Eliminating Chain Drift in Price Indexes Based on Scanner Data

Jan de Haan[a] and Heymerik van der Grient[b]

16 April 2009

Abstract: The use of scanner data in the CPI makes it possible to compile superlative price indexes at detailed aggregation levels since prices and quantities are available. A potential drawback is the high attrition rate of items. The usual solution to handle this problem is monthly chaining. Price and quantity bouncing due to sales, which is another feature of scanner data, can create drift in chained indexes, however. Ivancic, Fox and Diewert (2009) have recently proposed a novel approach, by adapting multilateral index number theory, that provides drift free, superlative-type indexes. In this paper we apply their proposal to seven product groups and find very promising results. We compare the results with those obtained by using the Dutch method to deal with supermarket scanner data in the CPI.

Keywords: consumer price index (CPI), chain drift, multilateral index number methods, scanner data, superlative indexes.

---

[a] Corresponding author; Statistics Netherlands, Division of Macro-economic Statistics and Dissemination, P.O. Box 24500, 2490 HA The Hague, The Netherlands; j.dehaan@cbs.nl

[b] We thank Lorraine Ivancic for letting us use her SAS program to compute the price index numbers. The views expressed in this paper are those of the authors and do not necessarily reflect the views of Statistics Netherlands.

1. Introduction

The advantage of using scanner data in the Consumer Price Index (CPI) is that prices and quantities on all goods are available so that the construction of weighted (preferably superlative) price indexes at detailed aggregation levels becomes feasible. But scanner data also have a number of potential drawbacks, such as a high attrition rate of goods and volatility of the prices and quantities due to sales. High-frequency chaining seems a natural solution at first sight to handle new and disappearing goods, but that could lead to drift in weighted indexes when prices and quantities oscillate or 'bounce'.[1] Quantity bouncing arises from the fact that households tend to stock up during sale periods and consume from inventory at times when the goods are not on sale. According to Triplett (2003, p. 152) we require "a theory that adequately describes search, storage, shopping, and other household activities that drive a wedge between acquisitions periodicity and consumption periodicity." While that may be true, in our opinion it is unnecessary to wait until all problems associated with the use of scanner data are resolved. Producing official statistics will always involve making assumptions and pragmatic choices.

In particular, we assume that for a homogeneous good the unit value computed across all purchases from a single retail chain during a month is an acceptable measure of the average price paid by the representative consumer.[2] Essentially we are assuming that price and quantity variation within a month represents noise in the data and is not meaningful in the context of a CPI. Still, sales cause considerable bouncing of monthly unit values and quantities. A trivial solution to the problem of drift is not to chain at all and use a direct index, as suggested by Feenstra and Shapiro (2003). This is problematic considering the small number of products that match over time. Another solution would be to exclude goods that are on sale, which is what Statistics Norway does to compute monthly chained price indexes from scanner data; see Rodriguez and Haraldsen (2006). This is unsatisfactory too: it often happens that popular goods go on sale and excluding such goods leads to biased indexes unless long-run price trends are unaffected.[3]

[1] Szulc (1983) seems to have been the first to address the problem of price bouncing and chaining.

[2] Thus we aggregate across stores belonging to one chain, which often have a common pricing policy, but we do not aggregate across different chains. This is consistent with empirical findings by Ivancic (2007). For more information on the use of unit values, see Diewert (1995), Balk (1998), and ILO et al. (2004).

[3] De Haan (2008a) investigated a third option where the superlative index number formula in a chained index is allowed to change over time.

An interesting approach has recently been proposed by Ivancic, Fox and Diewert (2009). They adapt multilateral index number theory to provide weighted indexes which make maximum use of all possible matches in the data between any two months and are free of drift. They write: “Discussion of methods of how best to use scanner data in the context of constructing consumer price indexes is particularly important at the present moment as statistical agencies worldwide are becoming increasingly interested in using scanner data in their official CPI figures. To our knowledge, scanner data are currently used only by two statistical agencies: the Central Bureau of Statistics in the Netherlands and Statistics Norway.” Statistics Netherlands is going to expand the use of scanner data to nearly all major supermarket chains as part of the re-design of the CPI (see De Haan, 2006). This should take place during the second half of 2009. The method developed by Ivancic, Fox and Diewert (2009) will not be used, however, for reasons we will explain later on. The aim of the present paper is to give some background material on this novel approach, apply it to a large Dutch scanner data set to investigate whether it works as expected, and compare the results with those obtained using the new Dutch method for treating scanner data.

The paper is structured as follows. Section 2 describes the scanner data set we will utilize, which covers seven product categories and 44 months. The data come from a single supermarket chain in the Netherlands and are currently inputs to the CPI. We focus on aspects like price and quantity bouncing, the lack of matching over time, and temporarily unavailable products. Section 3 confirms what others found earlier, namely that high-frequency chaining of price indexes, including superlative ones, can lead to drift when sales occur. For the monthly chained Törnqvist index we observe downward drift in most cases. In Section 4 we discuss the method proposed by Ivancic, Fox and Diewert (2009) to eliminate chain drift and find promising results. A slightly amended version is also presented. Section 5 addresses the Dutch method to handle supermarket scanner data in the future and compares the results with those obtained by applying the Ivancic, Fox and Diewert (2009) method. The Dutch approach is based on a monthly-chained (matched-items) Jevons price index with two modifications: the use of cut-off sampling to remove items with extremely low expenditure shares and imputations for temporarily ‘missing’ prices. Section 6 concludes and points to future work.

2. Features of Scanner Data

Supermarket scanner data have three important features which should be borne in mind when compiling price index numbers: price and quantity bouncing as a result of sales, a high attrition rate of new and disappearing items, and temporarily unavailable items (or 'missing prices'). In this section we present illustrative examples of those features. Our data set covers 191 weeks (44 months) of data on seven product categories: detergents, toilet paper, diapers, candybars, nuts and peanuts, beef and eggs. The product categories are not a random selection; we selected them for their heavy price bouncing behaviour. The data come from a large sample of stores belonging to one of the major supermarket chains in the Netherlands and are currently used in the CPI.[4]

Individual items are identified by the European Article Number (EAN). For all EANs, aggregate weekly expenditures and quantities are known, as well as a short item description. Dividing expenditures by quantities purchased gives the unit value, which is our measure of (average) price. Figure 1 shows the weekly unit values, quantities and expenditures for a detergent referred to as *XXX tablets*. There seems to be a 'regular price' of slightly more than 6.5 euros. In quite a number of weeks the item is on sale, with price reductions up to 50%. From our own experience we know that a sales period in this particular supermarket chain lasts for exactly a week (Monday through Sunday), which coincides with our weekly data. Nevertheless, the unit value for the week after a heavy discount is consistently much lower than the 'regular price'. This might be due to the fact that people who wish to buy a good that is on sale but happens to be sold out are entitled to purchase it at the sale price during the next week. So the unit values for post-sales weeks often include sale prices.[5] Figure 1 also shows what we have called quantity bouncing. The quantity shifts associated with sales are really dramatic. Consumers react instantaneously to discounts and purchase large quantities of the good – as a matter of fact, they hardly buy the good when it is not on sale. In this respect it is inappropriate to speak of a regular price during non-sale weeks. Note that the pattern of expenditures is almost identical to the pattern of quantities.

---

[4] The scanner data are provided to Statistics Netherlands at marginal cost. The agency has a policy of not paying for data which are directly used for the compilation of statistics. Scanner data are confidential and cannot be made publicly available.

[5] This explanation was suggested to us by Lida Martens. In the post-sales week there may also be some goods left on the shelves that can still be bought at the sale price.

Insert Figure 1

A priori one might expect the volatility of price and quantity data to diminish if we aggregated across months instead of weeks. This is not the case for *XXX tablets*, as Figure 2 makes clear. The monthly prices and quantities exhibit bouncing similar to the weekly data. For the larger part this is a result of the irregular pattern of weekly sales. Looking at the monthly values, the term regular price is indeed a misnomer: sale prices (unit values) are now just as common as non-sale prices.

Insert Figure 2

Another aspect of supermarket scanner data is the huge attrition rate: the number of disappearing and new items is usually large. Conversely, the number of items that are available in the stores for many weeks in a row is typically low. Figure 3 displays the number of matched items for monthly data on detergents in three ways. The downward sloping curve shows how the set of items at the beginning of the period (January 2005) shrinks over time. Only seven out of the 58 initial items can still be purchased at the end of the period (August 2008).[6] The upward sloping curve should be read in reverse order: it depicts the number of matches between the last month (August 2008) and each earlier month. A comparison with the downward sloping curve indicates that the total number of different types of detergent changes little in the long run because there are almost as many entries as exits. The third curve depicts the number of monthly matched items, i.e. items which are available in consecutive months. In the short run some marked changes occur. For example, it seems as if in August 2005 the supermarket chain removed part of its detergents assortment and replenished it gradually.

Insert Figure 3

[6] The obvious lesson for price measurement is that adhering to a strict matched-item principle – in other words, using a completely fixed sample of items – is impossible. This point is also stressed by Silver and Heravi (2005). They are especially interested in the use of quality adjustment methods to account for new and disappearing items.

Figure 4 plots monthly unit values for *YYY toilet paper*. This product has been unavailable during many months – the quantities are zero, giving rise to 'holes' in the data set. Practitioners would probably say that the prices are temporarily missing. Any monthly chained, matched-item index number method misses the price change between the last month the item was available and the month it re-enters the stores. For instance, the price increase between April 2005 and October 2007 in Figure 4 would be left out from the computation. The practical solution is to impute the 'missing prices'. We will return to this issue in Section 5 when discussing the new Dutch method.

Insert Figure 4

The EAN is a unique identifier at the lowest level of aggregation. In some cases this level may be too detailed: goods that are identical from the consumer's perspective may have different EANs. A fraction of the 'holes' in the data set could be attributable to this effect. Matching by EAN might thus understate the number of matched products and overstate the rate of turnover of new and disappearing products. This is perhaps just a minor issue.

3. Chained Superlative Indexes

3.1 The Problem of Chain Drift

Chained indexes may suffer from what is known as chain drift or chain link bias. Chain drift occurs if a chained index "does not return to unity when prices in the current period return to their levels in the base period" (ILO, 2004, p. 445). In this section we address chain drift in superlative price indexes.[7] Let p_i^0 and s_i^0 denote the price and expenditure share of good *i* in the base period 0; p_i^t and s_i^t denote the corresponding values in the comparison period *t* ($t > 0$). For a fixed set of goods *U* the Fisher and Törnqvist price indexes are defined as

[7] The attraction of superlative price indexes is that they approximate the underlying cost of living index to the second order while being easy to compute (Diewert, 1976). These indexes also have many desirable axiomatic properties; see e.g. and ILO et al. (2004). The Fisher and Törnqvist indexes are the best known superlative indexes. Ehemann (2005) addresses chain drift in Fisher and Törnqvist indexes. On chaining, see also Forsyth and Fowler (1981).

$$P_F^{0t} = \left[\frac{\sum_{i \in U} s_i^0 (p_i^t / p_i^0)}{\sum_{i \in U} s_i^t (p_i^t / p_i^0)^{-1}} \right]^{1/2} ; \tag{1}$$

$$P_T^{0t} = \prod_{i \in U} (p_i^t / p_i^0)^{(s_i^0 + s_i^t)/2} . \tag{2}$$

If the expenditure shares of all goods would coincide ($s_i^t = s_i^0 = 1/N$, where N denotes the number of goods), the Törnqvist index reduces to the Jevons index

$$P_J^{0t} = \prod_{i \in U} (p_i^t / p_i^0)^{1/N} . \tag{3}$$

Many statistical agencies are nowadays using the Jevons index to compile price indexes at the elementary level if expenditure data are lacking. For scanner data an unweighted index number formula seems irrelevant, but the new Dutch method for the treatment of scanner data does apply the Jevons formula, as will be outlined in Section 5.

We will start by distinguishing three periods: 0, 1 and 2. The chained Fisher and Törnqvist price indexes going from period 0 to period 2 are

$$P_{F,chain}^{02} = \left[\frac{\sum_{i \in U} s_i^0 (p_i^1 / p_i^0)^1}{\sum_{i \in U} s_i^1 (p_i^1 / p_i^0)^{-1}} \right]^{1/2} \left[\frac{\sum_{i \in U} s_i^1 (p_i^2 / p_i^1)^1}{\sum_{i \in U} s_i^2 (p_i^2 / p_i^1)^{-1}} \right]^{1/2} ; \tag{4}$$

$$P_{T,chain}^{02} = \prod_{i \in U} (p_i^1 / p_i^0)^{(s_i^0 + s_i^1)/2} \prod_{i \in U} (p_i^2 / p_i^1)^{(s_i^1 + s_i^2)/2} . \tag{5}$$

Price bouncing for a single good is a stylized version of a situation we often observe in supermarket scanner data. Suppose good 1 has been on sale in period 1 and its price has decreased considerably ($p_1^1 / p_1^0 < 1$) while in period 2 the price returned to the initial value ($p_1^2 = p_1^0$ or $p_1^2 / p_1^1 = p_1^0 / p_1^1$). The prices of all other goods are assumed fixed. Expressions (4) and (5) then simplify to

$$P_{F,chain}^{02} = \left[\frac{s_1^0 \{(p_1^1 / p_1^0) - 1\} + 1}{s_1^2 \{(p_1^1 / p_1^0) - 1\} + 1} \right]^{1/2} ; \tag{6}$$

$$P_{T,chain}^{02} = (p_1^1 / p_1^0)^{(s_1^0 - s_1^2)/2} . \tag{7}$$

Standard micro-economic theory assumes that, given a set of prices, the quantities are uniquely determined. So if prices bounce we would expect the quantities, and hence the expenditure shares, to return to their initial levels ($s_1^2 = s_1^0$) so that $P_{F,chain}^{02} = P_{T,chain}^{02} = 1$. However, 'distortions' may give rise to a difference between s_1^0 and s_1^2. In this stylized example we have $P_{F,chain}^{02} < 1$ and $P_{T,chain}^{02} < 1$ for $s_1^2 < s_1^0$, and $P_{F,chain}^{02} > 1$ and $P_{T,chain}^{02} > 1$ for $s_1^2 > s_1^0$.

