Discussion Paper

The SPAR index and some alternative house price indices

Léon Willenborg and Sander Scholtus

November 26, 2018

Various methods are described to compute house price indices. They have in common that hypothetical selling prices are used. A hypothetical selling price of a house is an estimate of the price it would have fetched when it would have been sold in another month. These hypothetical selling prices can be compared with the actual selling price of a house and thus provide the basis for a price index. Various models are proposed to compute hypothetical selling prices. They all have in common that they use appraisal values ('WOZ valuations') of the houses sold as an auxiliary variable. Two possibilities are explored to produce a long index, namely by chaining short indices and by computing the long index directly. Furthermore the use of imputation methods is suggested to estimate selling prices. In fact, models for hypothetical selling prices can also be used as imputation models. The SPAR index emerges among the indices that are derived. Due to its special status in house price statistics it is used as a reference index, for comparison with the various indices that are derived from the hypothetical selling price approach.

1 Introduction

This paper[1] discusses various ways to compute house price indices and compares them to the SPAR index (see Eurostat (2013) or Eurostat (2017)). This index is used in the present paper as a reference index for measurement of the development of house prices. It is determined on the basis of observed prices of sold houses. The target population consists of existing properties only; new builds are excluded. The assumption underlying this approach is that the sold properties form a representative sample of the entire stock of houses.[2] It is not clear that this is a reasonable assumption. It is likely that the more popular objects tend to be sold more quickly than the less attractive ones. Or that they are sold only after a long waiting period and after a considerable drop of the initial asking price. But the owners could also decide to withdraw such an object from the market temporarily and wait for better times, rather than to sell the property at a considerable loss.

A little bit about the genesis of the present paper might be helpful to position it. First of all, it is a spin-off from an early version of Willenborg and Scholtus (2018). It started its life as an appendix of that paper, but, in due course, it grew out of proportion. It was then decided to separate it from that paper and give it a life of its own. This fledgling matured over time into what it is now.

Initially, its goal was just to provide an intuitive understanding of the SPAR index, which was not self-evident to us. A more natural choice to us seemed to be to use hypothetical selling prices, which are estimated prices for houses when they would have been sold at a different month in the reference period, coinciding with a calender year.[3] A simple model implementing this idea uses the ratio of the selling prices of houses within a certain stratum (defined by region, type of house and year-month) and their WOZ valuations. This basic model is simpler than the ones

[1] The authors would like to thank their CBS colleagues Ron van Schie and Farley Ishaak (Department of House Prices) for helpful discussions and feedback while this paper was prepared. Also we are grateful to Frank Pijpers and to Arnout van Delden (both of the Methodology Department) for reviewing mature versions of this paper.

[2] The appraisal value - or WOZ valuation, as it usually will be called here - is available for all properties and would be another useful source for this purpose. But the developments within a WOZ period - currently a year - are not covered by this valuation.

[3] In fact, it is a WOZ period, the period in which WOZ valuations do not change. Currently these periods are calender years, but in the past they were longer periods, namely 4 year periods.

typically used in the literature on house price indices, namely hedonic models, which are regression models. See e.g. Fisher et al. (2007), Devaney and Martinez Diaz (2011) and De Haan and Hendriks (2013) as a small sample from this literature.[4)]

These papers as well as Eurostat (2013) and Eurostat (2017) suggest that the preferred type of model in this area is the hedonic model. But for our purposes in the present paper these models are too complicated. We were looking for a simpler class of models to understand the SPAR index. The idea of hypothetical selling prices proved to serve us well, as it gave us what we needed, namely an intuitive idea that, among other price indices, led to the SPAR index. With some alternative models to the SPAR model available, it was natural to use them as comparisons to the SPAR index, by applying all these price indices to the same data. In particular we were curious about the cumulative effect of errors in the long SPAR index due to chaining. The hope was that the alternative models would not be affected by this phenomenon. However, we did not see dramatically diverging index series, as time advanced.

The structure of the paper before you is as follows. In Section 2 some background information on selling prices and WOZ valuations of houses is given. In Section 3 the input data that will be used for the computations of the indices is briefly described. The description focusses on those parts that are useful for the present paper, discarding the rest. Also the components of the SPAR index are discussed, which are based on selling prices or WOZ valuations. In Section 4 the SPAR index is scrutinized. This chapter also contains the germ of the idea to look for other indices, namely those based on hypothetical selling prices. These prices are used for comparison with the selling price of the same house. Comparison to the consumer price index (CPI) is enlightening. In the CPI case one is dealing with a multitude of similar products (e.g. pots of peanut butter) that are sold in quantities. But houses are essentailly unique: they cannot readily be compared with each other. Besides, once a house is sold, it will most likely not be sold within a period of several, if not many, years. So if a house is to be compared, it is best compared with itself, at different points in time. So the idea is to estimate hypothetical selling prices of houses in months other than the one in which they were actually sold. This allows one to compare the selling price of a house with virtual, or hypothetical selling prices of the same house, in different months. This requires a model to estimate these hypothetical selling prices. For this it seems natural to exploit the relationship between WOZ valuations and selling prices of houses. Such a relationship can also be used to impute a WOZ valuation in case a selling price is available but a WOZ valuation is missing.[5)] The model is then used to impute missing WOZ valuations. These issues are discussed in Section 5. Various types of models are discussed here that model the relationship between WOZ valuation and selling price. In Section 6 the estimation of hypothetical selling prices is discussed, using various models for the relationship between selling price and WOZ valuation. Section 7 discusses several price indices based on hypothetical selling prices, including robust and nonrobust variants. In Section 8 chaining is considered. This method is used for the SPAR index. Short indices defined for WOZ periods are ’stitched together’ (chained) to a long one, using overlap months of consecutive WOZ periods. For the houses in such an overlap month two WOZ valuations are available: for the previous and the current WOZ period. To produce the long index the overlap months are crucial. The longer this series gets the more uncertainty one can expect from the matching factors used in the chaining process. Section 9 discusses transitive closure that allows one to complete a partial index series, that is, one for which the index values are known

[4)] Farley Ishaak provided us with these references. We refrained from delving deeper into the literature on this subject as it seemed pointless, and a waste of time, for our rather modest goals.

[5)] Or a selling price in case this is missing and a WOZ valuation is available. But this case does not occur in practice.

for only a subset of pairs of months. In case of a SPAR short index, one only compares the value of a particular month to that of the first month of the corresponding WOZ period, for example. Using transitive closure one is able to compute missing index values for both short index series as well as the long index series. Section 10 is a switch to other types of indices, namely those that compute long index series directly, using hypothetical selling prices. Several methods can be devised that use hypothetical selling prices to compute house price indices. They differ in the choice of the periods used to compute hypothetical selling prices. Several examples are discussed explicitly, but many more can be constructed. To allow one to switch from one WOZ period to the next one is still dependent on the overlap month, as this is the only month with WOZ valuations from both periods. Again, these overlap months are crucial for the index method used, but in this case there is no cumulative effect as in case of a long index (compounding match factors). One only has to deal with a WOZ period and its immediate successor.

So far the discussion of the various methods was only theoretical and conceptual. In Section 11, however, results are shown for various methods discussed before when applied to real data. These data cover a period of about 20 years, and contain data of two regions (municipalities) in The Netherlands and 2 types of houses (apartments and terraced houses). The results shown in that chapter are intended to give an impression of the performance of the various indices. But it is apparant that some phenomena came to light that are somewhat mysterious to the present authors. These must be attributed to peculiarities of the data that were used and that we are not thoroughly familiar with. Further study would be needed to shed some light on these mysteries. Section 12 completes the main body of the paper. It discusses the most important findings and gives some suggestions for future research. The conclusion is that there is enough left to explore further. A list of references as well as an appendix with terminology and notation complement the paper.

2 Selling prices and WOZ valuations

In theory, appraisal values - to be referred to as WOZ valuations in the present paper - have a value reference date which is the first of January of the year to which they apply.[6]

For various reasons there is a *value reference date* for the valuation of all objects:

- A fixed reference date is a kind of common ideal. But one has to be realistic: it is impossible to appraise all objects on the same day,
- To have a fixed reference date when analyzing realized sales and rental prices,
- To create legal equality.

There is also a *status date* at which the value of a property is re-evaluated. This applies if changes have occurred between the value reference date and the beginning of the levy period in the sense of construction, renovation, change of destination, improvement, demolition, etc. In that situation the start of the levy period is used as the status date.

[6] see `http://wetten.overheid.nl/BWBR0007119/2016-10-01`.

It depends on the circumstances when a WOZ valuation becomes available. In practice there may be delays, so that this value is occasionally missing. It is also possible that such a valuation is incorrect, due to an administrative mistake. Or it is delayed because a property has changed and it needs to be re-evaluated.

