, the unit value index deviates by 3.7 percentage points in the first case ($p_t \leftarrow p_{t-1}$) and by 4.5 percentage points in the second case ($p_{t-1} \leftarrow p_t$).

6 Conclusion

We have examined the failure at the identity test restricted to two consecutive periods for the RY-GEKS price index. In fact, the deviation from the identity depends on the covariance between price change and quantity change. In practice, the variance of price relatives is limited and consequently the movement made by the RY-GEKS is in general small. This contrasts with the unit value price index for which the deviation based on the covariance between price levels and quantity relatives can be substantially larger.

Problems may arise with the RY-GEKS price index if there is a high variance of price relatives or of quantity relatives. This may be the case for instance for seasonal products. However, for such volatile products, it is rather unlikely that between two consecutive months prices stay constant for all products, which is the underlying assumption of the identity test that is analyzed here.

References

[1] B. Balk (1998). On the use of unit value indices as consumer price subindices. Ottawa Group Meeting in Washington.

[2] L. Bortkiewicz (1923). Zweck und Struktur einer Preisindexzahl, Nordisk Statistik Tidsskrift.

[3] H. van der Grient, J. de Haan (2011). Scanner Data Price Indexes: The Dutch Method versus Rolling Year GEKS. Ottawa Group Meeting in Wellington.

[4] J. de Haan, H. van der Grient (2011). Eliminating chain drift in price indexes based on scanner data. Journal of Econometrics 161, 36-46.

[5] ILO/IMF/OECD/UNECE/Eurostat/The World Bank (2004). Consumer Price Index Manaual: Theory and Practice. Geneva, International Labour Office.

[6] L. Ivancic, W.E. Diewert, K.J. Fox (2011). Scanner data, time aggregation and the construction of price indexes. Journal of Econometrics 161, 24-35.

[7] M. Ribe (2012). Some properties of the RGEKS index for scanner data. Scanner data workshop Statistics Sweden, Stockholm.