This example does not represent our weekly data very well. From Section 2 the following pattern emerges. In week 0 good 1 is sold at the regular price and the quantity is very low or almost zero. In week 1, when the good is sold at the low sales price, the quantity is extremely high. In week 2 the price of good 1 is only slightly higher than in week 1 (though much lower than the regular price) but now the quantity is low, though not as low as in week 0. In week 3 both the price and the quantity return to their initial levels. Assuming again that the prices of the other goods stay the same, the four-period chained Törnqvist index can be written as

$$P_{T,chain}^{03} = (p_1^1 / p_1^0)^{(s_1^0 + s_1^1)/2} (p_1^2 / p_1^1)^{(s_1^1 + s_1^2)/2} (p_1^3 / p_1^2)^{(s_1^2 + s_1^3)/2}. \qquad (8)$$

Can anything be said a priori about the expected sign of chain drift in $P_{T,chain}^{03}$ in case of storable goods? The first component of (8) is probably the leading term: the strong price decrease p_1^1 / p_1^0 receives extraordinary large weight due to the high quantity purchased in period 1 (in particular when the quantities of the other goods have decreased, which is most likely for substitutable goods). Although the weight of the second component of (8) may even be greater, the price increase p_1^2 / p_1^1 is small and we expect its impact to be modest. The strong price increase p_1^2 / p_1^1 receives relatively small weight since the quantity in period 3 returns to the period 0 level. All in all, we would expect $P_{T,chain}^{03}$ to be below unity so that downward drift prevails.

In real life the situation is more complicated. The sign of the drift depends on the magnitude of the price decrease and the associated quantity shifts of all goods belonging to the product group, and on the periodicity of acquisition and consumption.[8] Different

[8] Feenstra and Shapiro (2003), using data on canned tuna, found that the weekly chained Törnqvist index had an upward drift: "in periods when the prices are low, but there are no advertisements, the quantities *are not* high […]. Because the ads occur in the final period of the sales, the price *increases* following the sales receive much greater weight than the price *decreases* at the beginning of each sale. This leads to the dramatic upward bias of the chained Törnqvist." That consumers are misinformed without advertisements surprises us a little bit. As was shown in Section 2, in our data set we observe instantaneous responses of consumers to strong price reductions: the quantities immediately increase dramatically and drop to almost zero in after-sales weeks.

goods can be on sale at different times. Furthermore, the set of goods *U* is typically not fixed. If it were, there was no use in chaining – direct superlative price indexes such as the Fisher and Törnqvist, given by (1) and (2), should then be used. Aggregation across time might help reduce the problem of chain drift, assuming that high frequency price and quantity variation represents noise in the data. Statistical agencies do not compile CPIs on a weekly basis anyway, so it is rather obvious to work with monthly unit values and quantities. In Section 3.2 we present some evidence on this topic.

3.2 Results

Figure 5 confirms what others have found before (Feenstra and Shapiro, 2003; Ivancic, 2007; De Haan, 2008; Ivancic, Fox and Diewert, 2009): weekly chaining of superlative indexes can lead to exceptionally large drift. For detergents we observe downward drift. Fisher and Törnqvist indexes measure a totally unrealistic price decrease of more than 90% in less than four years. The downward trend of the Jevons index is much smaller. This accords with expectations as it is the asymmetry of expenditure weights that drives chain drift in superlative price indexes. Still, the price decrease measured by the Jevons seems rather large.

Insert Figure 5

As can be seen from Figure 6, aggregating price and quantity data across months instead of weeks dramatically reduces chain drift. Although we cannot be sure that the monthly chained index numbers for detergents are completely free of drift, at least they look plausible. Notice that the Fisher and Törnqvist index numbers are almost identical, notwithstanding the volatility of the monthly price and quantity data. Monthly chaining raises the superlative indexes above the Jevons index. Nevertheless, the monthly Jevons price index numbers are higher than the weekly numbers. The sensitivity of the Jevons to time aggregation surprises us a bit.

Insert Figure 6

Figure 7 shows what happens if we further aggregate over time and use quarterly unit values and quantities to compute quarterly chained indexes. This is not very helpful for statistical agencies that compile monthly CPIs, but it may be considered in Australia, New Zealand and other countries where the CPI is published on a quarterly basis. The results for detergents are striking. Quarterly chained superlative indexes measure a price *increase* of 20% or more. We find this highly implausible. The Fisher and Törnqvist for the last quarter differ 5 points, which is remarkable too. Figure 7 seems to suggest that quarterly data suffer from 'too much' aggregation across time – the noise in the data has been eliminated but at the cost of messing up the trend.

Insert Figure 7

4. GEKS and Rolling Year GEKS Indexes

4.1 The Basic Idea and Some Background

Ivancic, Fox and Diewert (2009), henceforth IFD, have recently proposed a method for constructing price indexes that use all matches in the data between any two periods and that are, in contrast to high-frequency chained indexes, free of drift. The method is an adapted version of the multilateral GEKS (Gini, 1931; Eltetö and Köves; 1964; Szulc, 1964) approach. The GEKS index is the geometric mean of the ratios of all bilateral indexes (computed with the same index number formula) between a number of entities, where each entity is taken as the base. Let P^{jl} and P^{kl} be the bilateral indexes between entities *j* and *l* $(l = 1,...,M)$ and between entities *k* and *l*, respectively. The GEKS index between *j* and *k* can then be written as

$$P_{GEKS}^{jk} = \prod_{l=1}^{M} \left[P^{jl} / P^{kl} \right]^{1/M} = \prod_{l=1}^{M} \left[P^{jl} \times P^{lk} \right]^{1/M} , \qquad (9)$$

where the second expression holds when the bilateral indexes satisfy the 'entity reversal test', so that $P^{kl} = 1/P^{lk}$. It can easily be shown that

$$P_{GEKS}^{jk} = P_{GEKS}^{jl} / P_{GEKS}^{kl} . \qquad (10)$$

Expression (10) says that the GEKS price index satisfies the circularity or *transitivity* requirement: the same result is obtained if entities are compared with each other directly or via their relationships with other entities.

Multilateral indexes such as the GEKS are often used to make price comparisons across countries (or regions); see Diewert (1999a) and Balk (2001; 2008) for overviews. Transitivity is particularly useful to circumvent the choice of base or bridge country, but a drawback is that a transitive index for two countries depends on the data of all other countries – there is a *loss of characteristicity*.[9] The GEKS method can be justified as a means of preserving characteristicity as much as possible. More specifically, the GEKS price index is the solution to minimizing $\sum_{j=1}^{M}\sum_{k=1}^{M}(\ln P^{*jk} - \ln P^{jk})^2$, being the sum of squared differences between the logarithms of a (multilateral) index P^{*jk} for a pair of countries j,k and the direct (bilateral) index P^{jk}. Notice that the direct index 'counts twice' in equation (9), namely for $l = j$ and $l = k$.

IFD adapt the GEKS method to price indexes across time by treating each time period as an entity.[10] That is, *j* and *k* in expression (9) are now time periods and *l* is the link period. Suppose we have data on prices and quantities at our disposal for periods $0,1,...,T$. Choosing 0 as the index reference period and denoting the comparison periods by t $(t = 1,...,T)$, we can write the adapted GEKS index going from 0 to *t* as

$$P_{GEKS}^{0t} = \prod_{l=0}^{T}\left[P^{0l} / P^{tl}\right]^{1/(T+1)} = \prod_{l=0}^{T}\left[P^{0l} \times P^{lt}\right]^{1/(T+1)}, \tag{11}$$

provided that the bilateral indexes satisfy the time reversal test. In that case the GEKS index also satisfies this test, i.e. $P_{GEKS}^{t0} = 1/P_{GEKS}^{0t}$. The transitivity property implies that the GEKS index can be written as a period-to-period chained index, i.e.

$$P_{GEKS}^{0t} = \prod_{\tau=1}^{t} P_{GEKS}^{\tau-1,\tau}, \tag{12}$$

---

[9] Characteristicity is "the property that requires the transitive multilateral comparisons between members of a group of countries to retain the essential features of the intransitive binary comparisons that existed between them before transitivity" (Eurostat and OECD, 2006, p. 127). According to Caves, Christensen and Diewert (1982) characteristicity refers to the "degree to which weights are specific to the comparison at hand".

[10] They borrow an alternative method from the international comparisons literature, the Country Product Dummy (CPD) method, and adapt it to provide price indexes free of chain drift. The resulting estimates have standard errors associated with them. IFD argue that the lack of standard errors is a drawback of the GEKS methodology. We disagree with this view. The choice of index number formula is what matters. Index numbers that do not rely on sampling, as with scanner data, have no standard errors, or at least no sampling error (unless there would be imputations involved). The CPD approach, like any model-based approach, adds error because of the use of a stochastic model.

which should be free of chain drift.

The bilateral indexes are all matched-item indexes: only price relatives of items that are purchased in the two periods compared enter the indexes. IFD call this a flexible basket approach. The GEKS approach thus makes maximum use of all possible matches in the data between any two periods, which can be seen as its most important property. Imputations to deal with 'missing prices' are therefore unnecessary. Any matched-item index, including the GEKS, does not explicitly account for quality change.[11] For many fast-moving goods purchased in supermarkets quality change is arguably a minor issue. Even if quality changes are substantial, measuring prices of matched items might suffice under competitive market circumstances.

P_{GEKS}^{0t}, given by (11), depends on the price and quantity data of all time periods, including $t+1,...,T$. In real time we cannot produce an index based on future data. What we can do in practice is calculate the GEKS index for the current (most recent) period T using all the available data and update the time series as time passes. It is now more convenient to write the GEKS index going from period 0 to period T as

$$P_{GEKS}^{0T} = \prod_{t=0}^{T} \left[P^{0t} / P^{Tt} \right]^{1/(T+1)} = \prod_{t=0}^{T} \left[P^{0t} \times P^{tT} \right]^{1/(T+1)}. \tag{13}$$

Before discussing the updating of the time series we address one other issue first. While transitivity is a useful property, it is not a necessary requirement in a time series context where chronological ordering of the price indexes is the unique ordering. GEKS index P_{GEKS}^{0T} results from minimizing $\sum_{s=0}^{T} \sum_{t=0}^{T} (\ln P^{*ts} - \ln P^{ts})^2$ for any two periods *s* and *t*. But why should this be the optimal rule for deriving a price index going from 0 to *T*? Minimizing the sum of squared differences is a natural choice for a comparison between countries because the direct (bilateral) indexes are 'better' than other indexes. In a time series context, where a lack of matched items is the problem, the direct index may not be best. Suppose that the number of matches gradually decreases over time. The longer

[11] Quality change can best be seen as the appearance of new products and the disappearance of 'old' ones at the lowest possible aggregation level. From an index number point of view quality adjustment methods should therefore estimate what the prices of those products would have been if they had been available. Put otherwise, quality adjustment methods such as hedonic regression are essentially imputation methods; see Diewert, Heravi and Silver (2007) and De Haan (2008b). This raises the question whether the GEKS approach would still be of some use if we imputed all 'missing prices' through hedonic regression (or the like).

the period, the less we want to rely on the direct index. In other words, while in this case the direct index P^{0T} is less representative than the indirect indexes $P^{0t} \times P^{tT}$ $(t \neq 0, T)$, it has twice the weight.[12] We therefore alternatively consider the unweighted geometric mean of the direct and indirect indexes, which obviously also makes use of all matches in the data between any two time periods:

$$P_{ALT}^{0T} = \prod_{t=1}^{T} \left[P^{0t} \times P^{tT} \right]^{1/T} . \tag{14}$$

It can easily be shown that P_{ALT}^{0T} is not transitive. If the bilateral indexes satisfy the time reversal test then so does P_{ALT}^{0T}.

Now we turn to updating the time series. The GEKS index for period $T+1$ using price and quantity data pertaining to all periods $t = 0,...,T+1$ is

$$P_{GEKS}^{0,T+1} = \prod_{t=0}^{T+1} \left[P^{0t} / P^{T+1,t} \right]^{1/(T+2)} = \prod_{t=0}^{T+1} \left[P^{0t} \times P^{t,T+1} \right]^{1/(T+2)} . \tag{15}$$

A drawback is that the index number for period *T* would be revised if we re-computed it using the extended data set.[13] We denote the revised index number by $P_{GEKS(0,T+1)}^{0T}$. There is however no need to publish the revised numbers. Since the time series is free of drift, we may use the change in the GEKS index between $T+1$ and *T*, $P_{GEKS}^{0,T+1} / P_{GEKS(0,T+1)}^{0T}$, as the chain link to update the time series. Due to transitivity we have (for bilateral indexes that satisfy the time reversal test)

$$P_{GEKS}^{0,T+1} / P_{GEKS(0,T+1)}^{0T} = \prod_{t=0}^{T+1} \left[P^{t,T+1} / P^{tT} \right]^{1/(T+2)} , \tag{16}$$

so that the index for period $T+1$ would become

$$P_{GEKS}^{0,T+1} = P_{GEKS}^{0T} \prod_{t=0}^{T+1} \left[P^{t,T+1} / P^{tT} \right]^{1/(T+2)} . \tag{17}$$

---

[12] On the other hand, if (nearly) all items do match between period 0 and period *T*, then we would in fact prefer the direct index. This suggests taking a weighted average of the direct and indirect indexes, where the weights somehow depend on the number of matches. Weights can be inserted into the minimization rule (see e.g. Balk, 2008, Ch. 7), but it is not easy to see how to derive weights without making arbitrary choices.