By its very nature, one should expect this estimation to be conservative, as the owners can appeal against the WOZ valuation for their property, and it is unlikely that they accept WOZ valuations that are too high in their opinion. But it should be considered as a proxy of the value of the corresponding house at the moment it is issued, (ideally) in January each year. This value is used for the entire year in which it is issued. The next year a new valuation is made, based on new information about the property market.

WOZ valuations are estimates produced by all municipalities in The Netherlands, and they are the basis for real estate taxation.

3 The input data

In principle, every month a complete list of houses sold is available, together with some background information for each house, such as the type of property, location (municipality) and the (most recent) WOZ valuation.

The population of existing properties is divided into strata defined by municipality where a house is located and type of object (apartment / flat, (mid-)terrace house, detached house, semi-detached house, mansion, etc).

We give a description of the data that are available. The discussion is describing the data as we need them in the present paper, leaving out irrelevant details. We assume that the 13-th month approach (see section 8.2) is used to chain short series into a long one.[7)]

The months $m = 2, \ldots, 12$ are treated differently from the first month $m = 1$ each year. The reason is that each year short, yearly series of indices are produced, that are chained together to a long series, spanning several years. This is done by using an overlapping month and multiplying the new short series with a matching factor so that it matches the long series at the overlapping month. In the 13-th month approach the first months of each year are the overlapping months. How this works is explained in section 8.2.

As to the data this implies that in January the WOZ valuation of the current year, as well as that of the previous year should be stored in each record. So due to this difference we have two slightly different data sets. In both cases we are dealing with rectangular data sets, matrices if you like. In both cases, the rows correspond to the houses sold, and the columns to a number of variables. The houses are sold in a particular period (j, m), where j denotes the year and m the month. For each house sold, in principle, its selling price is known as well as its WOZ valuation which is issued in January of that year.

[7)] For the -1-st month approach the details are different, but the underlying ideas are exactly the same. The choice of the overlapping month is different. Both approaches lead to the same long index, however.

For the months $m = 2, \ldots, 12$ of each year j - the non-overlapping ones in the 13-th month approach (see subsection 8.2.1) - the **first data set** contains the following variables:

- Region where house i is located. The values of this variable are municipalities.
- Type of house i sold. The values of this variable are: apartment, terrace house, etc. The variables region and type of house together form a stratum variable. A typical stratum is denoted by g.
- Year j and month m when house i was sold. The variables j and m together define a period, that we denote as period (j, m).
- Selling price V. For house i in stratum g that is sold in period (j, m) the selling price is denoted as $V^g_{j,m,i}$.
- WOZ valuation W. For house i in stratum g that is sold in period (j, m) the WOZ valuation is denoted as $W^g_{j,m,i}$.[8)]

For month $m = 1$ of each year $j + 1$, the overlapping months, there is a separate data set, the **second data set**, that contains the following variables:

- Region where house i is located. The values of this variable are municipalities.
- Type of house i sold. The values of this variable are: apartment, terrace house, etc. The variables region and type of house together form a stratum variable. A typical stratum is denoted by g.
- Selling price V. For house i in stratum g that is sold in year $j + 1$ (in month $m = 1$) the selling price is denoted as $V^g_{j+1,1,i}$.
- WOZ valuation W. For house i in stratum g that is sold in year $j + 1$ and month $m = 1$ the WOZ valuation is denoted as $W^g_{j+1,1,i}$.
- WOZ valuation for the year previous to the year when a house was sold. For house i in stratum g sold in year $j + 1$ (in month $m = 1$) the value is denoted as $W^g_{j,1,i}$.

3.1 The components of the SPAR index

The SPAR index starts in month $m = 1$ at the value 1, for each short sequence that corresponds with a WOZ period.[9)] Therefore to start the short series for year $j + 1$ there is no reason to keep the WOZ valuation of this year in the second data set. But to compute the so-called matching factors (see (46)) the WOZ valuations of the year $j + 1$ have to be known.[10)] Also to determine outliers in the overlapping months one needs the two WOZ valuations. See section 5.1 for details.

The SPAR index for stratum g and year j and month m can be written as the following fraction:

$$s^g_{j,m} \triangleq \frac{\overline{V^g_{j,m}} / \overline{W^g_{j,m}}}{\overline{V^g_{j,1}} / \overline{W^g_{j,1}}}. \tag{1}$$

[8)] The WOZ valuations are issued every year by the municipalities in January of each year and are 'valid' until the next valuation.

[9)] Currently WOZ periods are one year long. In the past these periods used to be longer.

[10)] How the linking of short series into a long one is actually done is explained in subsection 8.3.

The overlines in 1 denote averages over houses in the various strata that are being distinguished: sold properties in month m, say, where the selling price ánd the appropriate WOZ valuation are known and are considered plausible. Note that the series starts with the value 1 each year j, that is $s^g_{j,1} = 1$, for each j and g.[11)]

We now take a closer look at the components of the SPAR index (1). Because of the appearance of missing values for selling prices and WOZ valuations there are different subsets of non-missing records to take into account.

We start with the set of records for year j and month m: $H^g_{j,m}$. The set of complete records with respect to the selling price is denoted by $HV^g_{j,m}$, with $|HV^g_{j,m}| = NV^g_{j,m}$. The set of complete records with respect to the WOZ valuation is denoted by $HW^g_{j,m}$, with $|HW^g_{j,m}| = NW^g_{j,m}$. Special cases pertain to the first month of the year, that is with $m = 1$.

We are now in the position to explain $\overline{V^g_{j,m}}$ and $\overline{W^g_{j,m}}$:

$$\overline{V^g_{j,m}} \triangleq \frac{1}{NV^g_{j,m}} \sum_{i \in HV^g_{j,m}} V^g_{j,m,i} \tag{2}$$

and

$$\overline{W^g_{j,m}} \triangleq \frac{1}{NW^g_{j,m}} \sum_{i \in HW^g_{j,m}} W^g_{j,m,i}. \tag{3}$$

Of course, $\overline{V^g_{j,1}}$ and $\overline{W^g_{j,1}}$ are special values of (2) and (3).

4 A closer look at the SPAR index

4.1 Motivation

In this section we look more closely at the SPAR index, and more widely to the SPAR method. The former refers to the formula used, the latter to the way the index is applied, in particular in view of the treatment of missing values. In the present section we only bring a few critical points to attention. They will be elaborated in subsequent sections.

[11)] Under the assumption that houses have been sold in stratum g in period $(j, 1)$.

4.2 Selling price and WOZ valuation

The SPAR index uses both selling prices and WOZ valuations. From the way they are used in the SPAR index formula it is not immediately clear what their roles are. Selling prices and WOZ valuations are clearly related variables, but they are not the same. A WOZ valuation is a proxy to the selling price of a house. It is based on exterior properties of houses: roof extensions are included in WOZ valuations, the quality of the kitchens are not. Of course, interior properties of a house (state of maintenance of the interior, quality of the kitchen, the bathroom(s), etc.) affect the selling price, but not the WOZ valuation. As a WOZ valuation is issued once a year[12] - in January[13] - one can expect WOZ valuations and the selling prices of the houses sold in January to be closest. Also, one may expect the discrepancy between the selling price and WOZ valuation of a house to increase, the later in the year it is sold. This is due to two effects: the changes in the housing market (external factor) and the possible changes of the houses (internal factor). This latter effect is particularly noticeable for houses that have been renovated or extended in the period after the last WOZ valuation and the sale of the house. In fact, if something like this has happened we are dealing with a house with different properties before and after the renovation.[14] It is clear that in this case the ’old’ WOZ valuation (before the transformation) and the selling price may be quite different. The re-valuation and the selling price should be much closer.

4.3 Description of the SPAR index

The SPAR index first derives yearly short series, which are then ’stitched together’ into a long series. In the present section we consider how the yearly short series are computed. How these are chained into a long series is a separate issue that is discussed in subsection 8.3.

As was remarked earlier, the problem with defining a price index for houses is that each house is, in a sense, unique. And also, a house is usually not sold twice in a year. So in order to compute an indicator for house price development one has no selling prices to compare. A ’trick’ is needed: hypothetical selling prices. This exploits the existing combinations of WOZ valuations and selling prices. The former can be viewed as proxies to selling prices. So the idea is to consider, within a WOZ period, say a year, j, the average selling prices of houses sold in month m, to the corresponding WOZ valuations of these houses which are a kind of virtual selling prices at month $m = 1$ of year j.