[13] In the words of Hill (2004), the GEKS index violates time fixity. Most statistical agencies would find this unacceptable.

The same approach could be followed to extend the time series to periods $T+2$, $T+3$, etc. Clearly, any index changes derived from the time series constructed in this way, for instance the annual inflation rate, are affected by the prices and quantities pertaining to earlier periods. To diminish the loss of characteristicity, IFD use a so-called *rolling year approach*.

We assume that, like in most countries, the CPI is a monthly statistic. The rolling year approach uses the price and quantity data for the last 13 months to compute GEKS indexes. As in (17), the most recent month-to-month index change is then chain linked to the existing time series. The choice for a 13 month moving window is optimal in the sense that it allows a comparison of strongly seasonal items.[14] Longer windows could be chosen, but that would lead to a greater loss of characteristicity. Using $P_{GEKS}^{0,12}$ as the starting point for constructing a monthly time series, the rolling year GEKS (RGEKS) index for month $T+1$ becomes

$$P_{RGEKS}^{0,13} = P_{GEKS}^{0,12} \prod_{t=1}^{13} \left[P^{12,t} / P^{13,t} \right]^{1/13} = \prod_{t=0}^{12} \left[P^{0t} / P^{12,t} \right]^{1/13} \prod_{t=1}^{13} \left[P^{12,t} / P^{13,t} \right]^{1/13} . \qquad (18)$$

The general expression for the RGEKS index going from an arbitrary base month 0 to the current month T $(T > 12)$ is

$$P_{RGEKS}^{0T} = \prod_{t=0}^{12} \left[P^{0t} / P^{12,t} \right]^{1/13} \prod_{t=13}^{T} \prod_{t=T-12}^{T} \left[P^{T-1,t} / P^{T,t} \right]^{1/13} . \qquad (19)$$

The rolling year method can also be applied to the alternative index given by expression (14), using $P_{ALT}^{0,12}$ as the starting point.

GEKS and RGEKS indexes are preferably based on superlative bilateral indexes because they satisfy the time reversal test and have other desirable axiomatic properties. IFD calculate GEKS indexes using bilateral Fisher indexes. They also estimate RGEKS indexes for (no more than) three months – their data series is only 15 months long. We chose to work with Törnqvist price indexes and compute GEKS and RGEKS for a much longer time period. In addition we will use Jevons bilateral price indexes to investigate the impact of weighting and to compare the results with monthly chained Jevons price indexes presented in Section 5. The Jevons also satisfies the time reversal test.

[14] Strongly seasonal goods can only be purchased during some months of the year. For a discussion on the problems associated with seasonality, see Diewert (1999b),

4.2 Results

To get an idea of the potential effects of revisions, Figure 8 depicts two monthly GEKS-Törnqvist indexes for detergent during January 2005 – January 2006. The first one uses the data of those 13 months only, the second one is based on all data that is available to us (44 months), including data from February 2006 through August 2008. The revision is downward. While being small as compared to the volatility of the index numbers, it cannot be ignored.

Insert Figure 8

Figure 9 shows monthly RGEKS-Törnqvist and RGEKS-Jevons indexes for all seven product categories. The alternative indexes in which the direct bilateral (Törnqvist or Jevons) index counts once, are also shown. The RGEKS-Törnqvist indexes show no obvious sign of drift, as expected. The highly volatile pattern is somewhat surprising as we would expect the RGEKS approach to smooth price fluctuations. In most cases the RGEKS-Jevons is much lower than the RGEKS-Törnqvist; apparently, low expenditure items exhibit relatively small price increases or large price decreases. This underlines the importance of weighting. The volatility of the RGEKS-Jevons is less than that of the RGEKS-Törnqvist but still substantial. Notice that in general the alternative indexes are slightly higher than their RGEKS counterparts.

Insert Figure 9

Figure 10 compares the RGEKS-Törnqvist indexes (presented in Figure 9) with monthly-chained Törnqvist indexes and direct Törnqvist indexes. Except for detergents, where we find no obvious sign of drift, monthly chaining leads to downward drift. In a number of cases the drift is severe; for toilet paper the difference between the RGEKS-Törnqvist and the chained Törnqvist has risen to 30 index points in August 2008. Direct price indexes are of course free of chain link bias but have the drawback of relying on an increasingly smaller set of items. Figure 10 confirms that the direct (matched items) Törnqvist index should not be used.

Insert Figure 10

5. Chained Jevons Indexes

Scanner data were first introduced into the Dutch CPI in 2002. Price index numbers for two supermarket chains were calculated with the Lowe formula, based on a large cut-off sample of items (EANs) for each product group. The expenditure weights of the items were updated annually, or sometimes bi-annually, and the short-term index series were chained in December to obtain long-run series. Although weighting at the item level is a strong point, it had the drawback of 'amplifying' the impact of sales as often the more popular items go on sale, and thus led to volatile index numbers. More importantly, new items could only be introduced in December unless they were selected as replacements for disappearing items. Searching for replacement items and trying to adjust for quality changes was a very labour intensive and time consuming process. This was true also for the initial selection of the basket of items.

As from mid 2009 the use of scanner data will be extended to a large number of supermarket chains. The Jevons instead of the Lowe index number formula is going to be used. In order to update item samples as quickly as possible and enhance efficiency, monthly chained matched-item Jevons price indexes will be computed. The method has several potential drawbacks for which solutions had to be found.

Since the Jevons is an unweighted index, relatively unimportant items, in terms of their expenditure shares, would have the same impact on the index as more important items. To reduce this effect somewhat a crude type of implicit weighting will be applied through cut-off sampling: important items will be included in the sample with certainty whereas unimportant items will be excluded. An item *i* is selected for the index between month *t*-1 to month *t* if its average expenditure share (with respect to the set of matched items) in both months, $(s_i^{t-1} + s_i^t)/2$, is above a certain threshold value. The threshold is given by $1/(N^{t-1,t} \times \chi^{t-1,t})$, where $N^{t-1,t}$ denotes the number of matched items. Initially we chose $\chi^{t-1,t} = 2$. This means that, for example, if $N^{t-1,t} = 50$, then all items with an average expenditure share of more than 1% would be selected. Note that the number of matched items in the sample, $n^{t-1,t}$, as well as the sample aggregate expenditure share, $\sum_{i=1}^{n^{t-1,t}} (s_i^{t-1} + s_i^t)/2$, will change over time. Statistical agencies usually have fixed-size samples ('panels') to compute elementary aggregate price indexes (see e.g. Balk, 2004).

As mentioned earlier, the second disadvantage of a strict matched-items method implies that temporarily missing items are excluded from the computation so that price changes occurring between the last month these items were in the sample and the month they re-enter the sample will be missed. The 'missing prices' will be imputed, as usual, by multiplying the last observed price by the (Jevons) price index of the matched items within the product group in question. In a way we are forcing a panel element onto the dynamic matched-items approach.

Finally, like any matched-items method, the new method does not explicitly take quality changes into account. Since implicit quality-adjustment methods have been most prominent in the Dutch CPI in the past, in this respect the new method is similar to the old one. The newly-built computer system does allow for making explicit adjustments, just in case. In particular, quantity adjustments for changes in package size or contents could be made when deemed necessary. We expect this feature to be used infrequently (and hopefully not at all).

The impact of both adjustments, cut-off sampling and imputation, on the chained matched-items Jevons price index for toilet paper is shown in Figure 11. The unadjusted index clearly has a downward drift. Cut-off sampling $(\chi^{t-1,t} = 2)$ makes things worse. Imputing 'missing prices' turns the downward trend of the sample-based index into an upward trend, particularly during 2008.

Insert Figure 11

Figure 12 compares the adjusted chained Jevons indexes for all product groups with the RGEKS-Törnqvist indexes (from Figure 9) to assess whether both adjustments eliminate the downward bias. The evidence is a bit mixed. For toilet paper the adjusted Jevons ends at the same level as the RGEKS but in the middle of the observation period the difference is large. For detergents, diapers, candybars and beef the adjusted Jevons performs rather well. On the other hand, for nuts and peanuts and for eggs the adjusted Jevons has a severe downward bias. We conclude that although the new Dutch method is not without difficulties, it produces satisfactory results in most cases. Van der Grient (2009) provides more details. Based on our empirical work he also proposes to improve the method by reducing the cut-off sample (using $\chi^{t-1,t} = 1.5$ instead of $\chi^{t-1,t} = 2$).

Insert Figure 12

6. Conclusions and Future Work

In this paper we have applied the method developed by Ivancic, Fox and Diewert (2009) and computed rolling year GEKS price index numbers for seven product categories. The method performs as expected: in contrast to monthly chained superlative price indexes, the RGEKS indexes show no sign of (chain) drift.

In spite of the promising results, Statistics Netherlands will not use the RGEKS method in 2009 to incorporate scanner data from the major supermarket chains into the CPI. Even if we wanted to, it would be impossible due to time constraints – designing and testing an official computer system takes a lot of time and effort, and we will not be able to build such a system on time.[15] A drawback of the RGEKS method, which might make the CPI department reluctant to use it, is a lack of transparency. Practitioners may have difficulties in trying to come up with explanations for implausible price changes. In our opinion this is not a convincing argument against using the RGEKS approach; if a method is clearly better than others, it should be implemented, unless there are serious practical problems or high costs that would prevent this. There is one reason, apart from time constraints, why this new method cannot immediately be applied in the Dutch CPI. Statistics Netherlands has a policy of using only methods that are widely accepted. We interpret this rather vague statement as follows: methods do not necessarily have to be widely used, but they should be accepted as good practice by experts in the field and by the international statistical community. The RGEKS approach is obviously in an early stage, and more evidence is needed to get it widely accepted.

We encourage other statistical agencies – especially those that are already using scanner data and those that are interested in doing so in the near future – to consider the RGEKS method and present empirical evidence. Three issues could be addressed. First, it would be useful to compare RGEKS indexes for strongly seasonal goods such as fresh fruit with scanner data based price indexes calculated using traditional methods to cope with seasonality. Second, RGEKS indexes can be computed at various levels of product aggregation. Our computations were done at a detailed level but it would be worthwhile comparing them to indexes at higher aggregation levels. Third, in addition to monthly indexes, RGEKS indexes can be computed for weekly and quarterly data to investigate

[15] For this study we have used a statistical package (SAS) and a spreadsheet program. This would not be allowed for producing the Dutch CPI.

how increased aggregation over time affects the results. Since they should be drift free, we expect weekly, monthly and quarterly RGEKS indexes indexes to exhibit similar trends.

Statistical agencies that publish the CPI on a quarterly basis, like the Australian Bureau of Statistics and Statistics New Zealand, are most likely interested in quarterly aggregations. So far we have constructed quarterly RGEKS-Törnqvist price indexes for detergents only. In Figure 13 they are shown together with quarterly direct and quarterly chained Törnqvist indexes as well as monthly chained Törnqvist indexes. The latter are calculated as (re-scaled) three-month averages of the index numbers shown in Figure 6. The RGEKS method appears to be fairly insensitive to increased aggregation over time, though the quarterly RGEKS indexes are slightly higher than the monthly counterparts. The direct Törnqvist index is nearly identical to the RGEKS, so for detergents a direct comparison would suffice.

Insert Figure 13

References

Balk, B.M. (1998), On the Use of Unit Value Indices as Consumer Price Sub-Indices, in: W. Lane (ed.), Proceedings of the Fourth Meeting of the Ottawa Group, Bureau of Labor Statistics, Washington, D.C., pp. 112-120.

Balk, B.M. (2001), Aggregation Methods in International Comparisons: What Have We Learned?, ERIM report, Erasmus Research Institute of Management, Erasmus University Rotterdam.

Balk, B.M. (2005), Price Indexes for Elementary Aggregates: The Sampling Approach, *Journal of Official Statistics* 21, 675-699.

Balk, B.M. (2008), *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference*, New York: Cambridge University Press.

Caves, D.W., L.R. Christensen and W.E. Diewert (1982), Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers, *Economic Journal* 92, 73-86.

Diewert, W.E. (1976), Exact and Superlative Index Numbers, *Journal of Econometrics* 4, 115-145.

Diewert, W.E. (1995), Axiomatic and Economic Approaches to Elementary Price Indexes, Discussion Paper no. 95-01, Department of Economics, University of British Columbia.

Diewert, W.E. (1999a), Axiomatic and Economic Approaches to International Comparisons, in: *International and Interarea Comparisons of Income, Output and Prices*, A. Heston and R.E. Lipsey (eds.), pp. 13-87, Studies in Income and wealth, vol. 61, Chicago: The University of Chicago Press.

Diewert, W.E. (1999b), Index Number Approaches to Seasonal Adjustment, *Macroeconomic Dynamics* 3, 1-21.

Diewert, W.E., S. Heravi and M. Silver (2007), Hedonic Imputation versus Time Dummy Hedonic Indexes, IMF Working Paper no. 07/234, IMF, Washington, D.C.

Ehemann, C. (2005), Chain Drift in Leading Superlative Indexes, Working Paper no. 2005-09, Bureau of Economic Analysis, Washington, DC.

Eltetö, Ö, and P. Köves (1964), On a Problem of Index Number Construction Relating to International Comparisons (in Hungarian), *Statisztikai Szemle* 42, 507-518.

Eurostat and OECD (2006), *Methodological Manual on PPPs*.

Feenstra, R.C. and M.D. Shapiro (2003), High Frequency Substitution and the Measurement of Price Indexes, pp. 123-146, in: R.C. Feenstra and M.D. Shapiro (eds.), *Scanner Data and Price Indexes*, Chicago: University of Chicago Press.

Forsyth, F.G. and R.F. Fowler (1981), The Theory and Practice of Chain Price Index Numbers, *Journal of the Royal Statistical Society* A 144, 244-246.

Gini, C. (1931), On the Circular Test of Index Numbers, *Metron* 9:9, 3-24.

Grient, H. van der (2009), Price Index Numbers Based on Scanner Data: Justification of the Choices Made [in Dutch], Internal report, Statistics Netherlands, The Hague.

Haan, J. de (2006), The Re-design of the Dutch CPI, *Statistical Journal of the United Nations Economic Commission for Europe* 23, 101-118.

Haan, J. de (2008a), Reducing Drift in Chained Superlative Price Indexes for Highly Disaggregated Data, paper presented at the Economic Measurement Group Workshop, Sydney, 10-12 December 2008.

Haan, J. de (2008b), Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Other Approaches, Working Paper no. 2008/01, Centre for Applied Economic Research, University of New South Wales, Sydney.

Hill, R.J. (2004), Superlative Index Numbers: Not All of Them Are Super, *Journal of Econometrics* 130, 25-43.

ILO, IMF, OECD, Eurostat, United Nations, World Bank (2004), *Consumer Price Index Manual: Theory and Practice*, Geneva: ILO Publications.

Ivancic, L. (2007), Scanner Data and the Construction of Price Indices, PhD thesis, University of New South Wales, Sydney.

Ivancic, L., K.J. Fox and E.W. Diewert (2009), Scanner Data, Time Aggregation and the Construction of Price Indexes, Mimeo, School of Economics and Centre for Applied Economic Research, University of New South Wales, Sydney.

Rodriguez, J. and F. Haraldsen (2006), The Use of Scanner Data in the Norwegian CPI: The "New" Index for Food and Non-Alcoholic Beverages, *Economic Survey* 4, 21-28.

Silver, M. and S. Heravi (2005), A Failure in the Measurement of Inflation: Results from a Hedonic and Matched Experiment Using Scanner Data, *Journal of Business and Economic Statistic*s 3, 269-281.

Szulc, B. (1964), Indices for Multiregional Comparisons (in Polish), *Przeglad Statystyczny* 3, 239-254.

Szulc, B.J. (1983), Linking Price Index Numbers, in: W.E. Diewert and C. Montmarquette (eds.), *Price Level Measurement*, Statistics Canada, Ottawa.

Triplett, J.E. (2003), Using Scanner Data in Consumer Price Indexes: Some Neglected Conceptual Considerations, pp. 151-162, in: R.C. Feenstra and M.D. Shapiro (eds.), *Scanner Data and Price Indexes*, Chicago: University of Chicago Press.