4.4 Form of the SPAR index

The form of (1) does not make it immediately transparent that we are dealing with a price index that gives information of price developments of selling prices of houses, as it not only involves selling prices but also WOZ valuations, which are appraisals and not actual selling prices. So in the author’s opinion it should be understood why (1) is a reasonable choice for a price index for house prices.

[12] This is true since 2008. Before that the WOZ valuation was renewed once in several years.

[13] In most cases. Some houses may be re-valued during the year due to a major change in the house (renovation, extension, etc.).

[14] Or expressing it differently: the quality of the house has changed.

In view of the situation with houses and in particular the fact that they are unique (to a high degree) and are not sold with high frequency, the following type of index method would seem reasonable: compare the selling price of a house with a hypothetical selling price of the same house at the beginning of the year. The WOZ valuations provide the opportunity to take such an approach. This would result in the following ingredients for a price index for houses:

$$\frac{\text{average selling price in stratum } g \text{ in period } (j,m)}{\text{average selling price in stratum } g \text{ in period } (j,1)}, \tag{4}$$

where, for the moment, we abstract from the way the averaging is done. We can rewrite (1) so that it conforms to (4) and we obtain for the SPAR index

$$s^g_{j,m} = \frac{\overline{V^g_{j,m}}}{\widehat{V^g_{j,1}}}, \tag{5}$$

where

$$\widehat{V^g_{j,1}} \triangleq \frac{\overline{W^g_{j,m}}}{\overline{W^g_{j,1}}}\overline{V^g_{j,1}}. \tag{6}$$

The factor

$$\frac{\overline{W^g_{j,m}}}{\overline{W^g_{j,1}}} \tag{7}$$

is apparently intended as a correction factor for the average selling price $\overline{V^g_{j,1}}$ in month 1. It is not clear (to the present authors) what this factor (7) is actually correcting for. Why should the average WOZ valuations of the houses sold in month $m > 1$ in year j be systematically different from those sold in January of that same year? We would expect the averages of the WOZ valuations to be roughly the same. As a consequence, the factor (7) is likely to behave as a random variable fluctuating around 1.

So in case of the SPAR index the following question raises itself: Why not simply take $\overline{V^g_{j,1}}$ instead of (6)? Then the SPAR index would simplify to the following index formula:

$$\hat{s}^g_{j,m} = \frac{\overline{V^g_{j,m}}}{\overline{V^g_{j,1}}}, \tag{8}$$

which does not use WOZ valuations. We may expect this index to fluctuate as the houses 'sampled' in each month can be expected to be independently 'drawn'. We do not consider this a serious index, but we have used it in some graphs in the sequel just to see how it performs compared to other, more serious house price indices.

Now (4) is based on aggregates. We want to explore in the sequel what price indices we get if we look at ratios of selling prices of houses and hypothetical selling prices for month one of the same year these houses were sold. For a house i in stratum g sold in month m of year j:

$$\frac{\text{selling price of house } i \text{ in stratum } g \text{ in month } m}{\text{hypothetical selling price of house } i \text{ in stratum } g \text{ in month } 1}. \tag{9}$$

The hypothetical selling price applies to January of the year j, the same year as house i in stratum g was sold. It will appear that WOZ valuations are crucial in estimating hypothetical selling prices (see Section 6).

It will be shown in Section 7 that a similar index to the SPAR index can be derived based on (9). In this index, as in the original SPAR index, the numerator and denominator of (9) are aggregated separately, before taking the ratio; cf. (4). Alternative indices will also be obtained in Section 7 by aggregating the ratio (9) directly. For a given set of hypothetical selling prices, these two approaches in general do not yield identical indices. In Section 11, indices based on both approaches will be compared empirically.

4.5 Using completed data

The SPAR method as it is currently applied simply ignores any missing values in the data. However, this is certainly not necessary. It is also possible to apply an imputation procedure in case either the selling price or the WOZ valuation is present, and the other variable is missing in a record.[15] We propose to use a simple imputation model. The records where both values are missing are useless for index computation and they can be discarded from the start. The records with exactly one of these two quantities missing are the ones that we focus our attention on. The missing values in these records will be imputed, before the index computations start. So we may assume that for these computations we have a complete or completed file.[16]

4.6 Estimates for hypothetical selling prices

Another possible modification of the SPAR index concerns the use of both selling price and WOZ-value in the SPAR index formula. We propose a modification of the approach taken in the SPAR index, in the sense that the selling price of a house i sold in month m of year j is compared with a hypothetical selling price of i in month $m = 1$ in year j. In case of the SPAR index, the selling price of house i is compared to its WOZ valuation in the same year. But in our view, selling price and WOZ valuations are different variables, and should not be confused. A WOZ valuation is a proxy to a selling price, but it is not a substitute for selling price. So we propose to use the WOZ valuation of a house i to estimate a hypothetical selling price for month 1 of the same year, as they are correlated. Information about selling prices and WOZ valuations of houses sold in month 1 in stratum g can be used to estimate the relationship between these variables.

[15] As remarked earlier, WOZ valuations may be missing in the data, but not selling prices.

[16] We assume that the records with both values missing have been discarded before.

5 Imputing missing selling prices

In the SPAR method missing values are ignored. No imputation is used. The computation of average selling prices of WOZ valuations only uses regular (non-missing) data.

It should be noted that a record originally may have had a non-missing, regular, value. But during data editing one of the values in such a record was considered an outlier, and its value was replaced by a missing value. There are various reasons why missing values occur, or outliers. Selling prices or WOZ valuations may not be known at the time of reporting. Or an administrative error was made in specifying the selling price or the WOZ valuation. It may also occur that the WOZ valuation applies to a partly built house, and the entire record is rejected because the house is a new build, which is outside the target population, so the entire record is discarded. Instead of discarding incomplete records we can impute the missing values, and use the completed records in the price index computations as well.

The easiest assumption about missing values in data sets is that they are caused by a random, rather than a systematic, process. Serious biases could result from ignoring the nonresponse results requiring a fix (an imputation). However, fixing them needs an imputation model describing how the ’fixes’ should be computed from observed values.

Imputation is a processing step that should be carried out prior to the computation of the price indices. The computation of the SPAR indices is exactly as described in Section 3. Howeverut, some of the values used for the selling price or the WOZ valuation may be imputed instead of observed. For the computation of the price index this is no problem. When computing variances, one should keep in mind that some values are imputed; they are in fact random variables and contribute to an increase of the variance.

There are several ways in which missing selling prices or WOZ valuations can be imputed. We discuss some of them in the following subsections of the present section.

5.1 Missing values and outliers

5.1.1 Non-overlapping months

The input data at a particular moment in time may have missing WOZ valuations for certain records, due to administrative delays.[17] The input data are checked, in particular with respect to the ratio of the selling price of a house and the WOZ valuation. In house a record contains both a selling price V and a WOZ valuation W, in order to be acceptable it is required that $V, W > 0$ and furthermore[18]

[17] At a later time these values may become available. But at the moment we consider the state of the input data at a particular moment in time. Selling prices may also be missing but they do not enter the input data set for the SPAR index.

[18] This is what we had been told. However Le (2014, Section 4) mentions a more complicated formula, that corrects for the development of the SPAR index between two periods. We learned about this outlier criterion when the present document was almost finished. As the time was up we decided to keep this section of the document as it was, and in particular (10). Only the current footnote was added as a warning sign for the reader who wishes to delve deeper into these matters. For our approach there were no consequences, as we only used the indicator variables present in the data to identify outliers.

$$\frac{1}{2} \leq \frac{V}{W} \leq 2. \tag{10}$$

If a record violates requirement (10) at least one of the values V or W is incorrect, and considered an outlier. The department for house price statistics decides, which of these values is to be considered an outlier.[19)]

Obviously, we can rewrite (10) as

$$\frac{W}{2} \leq V \leq 2W, \tag{11}$$

which is a more convenient form, consisting of two linear inequalities.

5.1.2 Overlapping months

In some months, overlap is created between a short series and a previous one (or the long index up to and including this year) in order to extend the long index with the information for one year. In such overlapping months there are two WOZ valuations to reckon with, say W_1 and W_2, for year j and year $j+1$, respectively. The two conditions to take into account for such overlapping months:

$$\frac{W_1}{2} \leq V \leq 2W_1 \tag{12}$$

and

$$\frac{W_2}{2} \leq V \leq 2W_2. \tag{13}$$

So the following situations are possible for a selling price V in an overlapping month:

1. a selling price V is not an outlier with respect to either W_1 or W_2. In this case V can be used for computation of the SPAR index for the short series for j and $j+1$;
2. a selling price V is an outlier with respect to W_1, but not with respect to W_2. In this case V is is only used in the short series of year $j+1$ but not for that of year j;
3. a selling price V is an outlier with respect to W_2, but not with respect to W_1. In this case V is is only used in the short series of year j but not for that of year $j+1$;
4. a selling price V is an outlier with respect to both W_1 and W_2. In this case V is not used for index computations in year j or year $j+1$.

[19)] If both are considered incorrect, the record is of no use for index computations and can be discarded.

In the first case V in fact satisfies the condition

$$\frac{1}{2}\max\{W_1, W_2\} \leq V \leq 2\min\{W_1, W_2\}. \tag{14}$$

In fact the first case above and the condition (14) applies to the selling prices in the overlapping months to be acceptable. This shows that the conditions for records being acceptable in overlapping months are stricter than for those in the remaining months of the year.

In the current computation of the SPAR index the selling price V is checked against the WOZ valuation for the corresponding short series. Therefore it may, in principle, occur that situations 2 or 3 apply, so that different sets of records for an overlapping month are used for computing the short series of year j and year $j+1$.

In case missing values (original ones or induced ones, to replace outliers) are imputed, these problems disappear automatically.

It should be noted that if we have a set of pairs $\{(V_1, W_1), \dots, (V_n, W_n)\}$ that satisfy (10), then so also do the arithmetical and geometric averages of the ratios:

$$\frac{1}{2} \leq \sum_{i=1}^{n} \frac{V_i}{W_i} \leq 2, \tag{15}$$

and

$$\frac{1}{2} \leq \sqrt[n]{\prod_{i=1}^{n} \frac{V_i}{W_i}} \leq 2. \tag{16}$$

This is a useful observation in the light of imputing missing WOZ valuations (or selling prices).

5.2 Using averages of ratios

If we look at formula (1) we see that only the complete records in the data are used. But there is an alternative in case the selling price of a house is known and the WOZ valuation is missing. In that case we can try to impute the missing value by using the known selling price as an auxiliary variable. To be able to do this, we assume a simple imputation model, such as the one that estimates the ratio of the average selling price and the WOZ valuation per stratum g, using the complete records. The underlying assumption is that the missing values are not systematically generated, but randomly, independent of the original value. For stratum g we then would have the average ratio[20)]

[20)] We take the arithmetical average here, but other averages like the geometric or harmonic averages (etc.) are possible as well; or robust ones, like the median.

$$\hat{\phi}^g_{j,m} \triangleq \frac{1}{NVW^g_{j,m}} \sum_{i \in HVW^g_{j,m}} \frac{V^g_{j,m,i}}{W^g_{j,m,i}}, \tag{17}$$

where $HVW^g_{j,m}$ is the set of all houses in stratum g sold in year j and month m for which both selling price $V^g_{j,m,i}$ and WOZ value $W^g_{j,m,i}$ are available (i.e., $HVW^g_{j,m} = HV^g_{j,m} \cap HW^g_{j,m}$) and $NVW^g_{j,m} = |HVW^g_{j,m}|$. Of course, to use the stratum g as a resource of properties to compute $\hat{\phi}$ is a choice and not a necessity. Some other subset including g could be chosen instead. Or perhaps a subset of g, if one has extra characteristics of houses available. In fact the imputation could be applied in two cases, namely those in which either the price of a house or its WOZ valuation is missing, but not both.[21)]

In case a record i in stratum g for which the selling price is available as $V^g_{j,m,i}$ but its WOZ valuation $W^g_{j,m,i}$ is missing, we can estimate the latter by

$$\widehat{W}^g_{j,m,i} \triangleq \frac{V^g_{j,m,i}}{\hat{\phi}^g_{j,m}}. \tag{18}$$

In case the WOZ valuation $W^g_{j,m,i}$ is present but the selling price $V^g_{j,m,i}$ is missing, we can estimate the latter by

$$\widehat{V}^g_{j,m,i} \triangleq \hat{\phi}^g_{j,m} W^g_{j,m,i}. \tag{19}$$

For our data, the second case is hypothetical, but it could be handled just as easily as the first case.

So after application of this imputation technique the set of records with either selling price or WOZ valuation missing (and the other value present) can be considered complete.[22)]

5.3 Using ratios of averages

Instead of the method in subsection 5.2 it is also possible to use ratios of averages, as in the SPAR index itself are used. Instead of (17) we can use

$$\hat{\chi}^g_{j,m} \triangleq \frac{\overline{V^g_{j,m}}}{\overline{W^g_{j,m}}} = \frac{NW^g_{j,m}}{NV^g_{j,m}} \frac{\sum_{i \in HV^g_{j,m}} V^g_{j,m,i}}{\sum_{i \in HW^g_{j,m}} W^g_{j,m,i}}. \tag{20}$$

[21)] In case both WOZ valuation and selling price of a house are missing in a record, the entire record is discarded as useless for index computation.

[22)] If there are enough records to make the imputation work, which we shall assume to be the case.

Note that the sets over which is summed in numerator and denominator of (20) are different.[23] This is a result of the appearance of missing values for selling prices or WOZ valuations. Instead of summing over different sets of houses in numerator and denominator, one could sum over the same set, namely $HVW^g_{j,m}$ and obtain

$$\hat{\chi}^g_{j,m} \triangleq \frac{\sum_{i \in HVW^g_{j,m}} V^g_{j,m,i}}{\sum_{i \in HVW^g_{j,m}} W^g_{j,m,i}}, \tag{21}$$

which is probably to be preferred to the $\hat{\chi}^g_{j,m}$ in (20). If the overlapping set $HVW^g_{j,m}$ is small compared to $HW^g_{j,m}$ or $HV^g_{j,m}$ then (20) may seem to be a better choice than (21) because the summing in numerator and denominator is over bigger sets, but they have a small intersection, namely $HVW^g_{j,m}$.

The mechanics to impute missing values is now essentially the same as in subsection 5.2, except that the χ's have to be used instead of the ϕ's. After the application of imputation, again, the set of records with either selling price or WOZ valuation missing (and the other value present) can be considered complete.[24]

5.4 Using PCR

Linear regression analysis can be used to model the (assumed) linear relationship between selling price and WOZ valuation. This is an extension of the assumption of a fixed ratio between the two quantities, per stratum, as assumed above.

We propose principal component regression (PCR), in order to keep the problem symmetric: we are interested in the relationship between selling prices and WOZ valuations. We do not want to model as in linear regression, where one variable is viewed as a linear function of the other. This approach would destroy the symmetry of the roles played by both variables. PCR preserves this symmetry.

Before we start with PCR it is convenient to transform the data. We want to shift the center of gravity of the data cloud of $X^g_{j,m,i} \triangleq (V^g_{j,m,i}, W^g_{j,m,i})$ to the origin. The center of gravity $Z^g_{j,m}$ is

$$Z^g_{j,m} \triangleq \frac{1}{NVW^g_{j,m}} \sum_{i \in HVW^g_{j,m}} X^g_{j,m,i}. \tag{22}$$

If we map $X^g_{j,m,i} \mapsto X^g_{j,m,i} - Z^g_{j,m} \triangleq Y^g_{j,m,i}$ then $\frac{1}{NVW^g_{j,m}} \sum_{i \in HVW^g_{j,m}} Y^g_{j,m,i} = 0$.

PCR is achieved as follows. Put these values in a matrix $Y^g_{j,m}$ of size $NVW^g_{j,m} \times 2$. Then $Y'Y$ is a 2×2 positive, symmetric matrix. This is diagonalizable with real, nonnegative eigenvalues. We

[23] Or rather, they are not necessarily the same. Records where both the selling price and the WOZ valuation are missing or look suspicious, have already been eliminated, as they are of no use.

[24] Assuming that there are enough records to make the imputation work, which we shall assume for the moment.

are interested in the eigenvector corresponding to the largest eigenvalue. This gives the direction of the main axis of the cloud of points, consisting of pairs of selling price and WOZ valuation for houses i sold in year j and month m in stratum g.

The idea of PCR regression is to map a point in the data set to a point on the regression line by orthogonal projection. This yields the point on this line with the shortest distance to the data point. An easy way to obtain such projection is by first rotating the regression line so that it coincides with the x-axis. Let R denote this rotation. Apply R to the data points as well. Orthogonally projecting a point on the x-axis is easy: if (x, y) is such a (transformed) point, its orthogonal projection on the x-axis is $(x, 0)$. If we rotate this point with the reverse rotation R^{-1} we have obtained the orthogonal projection of the original data point: $R^{-1}(x, 0)'$ is the orthogonal projection of $R^{-1}(x, y)'$ on the PCR regression line.

6 Estimating hypothetical selling prices

Ideally we would compare the selling price of a house i sold in month m of year j with its selling price at month 1 in the same year j. But this is only very rarely possible, as properties are typically not sold twice a year. So we propose to estimate a hypothetical selling price of this house in month 1 of year j.