Figure 1. Weekly unit values, quantities and expenditures; *XXX tablets*

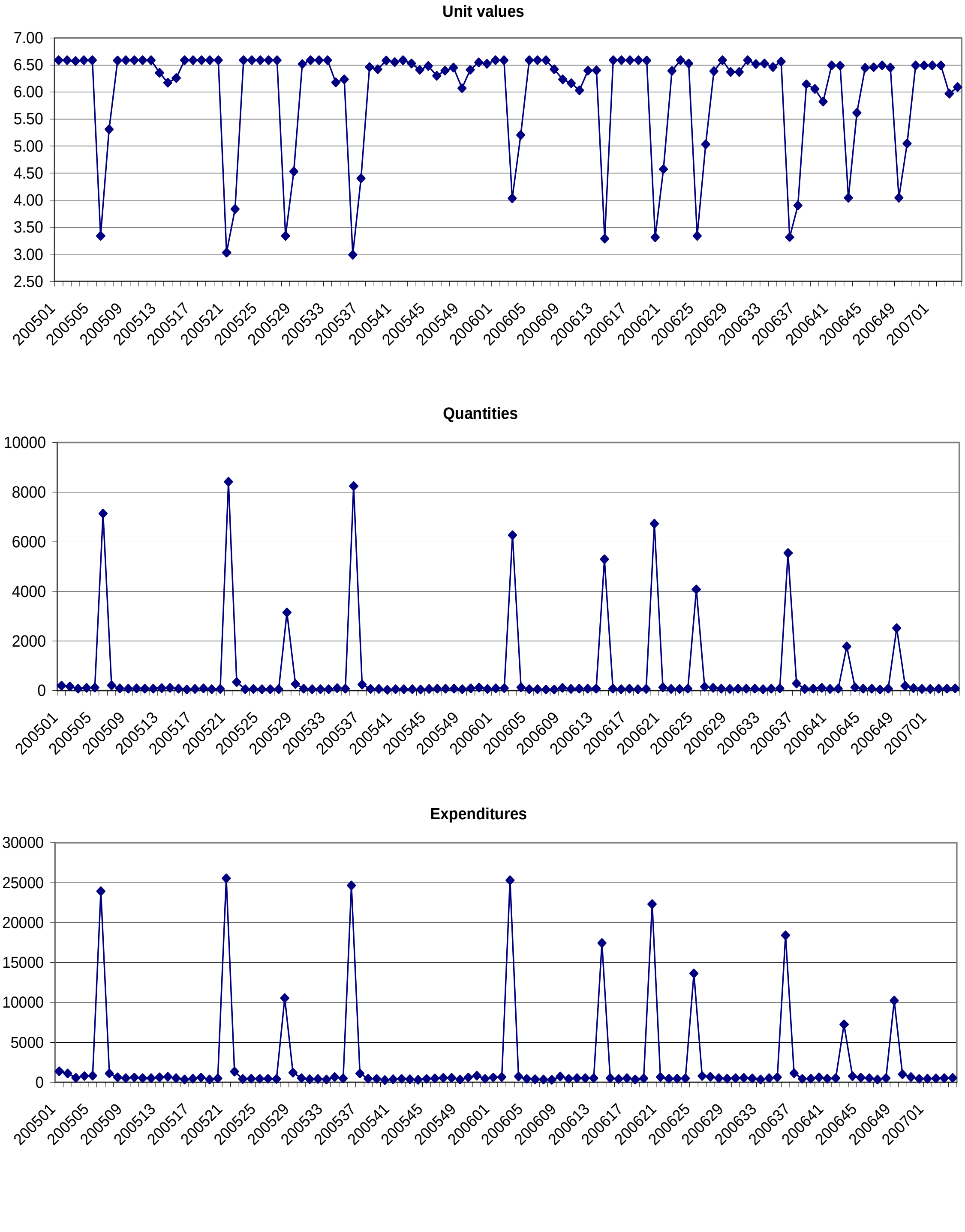


Figure 2. Monthly unit values, quantities and expenditures; *XXX tablets*

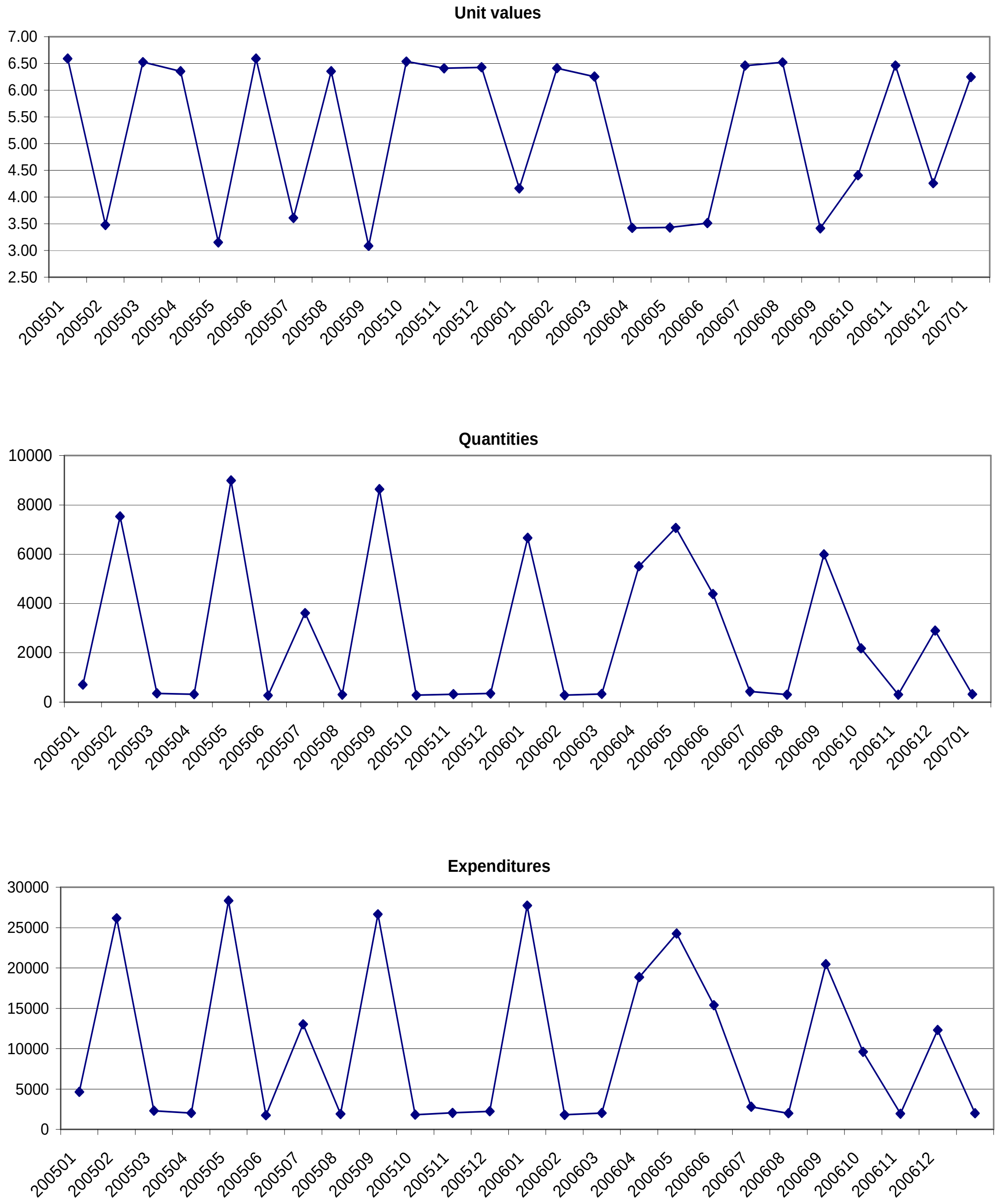


Figure 3. Number of matched items; detergents

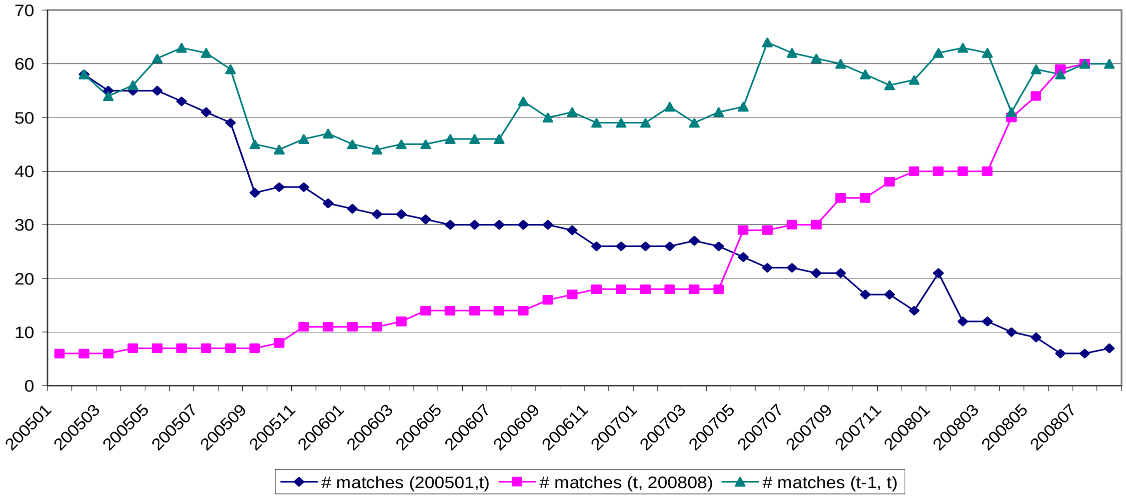


Figure 4. Monthly unit values; *YYY toilet paper*

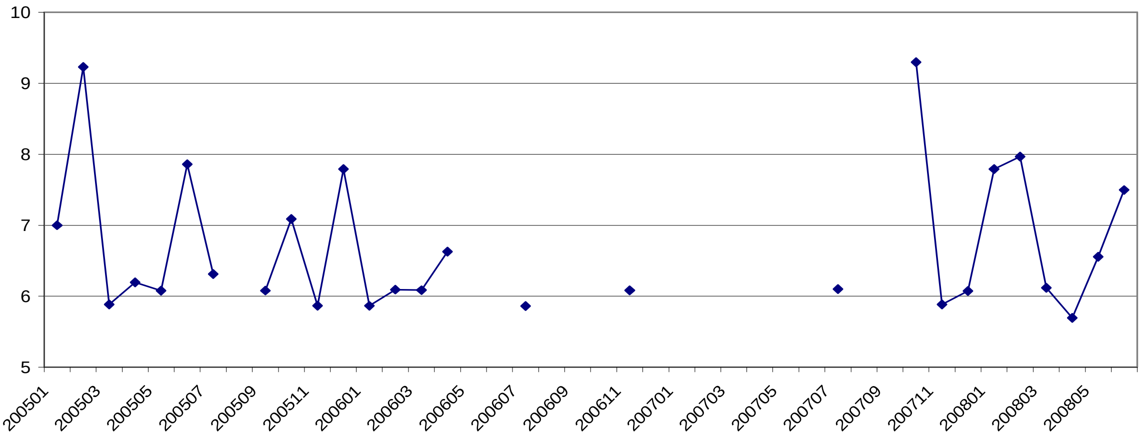


Figure 5. Weekly chained price indexes; detergents

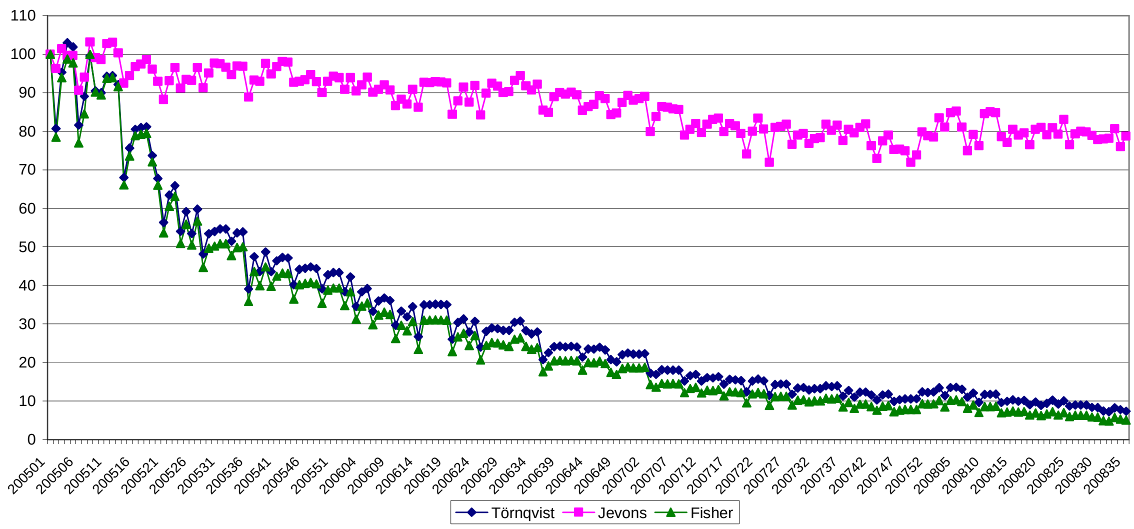


Figure 6. Monthly chained price indexes; detergents

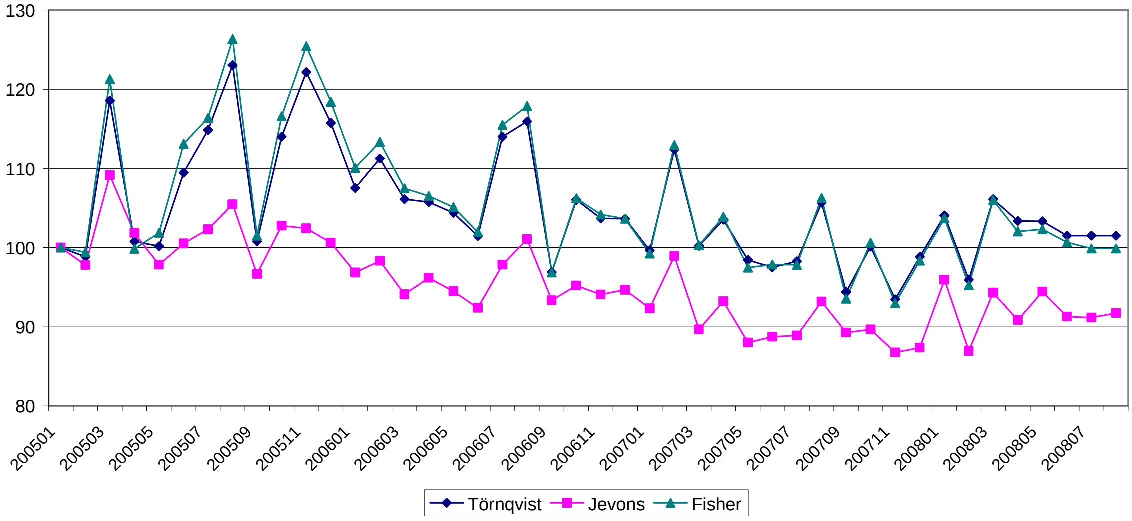


Figure 7. Quarterly chained price indexes; detergents

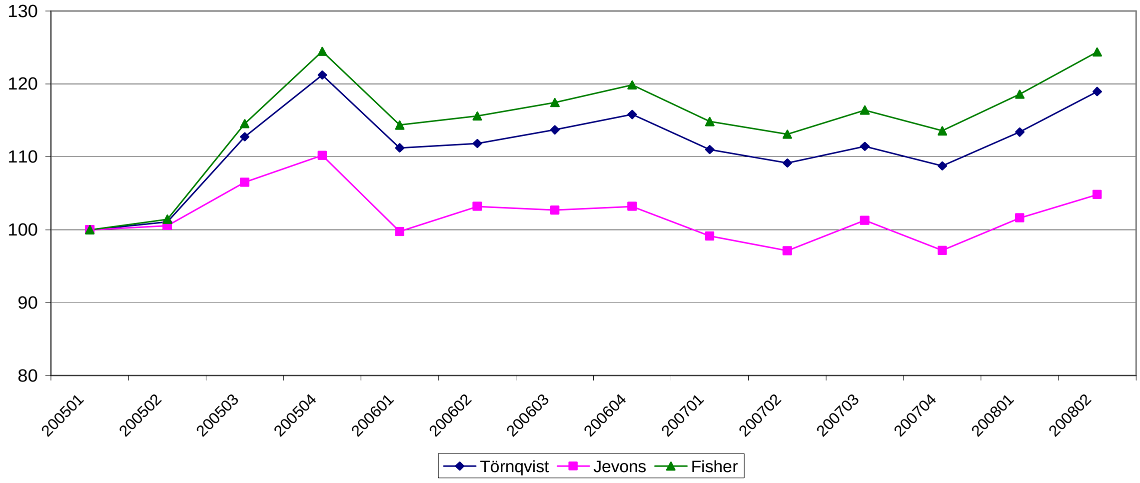


Figure 8. Initial and revised (44 months) GEKS-Törnqvist indexes; detergents

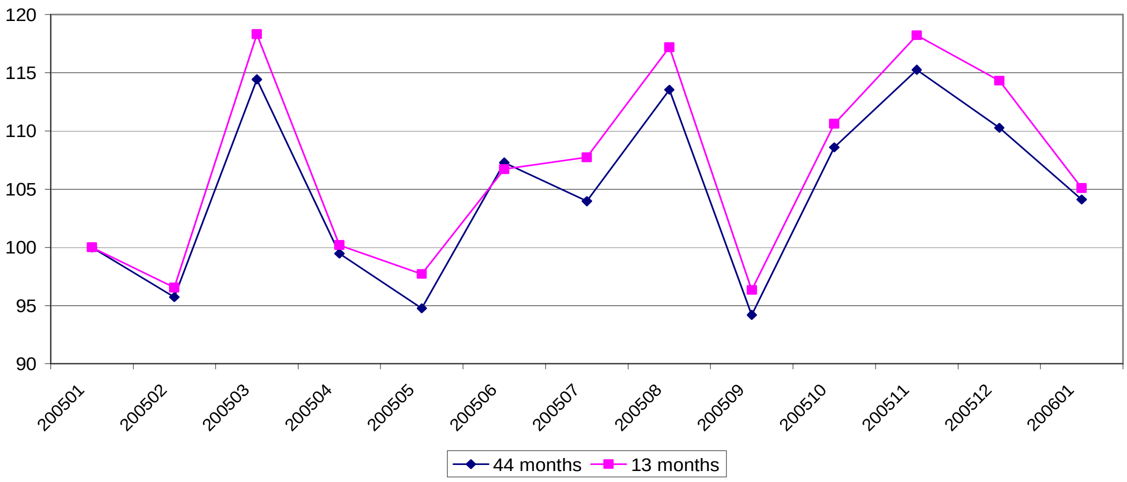
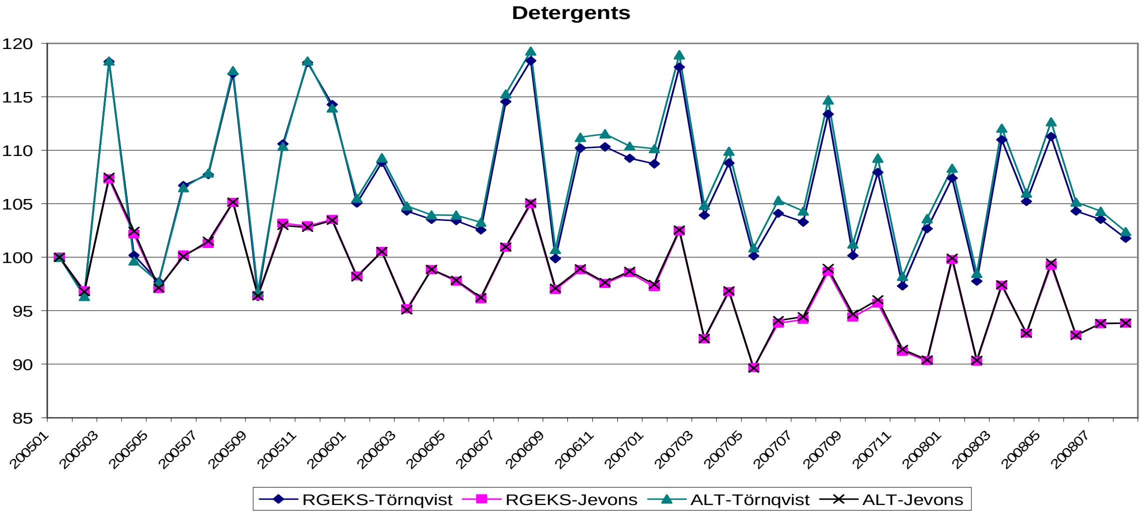
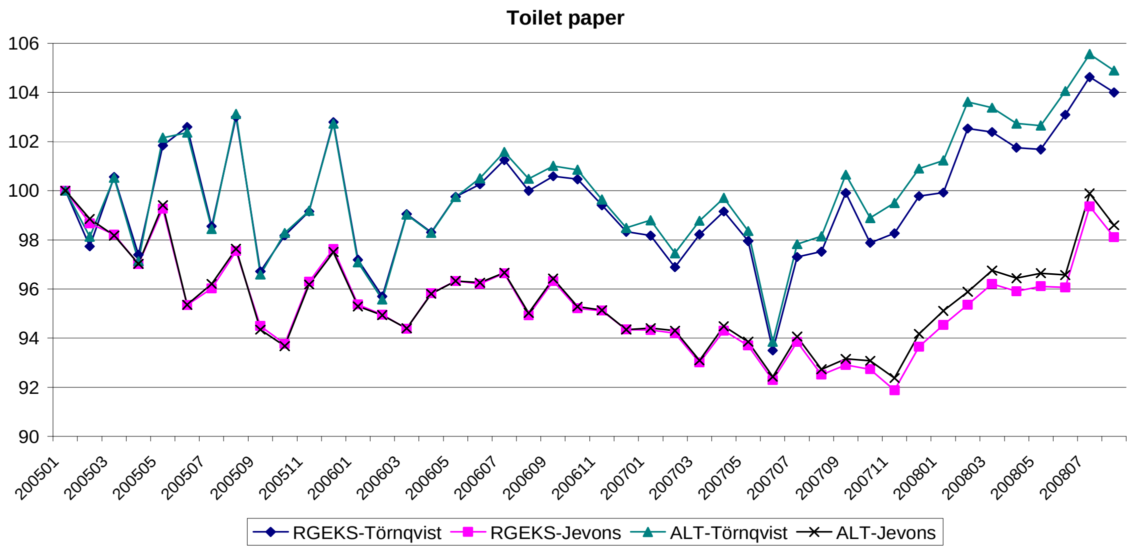
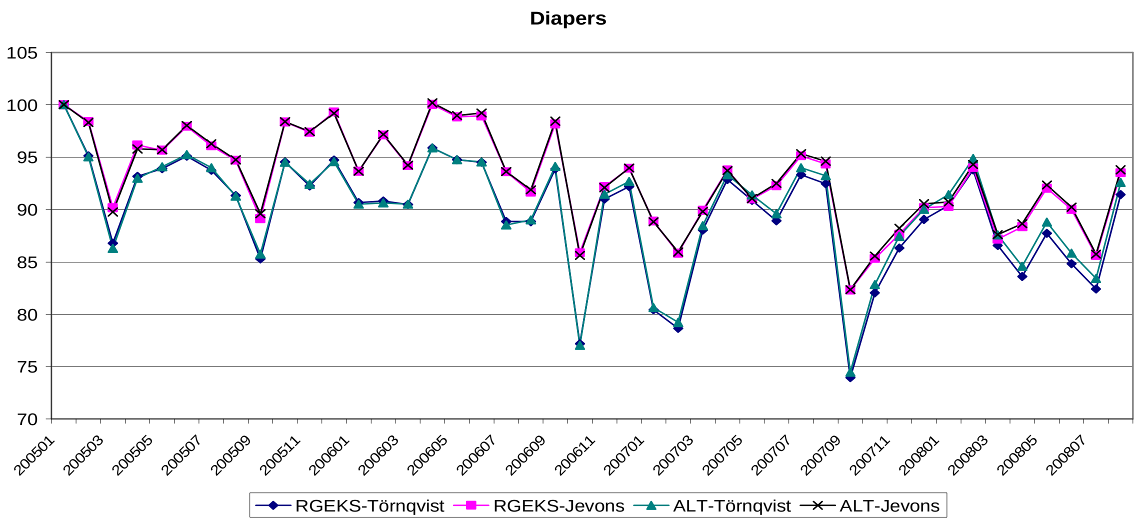
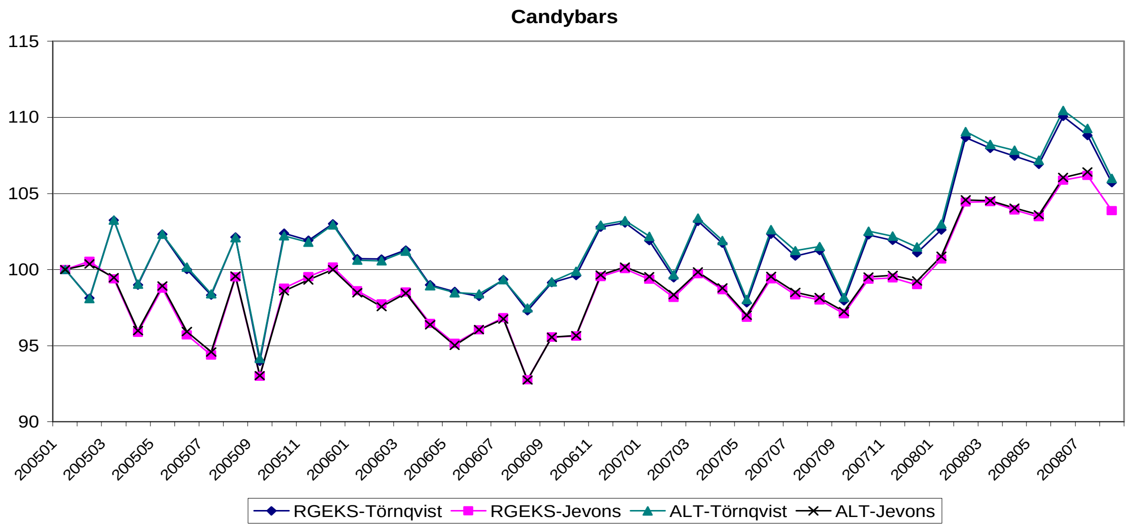


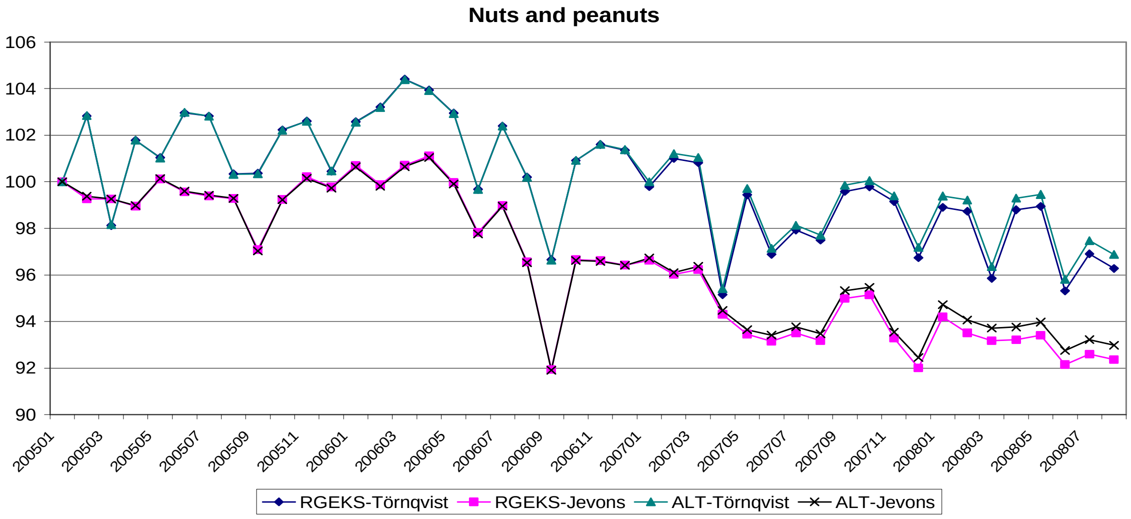
Figure 9. Rolling year GEKS-Törnqvist and GEKS-Jevons indexes

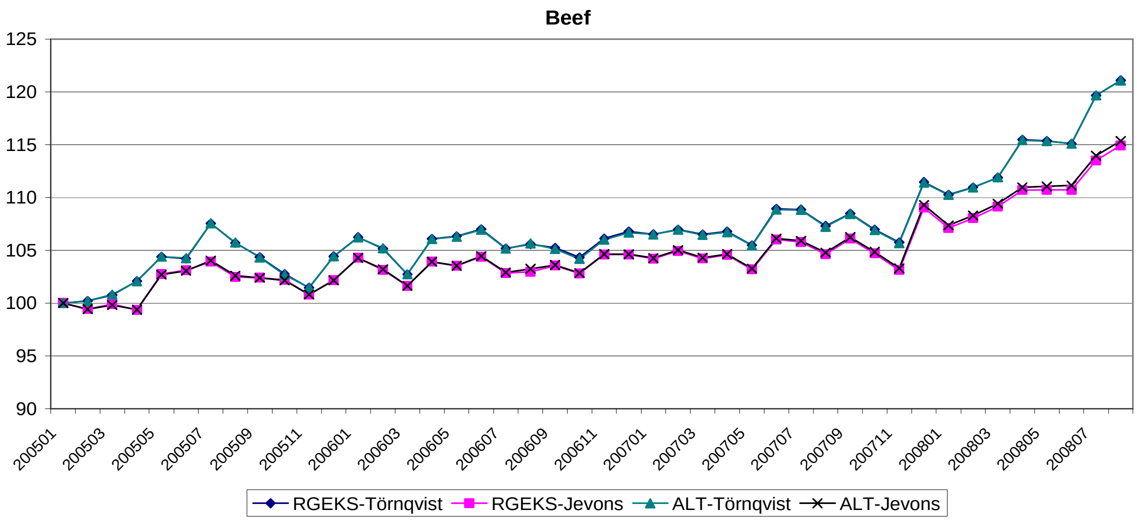







Candybars
115
110
105
100
95
90
200501
200503
200505
200507
200509
200511
200601
200603
200605
200607
200609
200611
200701
200703
200705
200707
200709
200711
200801
200803
200805
200807
RGEKS-Törnqvist
RGEKS-Jevons
ALT-Törnqvist
ALT-Jevons


Nuts and peanuts
106
104
102
100
98
96
94
92
90
200501
200503
200505
200507
200509
200511
200601
200603
200605
200607
200609
200611
200701
200703
200705
200707
200709
200711
200801
200803
200805
200807
RGEKS-Törnqvist
RGEKS-Jevons
ALT-Törnqvist
ALT-Jevons


Beef
125
120
115
110
105
100
95
90
200501
200503
200505
200507
200509
200511
200601
200603
200605
200607
200609
200611
200701
200703
200705
200707
200709
200711
200801
200803
200805
200807
RGEKS-Törnqvist
RGEKS-Jevons
ALT-Törnqvist
ALT-Jevons

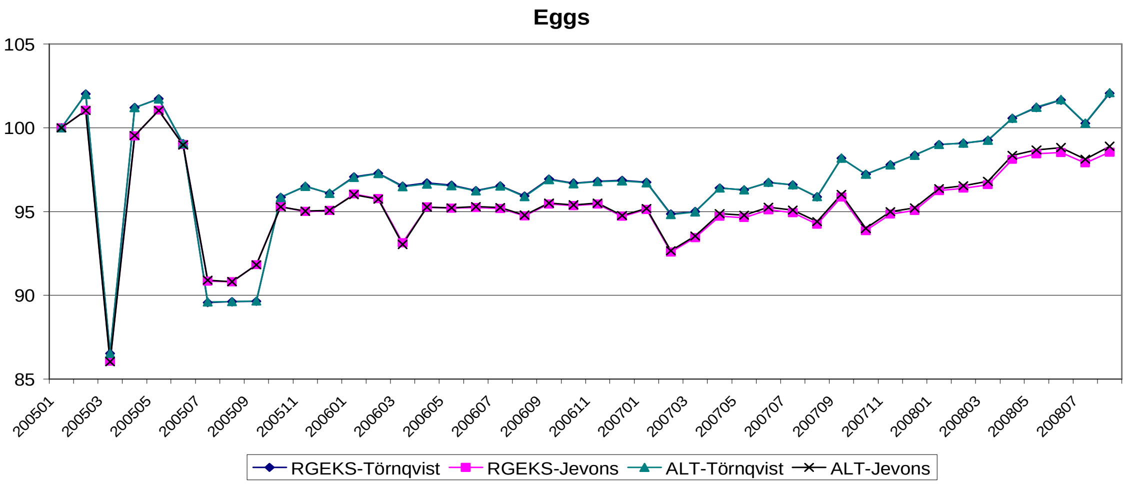
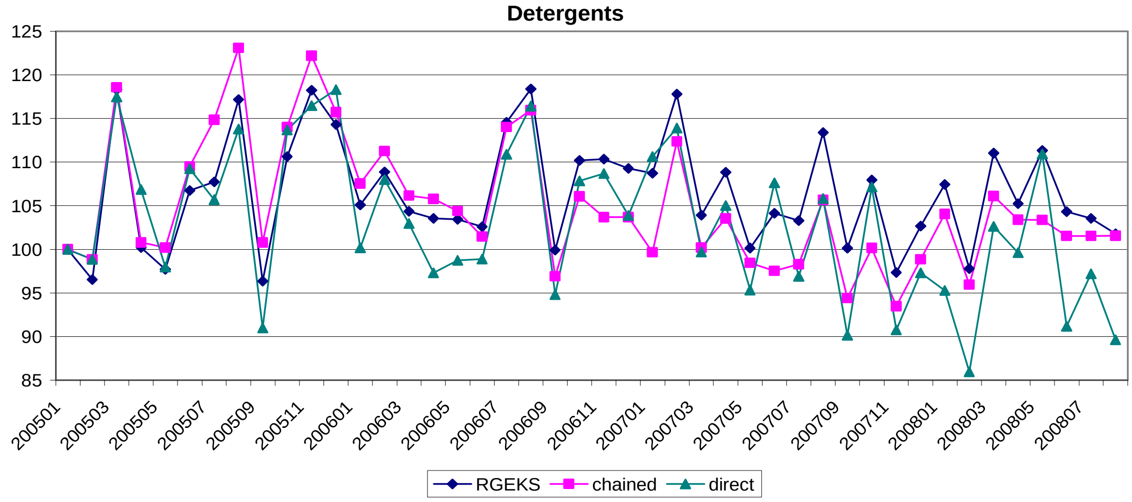
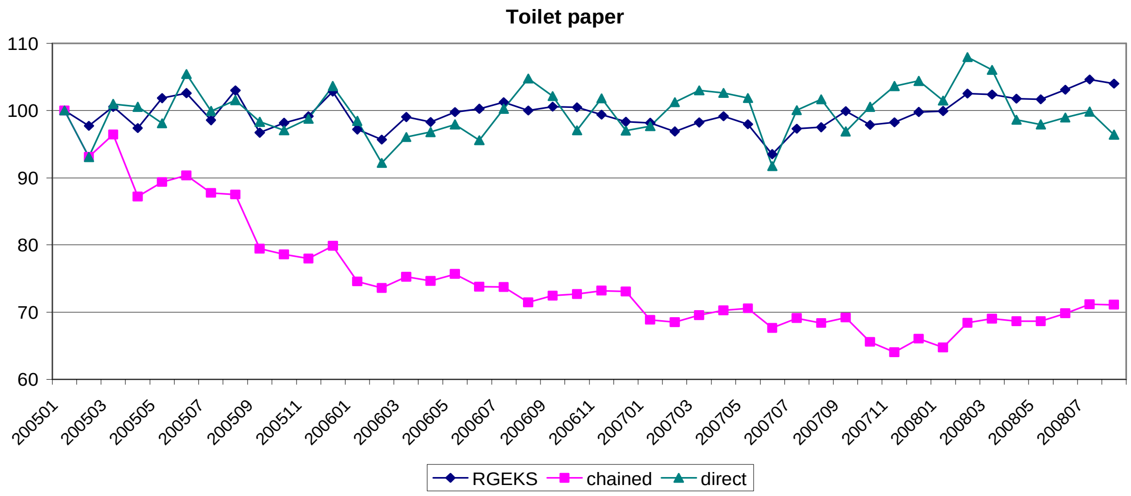
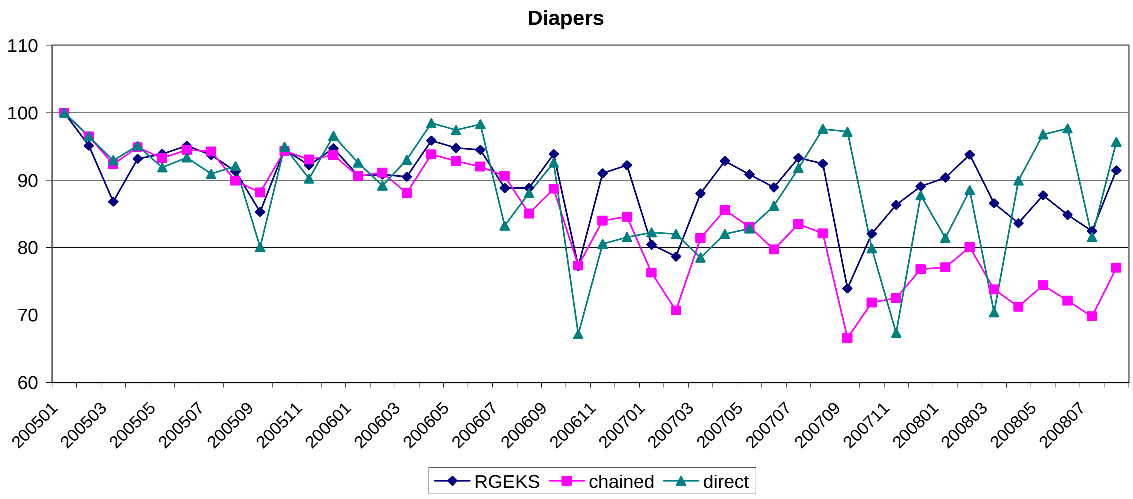
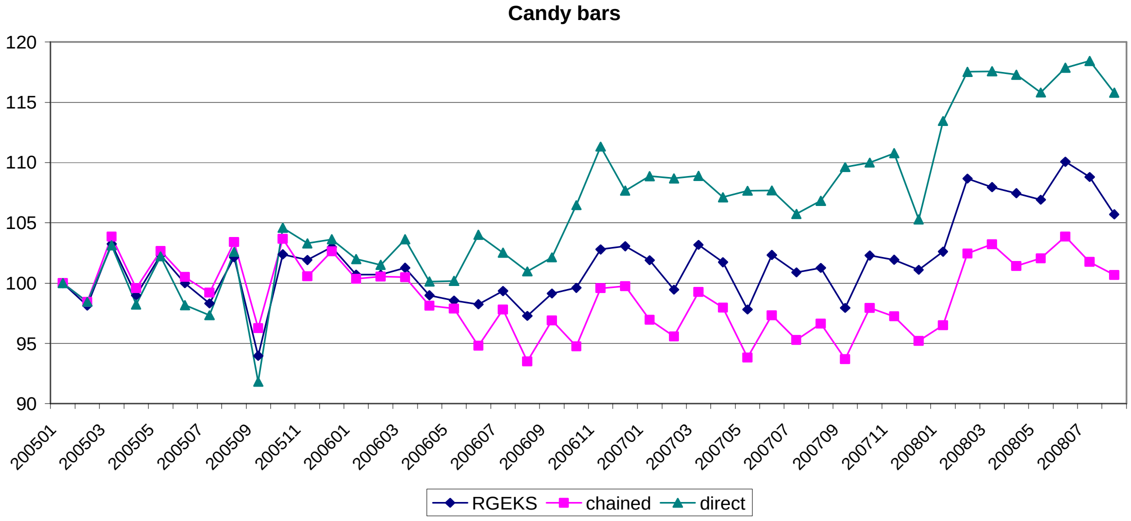