This involves two ideas. The first one is to use the WOZ valuations of the houses i selling in month m of year j. They provide a link to month 1 of year j for these house. The second idea is to use the WOZ valuations $W^g_{j,1,i}$ of these houses to find hypothetical selling prices, using a relationship between selling prices and WOZ valuations in January. This is the same idea as was used in the imputation method in section 5. Several methods are available to do this. We discuss three of them.

6.1 Using ϕ-factors

We want to use the same ϕ-factors as in subsection 5 to link the WOZ valuations to selling prices, but now for all records, not only for those with missing values. In particular we are interested in $\hat{\phi}^g_{j,1}$ as in (17).

Using this factor we then obtain estimates of hypothetical selling prices for properties sold later in the year. That is, an estimate of the hypothetical selling price for month 1 of year j for house i that was sold in month m in year j is:

$$\widehat{V}^g_{j,1,i} \triangleq \hat{\phi}^g_{j,1} W^g_{j,m,i}, \tag{23}$$

where $W^g_{j,m,i}$ is the WOZ valuation for this house i in stratum g sold in period (j, m).

Note that the $\hat{\phi}^g_{j,1}$'s are estimates based on houses sold in period $(j, 1)$, where selling prices and WOZ valuations can be expected to be most strongly correlated in that year.

6.2 Using approximate matching

Instead of the method in (23), we now suggest a method based on approximate matching. In this case the WOZ valuation of a house $i^g_{j,m}$ in stratum g sold in month in period (j,m) is used as a 'linking pin' to selling prices of houses sold in period $(j,1)$, as follows. The idea is to look at WOZ valuations of houses $i^g_{j,1}$ in stratum g sold in period $(j,1)$ which have roughly the same WOZ valuations. We then take the average of the selling prices of these houses and use this as an estimate for the hypothetical selling price in period $(j,1)$ of house $i^g_{j,m}$. Of course, one has to specify what is close enough. We shall not elaborate this method here as we have sufficient alternatives.

6.3 Using linear regression

Here we propose to use linear regression, where the selling price is the dependent variable and the WOZ valuation the independent variable. The symmetry of the PCR method of Section 5.4 is not needed here. In our application we need to predict a selling price for period $(j,1)$ of a house i sold in period (j,m), using the WOZ valuation of this house (in period $(j,1)$).

So in stratum g of period $(j,1)$, we assume a linear regression model for the selling prices $V^g_{j,1,i}$ for house i sold in that period, and the corresponding WOZ valuation for that house, $W^g_{j,1,i}$.

$$V^g_{j,1,i} = \alpha^g_{j,1} W^g_{j,1,i} + \beta^g_{j,1} + \epsilon^g_{j,1,i}, \tag{24}$$

where $\alpha^g_{j,1}$ and $\beta^g_{j,1}$ are coefficients that have to be estimated and $\epsilon^g_{j,1,i}$ is the error term. The estimates are determined by applying the ordinary least squares (OLS) technique, that is by minimizing $\sum_{i \in HVW^g_{j,1,i}} (\epsilon^g_{j,1,i})^2$, which is an unweighted sum.

Suppose that this method yields estimators $\hat{\alpha}^g_{j,1}$ and $\hat{\beta}^g_{j,1}$. Then we can use

$$\tilde{V}^g_{j,1,i} = \hat{\alpha}^g_{j,1} W^g_{j,m,i} + \hat{\beta}^g_{j,1}, \tag{25}$$

to predict the value of $V^g_{j,1,i}$ as a function of $W^g_{j,m,i}$ for a house sold in month m.

We conclude the present section with two remarks.

Remark When applying the linear regression model (24), it is tacitly assumed that the independent variables (the WOZ valuations) are without errors. But they are likely to be with errors, as they have to be estimated with a limited set of parameters that determine the value of a property. So in fact, an errors-in-variables model would be more appropriate. Or principal components regression (PCR) could be used. See Section 5.4. □

Remark For the consumer price index (CPI) goods are typically weighted with turnover values, if available. In case of house prices this approach would imply that houses should be weighted proportional to their selling prices, or in fact selling price × quantity, where quantity $= 1$. □

7 Price indices using hypothetical selling prices

7.1 Ground material

In line with the reasoning in subsection 4.4 we propose to consider price indices that use hypothetical selling prices for January of the year they were actually sold as comparison prices. Using this idea we can find various methods to compute price indices, including the SPAR method.

The ground material for the class of indices considered here are the ratios of house price in period (j, m) of house i in stratum g compared to the estimated hypothetical house price in period $(j, 1)$:

$$\frac{V_{j,m,i}^{g}}{\hat{V}_{j,1,i}^{g}} \tag{26}$$

As we have seen in subsection 6 we can estimate $\hat{V}_{j,1,i}^{g}$ in various ways. This leads to various price indices, some of which we consider in the remainder of the present section.

7.2 Carli-like price index

A Carli-like price index is obtained by taking the arithmetic mean of the ratios (26) for all houses i in stratum g in period (j, m):

$$z_{j,m}^{C,g} \triangleq \frac{1}{NVW_{j,m}^{g}} \sum_{i \in HVW_{j,m}^{g}} \frac{V_{j,m,i}^{g}}{\hat{V}_{j,1,i}^{g}}, \tag{27}$$

for $m = 2, \ldots, 12$. For $m = 1$ we assume that $\hat{V}_{j,1,i}^{g} = V_{j,1,i}^{g}$.

It is known that a Carli index does not obey the time reversal test (and hence it also fails the transitivity test), and therefore it is not very attractive. But the situation in our case is slightly different from the one usually encountered. See also the discussion in Section 9.

7.3 Dutot-like price index

Instead of (27) we can use a Dutot type price index, which is a ratio of averages. We then obtain:

$$z_{j,m}^{D,g} \triangleq \frac{\sum_{i \in HVW_{j,m}^{g}} V_{j,m,i}^{g}}{\sum_{i \in HVW_{j,m}^{g}} \hat{V}_{j,1,i}^{g}}, \tag{28}$$

where we have omitted the factor $1/NVW_{j,m}^{g}$ in numerator and denominator.

7.4 Jevons-like price index

An alternative to (28) is to define a Jevons-like price index, using geometric averages rather than arithmetic ones. We then have

$$z_{j,m}^{J,g} \triangleq \prod_{i \in HVW_{j,m}^{g}} \left(\frac{V_{j,m,i}^{g}}{\hat{V}_{j,1,i}^{g}} \right)^{1/NVW_{j,m}^{g}}, \tag{29}$$

Formula (29) should only be used when the denominators $\hat{V}_{j,1,i}^{g}$ in this expression are positive, which is actually the case.

7.5 SPAR-like price index

We now use (23) as an estimate for $\hat{V}_{j,1,i}^{g}$. Then we can write (26) as

$$\frac{V_{j,m,i}^{g}}{\hat{V}_{j,1,i}^{g}} = \frac{V_{j,m,i}^{g}}{\hat{\phi}_{j,1}^{g} W_{j,m,i}^{g}}. \tag{30}$$

We can take the arithmetical average of expressions in (30). We now recognize the arithmetical average of $V_{j,m,i}^{g}/W_{j,m,i}^{g}$ as $\hat{\phi}_{j,m}^{g}$. Then the following index for stratum g is obtained:

$$z_{j,m}^{S1,g} \triangleq \frac{\hat{\phi}_{j,m}^{g}}{\hat{\phi}_{j,1}^{g}}, \tag{31}$$

where $\hat{\phi}_{j,m}^{g}$ is as defined in (17). Using this definition we can write (31) in expanded form as

$$z_{j,m}^{S1,g} = \frac{NVW_{j,1}^{g}}{NVW_{j,m}^{g}} \frac{\sum_{i \in HVW_{j,m}^{g}} V_{j,m,i}^{g}/W_{j,m,i}^{g}}{\sum_{i \in HVW_{j,1}^{g}} V_{j,1,i}^{g}/W_{j,1,i}^{g}}, \tag{32}$$

for period (j, m), period $(j, 1)$ and stratum g.

7.6 Another SPAR-like price index

In (32) we can replace the averages of the ratios in numerator and denominator by corresponding ratios of averages

$$z_{j,m}^{S2,g} = \frac{\sum_{i \in HVW_{j,m}^g} V_{j,m,i}^g / \sum_{i \in HVW_{j,m}^g} W_{j,m,i}^g}{\sum_{i \in HVW_{j,1}^g} V_{j,1,i}^g / \sum_{i \in HVW_{j,1}^g} W_{j,1,i}^g} \tag{33}$$

Index (33) is like the SPAR index, except that in this case only complete records are used, that is $HVW_{j,m}^g$. We can write (33) in the form (5) and (6). By replacing the summation over different sets we can obtain the SPAR index.