Eggs
105
100
95
90
85
200501
200503
200505
200507
200509
200511
200601
200603
200605
200607
200609
200611
200701
200703
200705
200707
200709
200711
200801
200803
200805
200807
RGEKS-Törnqvist
RGEKS-Jevons
ALT-Törnqvist
ALT-Jevons

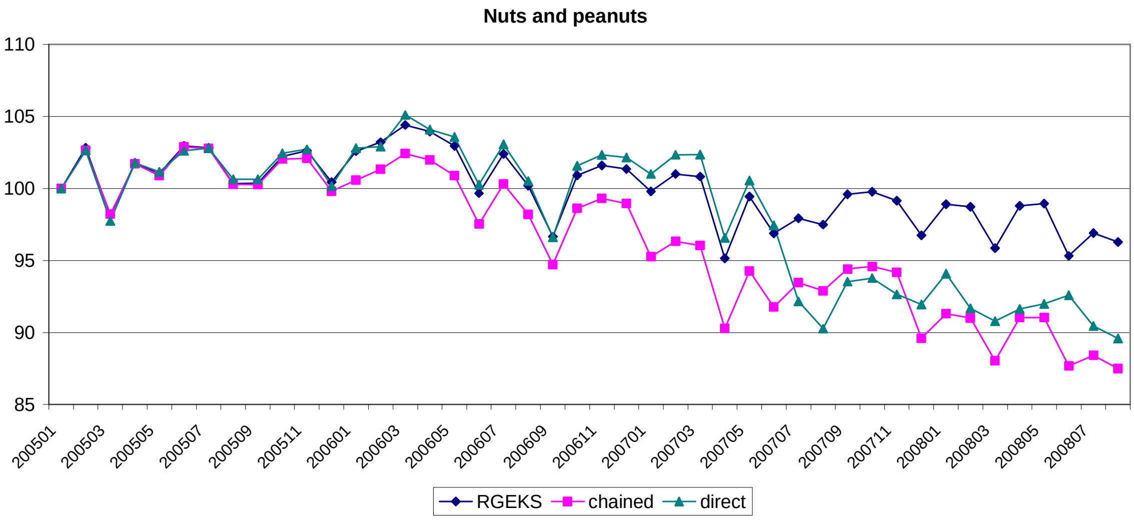
Figure 10. Rolling year GEKS-Törnqvist, chained Törnqvist and direct Törnqvist price indexes

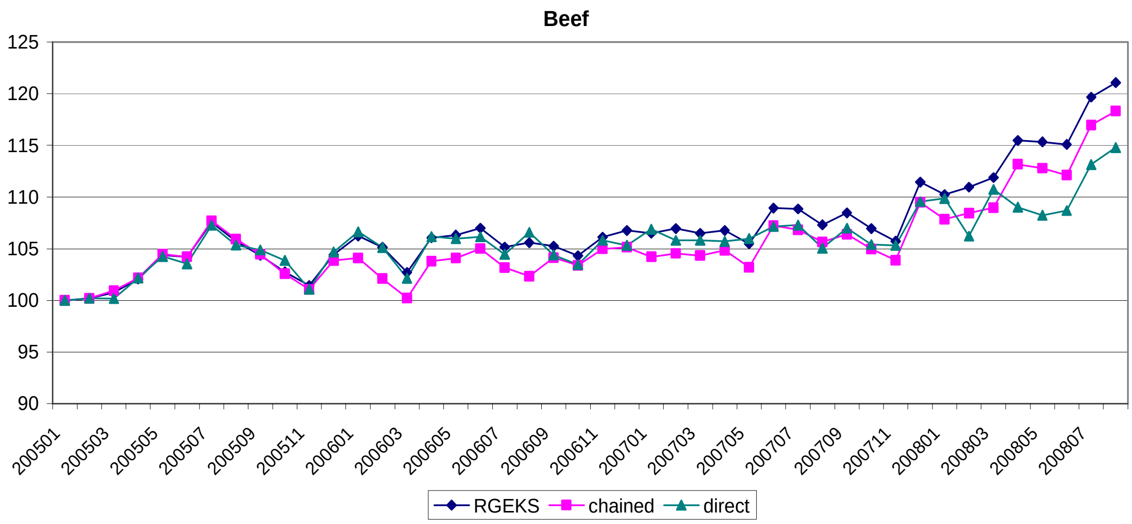







Candy bars
120
115
110
105
100
95
90
200501
200503
200505
200507
200509
200511
200601
200603
200605
200607
200609
200611
200701
200703
200705
200707
200709
200711
200801
200803
200805
200807
RGEKS
chained
direct


Nuts and peanuts
110
105
100
95
90
85
200501
200503
200505
200507
200509
200511
200601
200603
200605
200607
200609
200611
200701
200703
200705
200707
200709
200711
200801
200803
200805
200807
RGEKS
chained
direct


Beef
125
120
115
110
105
100
95
90
200501
200503
200505
200507
200509
200511
200601
200603
200605
200607
200609
200611
200701
200703
200705
200707
200709
200711
200801
200803
200805
200807
RGEKS
chained
direct