Through this derivation of the SPAR index one is able to understand how it fits in the approach taken in the present paper. However, it is also clear that it is not the price index one would immediately choose. The index in (31), or in expanded form in (32), would be a more likely choice. (33) could be seen as an approximation to (31), and the SPAR index itself as an approximation to (33).

Equation (33) can be expressed in terms of the χ factors defined in (21) as

$$z_{j,m}^{S2,g} \triangleq \frac{\hat{\chi}_{j,m}^g}{\hat{\chi}_{j,1}^g}. \tag{34}$$

Using the variant of the χ factors defined in (20) in this expression yields the SPAR index.

7.7 Robust price indices

The price indices presented in subsections 7.2 to 7.6 are all based on (arithmetical or geometric) averages. For all these indices robust versions can be defined, by replacing arithmetical or geometric averages by robust alternatives like the median. For all these indices we present 'robustified' versions of indices based on the median in separate subsections.

Robust indices could be used instead of outlier removal, achieved by replacing outliers by missing values, as is the current practice for the SPAR index.

7.7.1 Robust Carli-like price index

A robust version of (27) based on the median is:

$$z_{j,m}^{C,med,g} \triangleq \text{median}\{\frac{V_{j,m,i}^g}{\hat{V}_{j,1,i}^g} \mid i \in HVW_{j,m}^g\}, \tag{35}$$

7.7.2 Robust Dutot-like price index

A robust version of the index (28) is obtained by viewing the denominator and enumerator as arithmetic averages and replace these by corresponding medians. This yields the following index:

$$z_{j,m}^{D,med,g} \triangleq \frac{\text{median}\{V_{j,m,i}^{g} \mid i \in HVW_{j,m}^{g}\}}{\text{median}\{\hat{V}_{j,1,i}^{g} \mid i \in HVW_{j,m}^{g}\}}, \tag{36}$$

7.7.3 Robust Jevons-like price index

To produce a robust version of the Jevons index we first take (natural) logarithms and replace the arithmetical average we then obtain by medians. This yields a the logarithm of a robustified Jevons index:

$$\log z_{j,m}^{J,med,g} \triangleq \text{median}\{\log V_{j,m,i}^{g} - \log \hat{V}_{j,1,i}^{g} \mid i \in HVW_{j,m}^{g}\}. \tag{37}$$

7.7.4 Robust SPAR-like price index

The index (32) can be viewed as a ratio of two arithmetical means. Replacing these by medians, yields the following robust SPAR-like index:

$$z_{j,m}^{S1,med,g} \triangleq \frac{\text{median}\{V_{j,m,i}^{g}/W_{j,m,i}^{g} \mid i \in HVW_{j,m}^{g}\}}{\text{median}\{V_{j,1,i}^{g}/W_{j,1,i}^{g} \mid i \in HVW_{j,1}^{g}\}}. \tag{38}$$

7.7.5 Another robust SPAR-like price index

The index (33) can also be viewed as a ratio of two arithmetical averages, and replacing these by medians yields the following robust version of this index:

$$\begin{aligned} z_{j,m}^{S2,med,g} \triangleq & \frac{\text{median}\{V_{j,m,i}^{g} \mid i \in HVW_{j,m}^{g}\}}{\text{median}\{V_{j,1,i}^{g} \mid i \in HVW_{j,1}^{g}\}} \\ & \times \frac{\text{median}\{W_{j,1,i}^{g} \mid i \in HVW_{j,1}^{g}\}}{\text{median}\{W_{j,m,i}^{g} \mid i \in HVW_{j,m}^{g}\}}. \end{aligned} \tag{39}$$

7.8 A price index using PCR

In this case we take as a price index the ratio of the direction coefficient[25)] of the PCR regression line in period (j, m) and the corresponding coefficient in the same stratum and period $(j, 1)$:

$$z_{j,m}^{PCR,g} \triangleq \frac{\rho_{j,m}^{g}}{\rho_{j,1}^{g}}, \tag{40}$$

[25)] This is equal to the tangent of the angle of the regression line and the x-axis.

where $(1, \rho^g_{j,m})'$ is the eigenvector corresponding to the largest eigenvalue of of the matrix $Y'Y$ in Subsection 5.4 in stratum g, period (j, m). $(1, \rho^g_{j,1})'$ is similar but for period $(j, 1)$ and the same.

7.9 Using completed data

In the price indices treated so far only the nonmissing values in the data are used. But when imputation is first applied to the data, the data set is completed first and we can use this completed data set to estimate price indices. Essentially the same formulas as above can be applied only the set of records involved is different. In the formulas we treat the imputed values as if they are the observed values. But one should be aware that they are actually random variables for which specific values have been substituted, resulting from applying an imputation model. If this imputation model is estimated on the observed data - as is usually the case in practice[26] -, the effect of imputation has to be taken into account during variance estimation. In particular, when bootstrapping for variance estimation (see Willenborg and Scholtus (2018)), one should not treat the imputed values as if they were observed values but repeat the imputation step for each bootstrap sample separately.

To illustrate, we take the following price index above and show how they are modified when applied to a completed data set.

$$\hat{z}^{S1,g}_{j,m} = \frac{|\tilde{H}^g_{j,1}|}{|\tilde{H}^g_{j,m}|} \frac{\sum_{i \in \tilde{H}^g_{j,m}} V^g_{j,m,i}/W^g_{j,m,i}}{\sum_{i \in \tilde{H}^g_{j,1}} V^g_{j,1,i}/W^g_{j,1,i}}, \tag{41}$$

which is the same index formula as (32) except that it is applied to a completed data set, denoted as $\tilde{H}^g_{j,m}$. Of course, there are several ways in which the completion can be achieved.

7.10 Aggregating

So far we have only considered a single stratum and computed price indices for this stratum. But the price indices computed per stratum have to be aggregated to obtain a price index for the entire population. The easiest way is to give the strata equal weights and aggregate the strata using these weights to obtain unweighted price indices.

But it is possible to use weights based on the ’turnover’. For each house sold we know the price and quantity, namely 1, in each case. So for each stratum g and period (j, m) we can compute the total amount spent on house sales, namely the sum of house price (times 1, the quantity sold).

We define weights for stratum g and period (j, m) as follows:

[26] One somewhat common exception might be cold deck imputation, where the imputed values are based on donor records from a different data set.

$$w_{j,m}^{g} \triangleq \frac{\sum_{i \in HV_{j,m,i}^{g}} V_{j,m,i}^{g}}{\sum_{g} \sum_{i \in HV_{j,m,i}^{g}} V_{j,m,i}^{g}}. \tag{42}$$

These weights are nonnegative and add to 1. We can define the following aggregated index $z_{j,m}$ for period (j, m):

$$z_{j,m} \triangleq \frac{\prod_{g} \left(z_{j,m}^{g}\right)^{w_{j,m}^{g}}}{\prod_{g} \left(z_{j,1}^{g}\right)^{w_{j,1}^{g}}}, \tag{43}$$

where the products are taken over all the strata. $z_{j,m}^{g}$ and $z_{j,1}^{g}$ denote the indices at stratum g at period (j, m) and $(j, 1)$, respectively .

This index is transitive. In fact, any transitive index can be written in this form: the ratio of an average price for the reporting month and an average price for the reference month; see Willenborg (2018, Section 2).

8 Chaining

8.1 Introductory remarks

So far we have concentrated on various methods to compute short, yearly series of price indices. There is a yearly cycle because the WOZ evaluations are renewed every year.[27] As the methods above have shown, the WOZ valuation is an important auxiliary variable to the selling price. So in a new year, the new valuations should be used as they can be expected to be more closely related to the selling prices in the new year. But this means that we start another short series of indices starting with the value 1 in January. But we are basically interested in a long, multiannual series. This can be obtained from the short series by ’stitching them together’. This is done by using an overlapping month. (See section 8.2 for two seemingly differently approaches, which turn out to be the same.) This ’stitching together’ is known as chaining.

8.2 Overlapping months

Short series are first computed. But they need to be combined (chained) to form a long series. This chaining is done by considering an overlapping month, for which an index is computed using two consecutive WOZ valuations. Chaining is discussed in subsection 8.3. In the present subsection we discuss some preparatory work for this action.

[27] That is the ’renewal rate’ for the past several years. It used to be updated less frequently, as was remarked before.

We consider two ways to chain short series into a long one. At first sight they may seem different, but they in fact turn out to be the same. These approaches are called the 13-th month and the -1-st month, referring to the way this overlapping month is created: by extending a year to include the first month of the next year (which is viewed as the 13-th month of the current year j). Or it can be done or by starting a short series at period $(j, 12)$, which is the last month of the previous year j, viewed as the -1-st month of the next year $j + 1$.[28] Both methods are discussed, in separate sections. In fact the second one is the method that is adopted by the department for house price statistics at CBS. But we use the 13-th month approach in the present paper, as it feels at bit more intuitive (s). And, as said, both methods yield the same results, i.e. matching factors.

8.2.1 The 13-th month approach

The computations assume a stratification of the houses, in house type and municipality (and possibly other variables). A typical stratum is referred to by an indicator g, usually as a superscript. The SPAR method first produces a short series of 13 months. Each short series starts with the value 1. For each month m the average ratio of sale price and WOZ valuation of month m is compared to that of month 1 of that year as in formula (1), where $m = 1, \dots, 13$, with the understanding that month 13 in year j is the same as month 1 in year $j + 1$.

For month 13 of year j, i.e. month 1 of year $j + 1$, but compared with the WOZ valuation of year j, we have

$$s^g_{j,13} \triangleq \frac{\overline{V^g_{j+1,1}}/\overline{W^g_{j+1,1}}}{\overline{V^g_{j,1}}/\overline{W^g_{j,1}}}. \tag{44}$$

But once the WOZ valuation is available for year $j + 1$ a new short series of 13 months can start:

$$s^g_{j+1,m} = \frac{\overline{V^g_{j+1,m}}/\overline{W^g_{j+1,1}}}{\overline{V^g_{j+1,1}}/\overline{W^g_{j+1,1}}}, \tag{45}$$

for $m = 1, \dots, 13$.

As already indicated, this short series starts with the value of 1 for the first month of each year, i.e. $s^g_{j+1,1} = 1$. This holds for all short series in the 13-th month approach. We now have as the matching factors for this approach:

$$\frac{s^g_{j,13}}{s^g_{j+1,1}} = s^g_{j,13} = \frac{\overline{V^g_{j+1,1}}}{\overline{V^g_{j,1}}} \frac{\overline{W^g_{j,1}}}{\overline{W^g_{j+1,1}}}. \tag{46}$$

So these matching factors only relate to the first months of consecutive years j and $j + 1$.

[28] It should in fact be called the 0-th month, but this terminology seems to be confusing rather than illuminating.

8.2.2 The -1st month approach

Another option[29] for the choice of an overlapping month for year $j + 1$ is the last month of the previous year: period $(j, 12)$. So the reference month for December of year j is January of year $j + 1$. Instead of looking backwards, with this choice one looks forward. The reason for this choice is that the selling prices in December are closer to those of the next year. With the 13-th month approach in Subsection 8.2.1 there is more than a month between the comparison month and the reference month.

The SPAR index for the -1-st month of year $j + 1$ is:

$$s^g_{j+1,-1} \triangleq \frac{\overline{V^g_{j,12}}/\overline{W^g_{j,12}}}{\overline{V^g_{j+1,1}}/\overline{W^g_{j+1,1}}}. \tag{47}$$

As part of the short series for year j this index for month 12 is:

$$s^g_{j,12} = \frac{\overline{V^g_{j,12}}/\overline{W^g_{j,12}}}{\overline{V^g_{j,1}}/\overline{W^g_{j,1}}}. \tag{48}$$

The ratio of the values (47) and (48) is used in the -1-st month approach as the matching factor:

$$\frac{s^g_{j,12}}{s^g_{j+1,-1}} = \frac{\overline{V^g_{j+1,1}}/\overline{W^g_{j+1,1}}}{\overline{V^g_{j,1}}/\overline{W^g_{j,1}}} = \frac{\overline{V^g_{j+1,1}}}{\overline{V^g_{j,1}}}\,\frac{\overline{W^g_{j,1}}}{\overline{W^g_{j+1,1}}}, \tag{49}$$

which is the same as the matching factor (46).

8.3 A closer look at the long SPAR index series

We now consider the chaining for the SPAR index more closely. For other indices the chaining process is essentially the same, only differing in the matching factors used, that depend on the overlapping month (see subsection 8.2). This is explained below.

The yearly short series for the SPAR index are only semi-finished products. The interest is in the long series, running over several consecutive years. This long series is obtained by chaining various short series in a special way. This chaining is done by using the 13^{th} month of each year as the overlapping month and using the matching factor as a multiplication factor. The newly chained short series is made to fit the last value of the current long series.

[29] This is what is actually chosen for the SPAR index. But the results are actually the same for both choices, as the present subsection shows.

We start the long series by taking the short series for year 1. We use the matching factor $s_{1,13}$ as a multiplication factor for the series of the second year. So we start with the short series for the first two years:

$$1, s^g_{1,2}, \dots, s^g_{1,13} \tag{50}$$

for year 1 and

$$1, s^g_{2,2}, \dots, s^g_{2,13} \tag{51}$$

for year 2. $s^g_{j,m}$ for $m = 2, \dots, 12$ is as defined in (1) for the special case that $j = 2$, and $s^g_{2,13}$ as in (44) for the special case $j = 2$.

We combine them in such a way that at the overlapping month $s^g_{1,13} = s^g_{2,1}$ both series agree. This means that we multiply the short series for the second year with the matching factor $s^g_{1,13}$. Then the resulting long series for the first two years is:

$$1, s^g_{1,2}, \dots, s^g_{1,13}, s^g_{1,13} \cdot s^g_{2,2}, \dots, s^g_{1,13} \cdot s^g_{2,13} \tag{52}$$

To the long series for the first two years, the short series for the third year is added, viz:

$$1, s^g_{3,2}, \dots, s^g_{3,13} \tag{53}$$

Again a matching factor is used, so that at the overlapping month $j = 2, m = 13$ or $j = 3, m = 1$ the values of the long series for the first two years and the short series for the third year agree. This means that we have to multiply the short series for the third year by a factor $s^g_{1,13} \cdot s^g_{2,13}$. The result is the following long series for the first three years:

$$\begin{aligned} 1, s^g_{1,2}, \dots, s^g_{1,13}, \quad & (54) \\ s^g_{1,13} \cdot s^g_{2,2}, \dots, s^g_{1,13} \cdot s^g_{2,13}, \quad & (55) \\ s^g_{1,13} \cdot s^g_{2,13} \cdot s^g_{3,2}, \dots, s^g_{1,13} \cdot s^g_{2,13} \cdot s^g_{3,13}. \quad & (56) \end{aligned}$$

By now, the pattern of this chaining process should be clear. Obviously the matching factors $s^g_{j,13}$ play a key role. They accumulate when short series further removed from the first year are chained to the long series at the time. The factors used to fit a new short series to the current long one are, respectively

$$\kappa_2 \triangleq s^g_{1,13}, \tag{57}$$

$$\kappa_3 \triangleq s^g_{1,13} \cdot s^g_{2,13}, \tag{58}$$

$$\kappa_4 \triangleq s^g_{1,13} \cdot s^g_{2,13} \cdot s^g_{3,13}, \tag{59}$$

etc. The matching factors play a critical role in producing the long series from short ones. If one of these matching factors $s^g_{j,13}$ is not estimated well, its effect will be noticeable in the long series from that month onwards.

In case an index is used different from the SPAR index we have different matching factors, as the factors for the $13-th$ month are different. For instance in case of index $z^{S1,g}_{j,m}$ in (32) we have as the matching factor for year $j+1$

$$z^{S1,g}_{j,13} \triangleq \frac{NVW^g_{j,1}}{NVW^g_{j+1,1}} \frac{\sum_{i \in HVW^g_{j+1,1}} V^g_{j+1,1,i}/W^g_{j+1,1,i}}{\sum_{i \in HVW^g_{j,1}} V^g_{j,1,i}/W^g_{j,1,i}}. \tag{60}$$

The κ's in (57)-(59) are then modified accordingly. For other indices similar kinds of modifications apply.

8.4 About the matching factors

The method of using a single overlapping month is a general one to chain short index series to a long one. But one should realize that this makes this factor of crucial importance for the behavior of the long index series. The question arises how well this factor can be estimated. It is clear that there is more than a year between the reference month of the WOZ valuation of period $(j, 1)$ with which a house sold in period $(j+1, 1)$. In this period both market effects and innovations to (some of) the houses may have resulted in a divergence between selling prices of houses and corresponding WOZ valuations. The new WOZ valuation of period $(j+1, 1)$ may be closer to the hypothetical selling price of such a house in period $(j, 12)$. So an option would be to use these more recent WOZ valuations as the basis for the matching factors. Or to use these valuations in combination with the ones we used, based on the 13-th month, to compute a new matching factor, for instance by averaging them (taking the geometric average). This would affect the way the short series are chained to form a long one. We shall not study this topic in the present paper. However, it is an interesting problem for future research. In Willenborg and Scholtus (2018) some attention is paid to the behavior of the matching factors in a bootstrapping procedure.