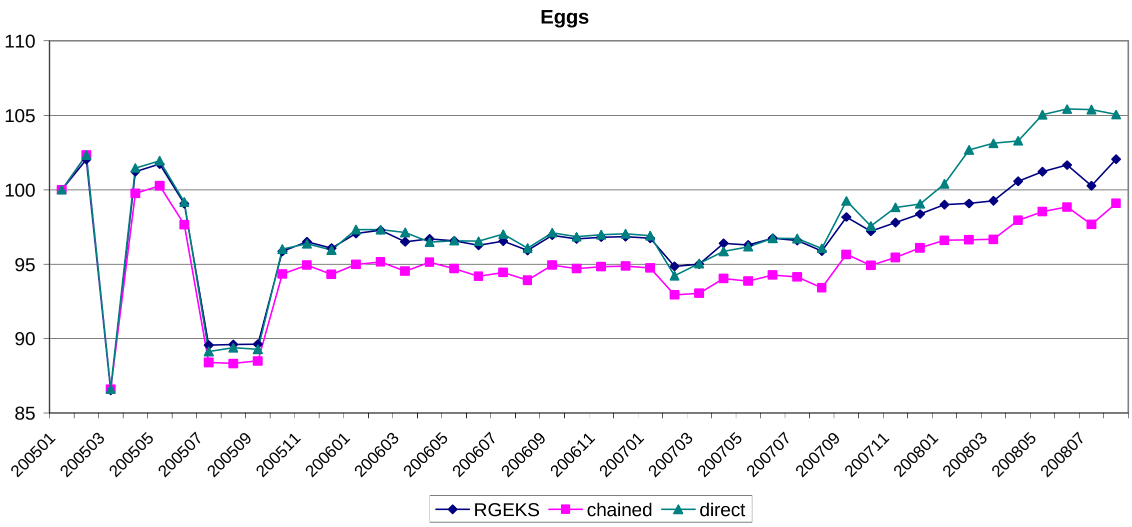


Figure 11. Chained Jevons price indexes; toilet paper

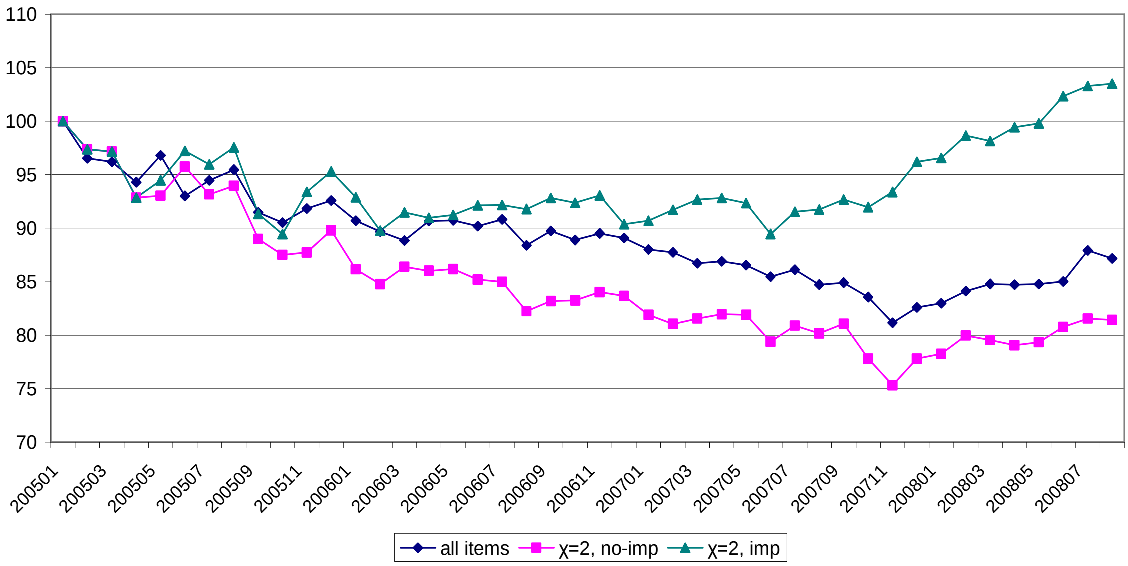
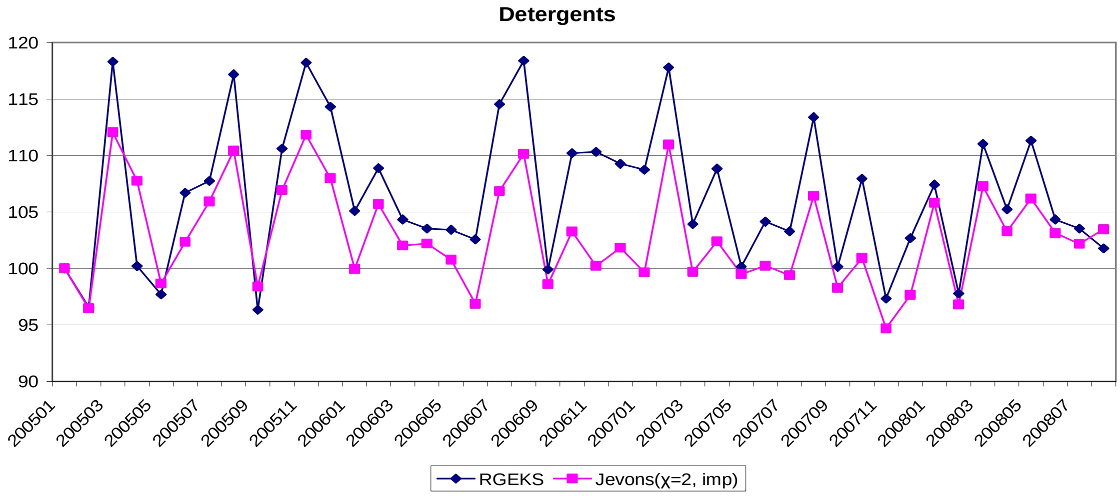
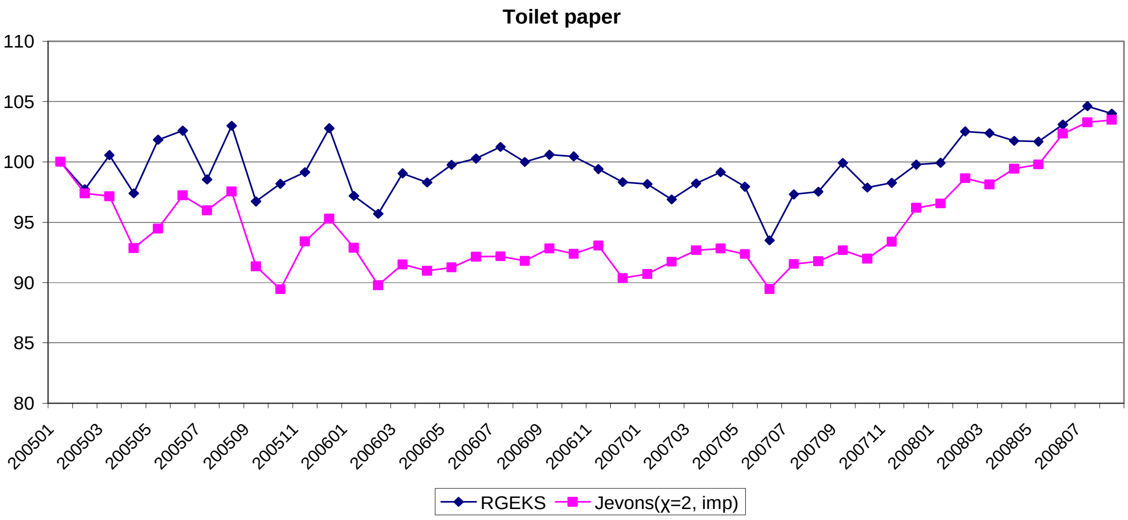
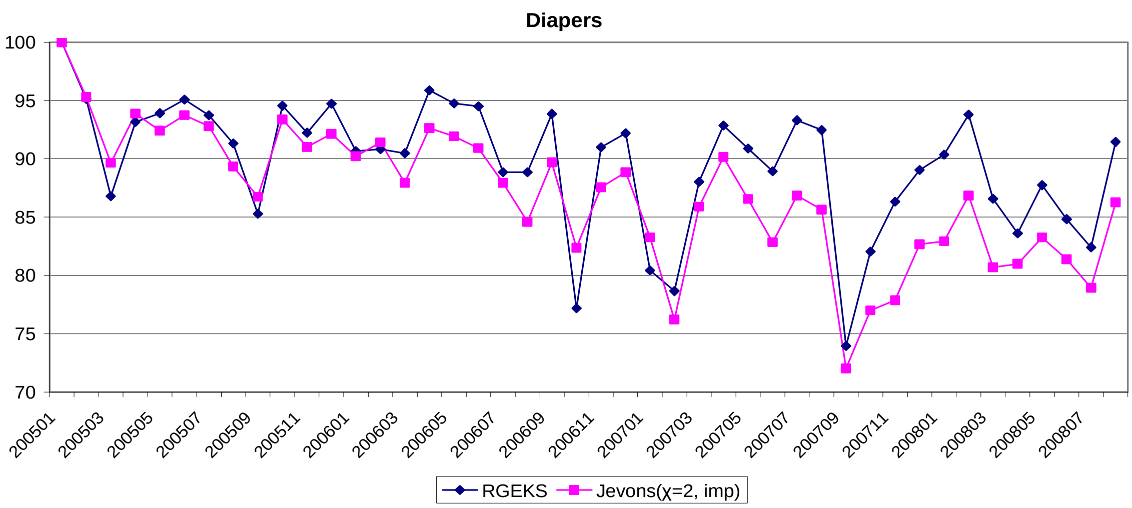
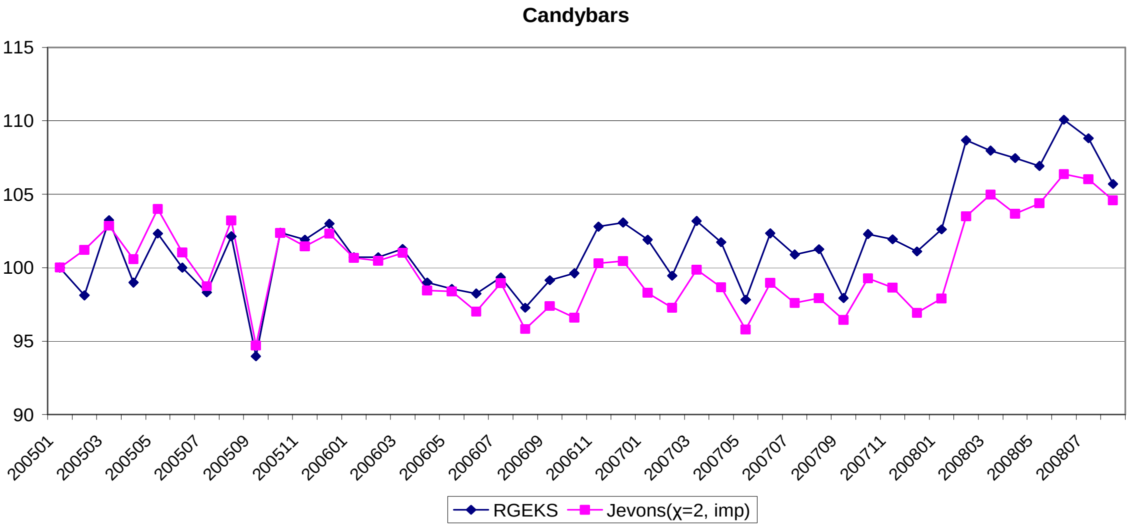


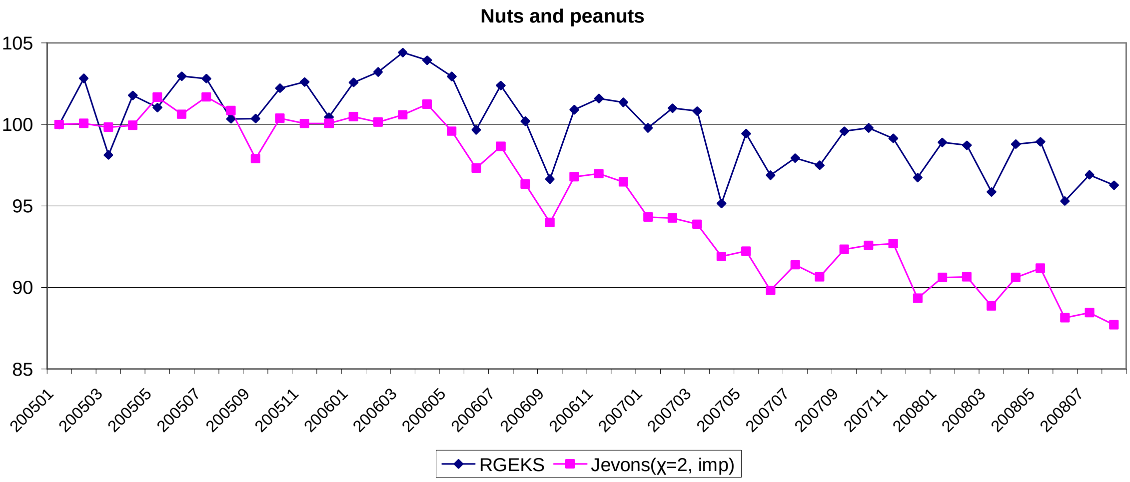
Figure 12. Rolling year GEKS-Törnqvist indexes and chained Jevons price indexes with imputations and based on a cut-off sample

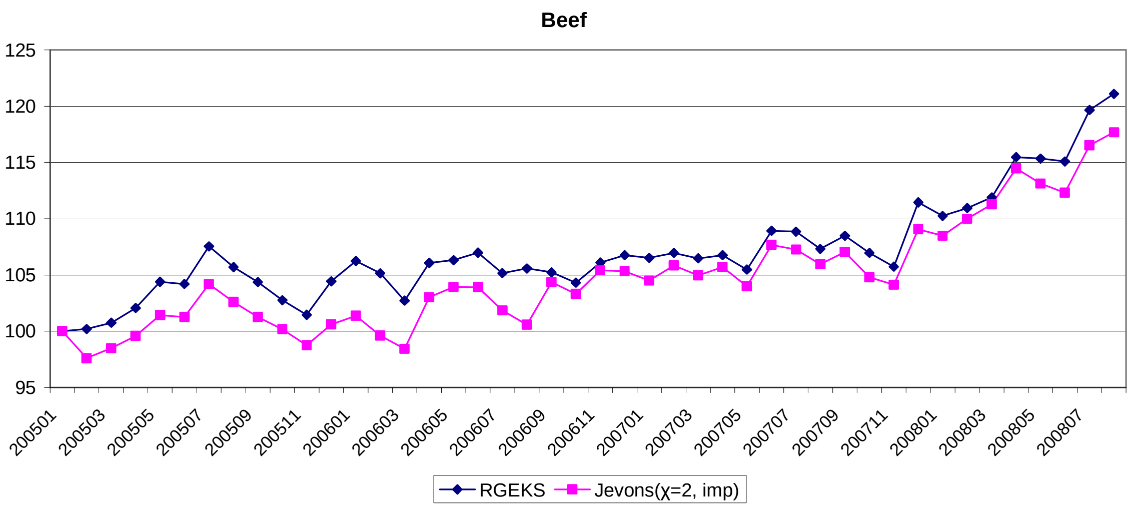







Candybars
115
110
105
100
95
90
200501
200503
200505
200507
200509
200511
200601
200603
200605
200607
200609
200611
200701
200703
200705
200707
200709
200711
200801
200803
200805
200807
RGEKS
Jevons(χ=2, imp)


Nuts and peanuts
105
100
95
90
85
200501
200503
200505
200507
200509
200511
200601
200603
200605
200607
200609
200611
200701
200703
200705
200707
200709
200711
200801
200803
200805
200807
RGEKS
Jevons(χ=2, imp)


Beef
125
120
115
110
105
100
95
200501
200503
200505
200507
200509
200511
200601
200603
200605
200607
200609
200611
200701
200703
200705
200707
200709
200711
200801
200803
200805
200807
RGEKS
Jevons(χ=2, imp)

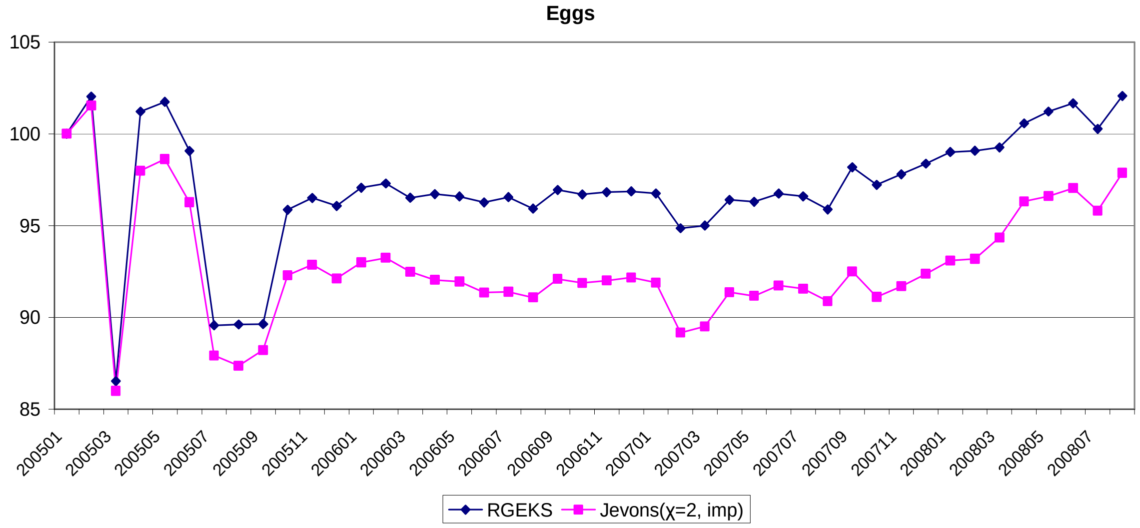


Figure 13. Monthly and quarterly rolling year GEKS-Törnqvist, quarterly chained Törnqvist and quarterly direct Törnqvist price indexes, detergents