9 Transitive closure

9.1 Preliminary remarks

Transitivity is a property that is preferable in an index. It means that the price development between two months i and j is independent of the match used to connect these months: if

$(i_1, \dots, i_l)$ is path connecting $i_1 = i$ and $i_l = j$, and $P^{(i_1,\dots,i_l)} \triangleq \Pi_{j=1}^{l-1} P^{i_j,i_{j+1}}$. If $(\iota_1, \dots, \iota_m)$ is another path connecting $\iota_1 = i$ and $\iota_m = j$ then $P^{(\iota_1,\dots,\iota_m)} \triangleq \Pi_{j=1}^{m-1} P^{\iota_j,\iota_{j+1}}$. In case the index is transitive we would have that $P^{(i_1,\dots,i_l)} = P^{(\iota_1,\dots,\iota_m)}$, so the index is independent of the path connecting i and j. In that case we can write $P^{i,j}$ to indicate that the index only depends on the base month i and reporting month j. Transitivity implies: $P^{i,i} = 1$ and $P^{i,j} \cdot P^{j,i} = 1$. Transitivity is a very desirable property of an index.[30] If an index is not transitive it can be made transitive (see Willenborg (2018)).

The situation with the SPAR (and other house price indices) is different from that for the CPI, due, roughly speaking, to the fact that houses are unique and important auxiliary information on selling prices is (as a rule) available (WOZ valuations) which also is regularly updated (in January). This implies that for the short series January acts as a fixed base month. By applying transitive closure one is able to deduce price index values for other pairs of months. This is discussed below. Furthermore, transitive closure for the long series is discussed as well.

9.2 Short index series

The indices for the short series are all with a fixed base within a year. It is easy to extend them to indices for any pair of months within the same year by applying transitive closure (see Willenborg (2018)). In this way we can define an index for any pair of months within the same year. This index is transitive, by definition. The method applies very generally.

Let $P^{i,j}$ be an index with base month i and reporting month j. In case of the indices considered in the present paper in each period $(j, 1)$ is the base month. For $i = 1, \dots, 13$ we have the price indices $P^{1,i}$. We can define $P^{i,j}$ for $i = 1, \dots, 13$ as follows

$$P^{i,j} \triangleq \frac{P^{1,j}}{P^{1,i}}. \tag{61}$$

It is easy to see that this defines a transitive price index: Let i, j, k be a triple of months within one year j then

$$P^{i,j} P^{j,k} = \frac{P^{1,j}}{P^{1,i}} \frac{P^{1,k}}{P^{1,j}} = \frac{P^{1,k}}{P^{1,i}} \triangleq P^{i,k} \tag{62}$$

holds. Special cases are when we take $i = j = k$ which yields that $P^{i,i} = 1$ for $i = 1, \dots, 13$ and $k = i$ which yields $P^{i,j} P^{j,i} = P^{i,i} = 1$, for $i, j = 1, \dots, 13$. This means that the price index defined by (61) satisfies the time reversal test.

The method works because the price index digraph (PIDG) with the set of points $\{1, \dots, n\}$ and the set of arcs being equal to $\{(1, 2), (1, 3), \dots, (1, n)\}$ form a spanning tree for the complete PIDG with the same number of nodes (corresponding to months). The indices corresponding to other arcs than arcs other than those that are part of the spanning tree can be computed from

[30] To put it mildly.

the ones part of the spanning tree. The process of completing the PIDG from the information in the spanning tree is called transitive closure.

It should be noted that this process also is possible if the price index used to compute the price indices for a spanning tree is not transitive, or intransitive. In case the PIDG is a tree (and therefore has no cycles, that is, closed paths) the index is transitive. In case the PIDG is a cyclic digraph the index may be intransitive, but from it an index can be computed that is transitive, for instance by using the GEKS method or the cycle method (see Willenborg (2018)).[31)]

9.3 Long index series

For every year we have an index with January as a fixed base month, as we have seen in the previous section. For this it is easy to determine a transitive closure within each year.

We obtain a long series by multiplying each short series with the right factor. This implies that if we compare two months s, t within the same year j we obtain the same results when comparing them for the short series for j. We have

$$P_L^{s,t} \triangleq \frac{P_L^{1,t}}{P_L^{1,s}} = \frac{\kappa_j P_S^{1,t}}{\kappa_j P_S^{1,s}} = \frac{P_S^{1,t}}{P_S^{1,s}}, \tag{63}$$

where the subscripts L, S refer to the long and short series respectively and κ_j is the matching factor to chain the short series for year j to the long series up until year $j-1$. In (57)-(59) one can find the matching factors $\kappa_2, \kappa_3, \kappa_4, \ldots$ for the SPAR index.

In case of two months s, t in different years, say j, k respectively, we have

$$P_L^{s,t} = \frac{P_L^{1,t}}{P_L^{1,s}} = \frac{\kappa_j P_S^{1,t}}{\kappa_k P_S^{1,s}}, \tag{64}$$

10 Methods without chaining

The methods discussed so far all had in common that short series of indices were built first and by chaining a long series was produced. The periods of short series coincide with WOZ periods. But actually there is no necessity to use this approach. An alternative is to produce indices using hypothetical selling prices. The most recent WOZ valuations are used as auxiliary variables to estimate hypothetical selling prices. For the index the key items are the selling prices and the hypothetical selling prices. The updating of the WOZ valuations affects the estimation of the hypothetical selling prices, but not the computation of the index using the hypothetical selling

[31)] With the cycle method one can control how close the 'transitivized' index is from the original index. With the GEKS method this is not possible.

prices. Using such an approach long index series are computed straightaway. One can only expect ’problems’ at transitions from one WOZ period to the next one, as this may prove to be somewhat disruptive.

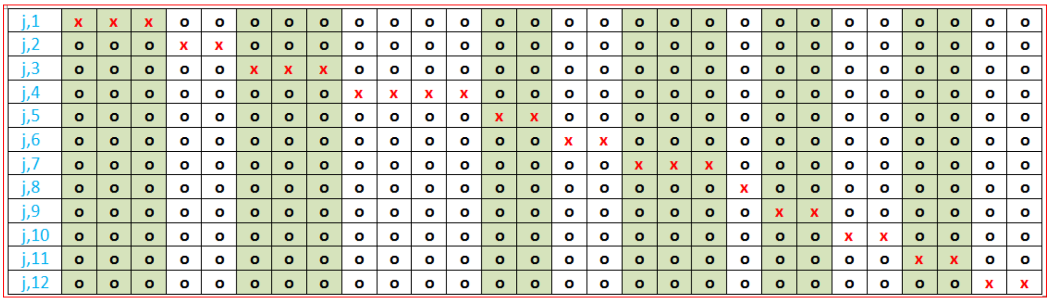
To indicate that there is continuity in the months when computing price indices, we use Greek letters to indicate the months, typically μ and ν. This extends to the sets that indicate houses sold in a particular period (year, month).

Each of the models presented in the remainder of the present section actually represents a class of models. The reason is that hypothetical selling prices can be estimated in a number of ways, as was shown in Section 6. In each case we pick a simple, specific method to make the discussion concrete. But it should be borne in mind that other methods could have been used instead, possibly supperior ones. The emphasis here is not on how to estimate hypothetical selling prices, but on their use to compute an index for house prices. For methods to estimate hypothetical selling prices Section 6 should be consulted.

10.1 Entire WOZ period

In the first model we take a WOZ period and consider all houses that have been sold in that period. For these houses we estimate the hypothetical selling prices in other months of that period, using a method described in Section 6. Each of these (and others) can be used to estimate these hypothetical selling prices. We take a simple model, based on a ratio estimator (cf. (17)).

Schematically the situation is depicted in Figure 10.1. In this figure every column corresponds to a house and every row to a month. A ’×’ indicates the selling price of a house in the month that it is actually sold and a ’o’ indicates a hypothetical selling price in another month, computed in one way or another. Selling prices and hypothetical selling prices in the same column correspond to the same house. The alternatingly coloured green and white columns indicate sets of houses sold in the same month. This notation is used in illustrations throughout the present section.



j,1	x	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
j,2	o	o	o	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
j,3	o	o	o	o	o	x	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
j,4	o	o	o	o	o	o	o	o	x	x	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
j,5	o	o	o	o	o	o	o	o	o	o	o	o	x	x	o	o	o	o	o	o	o	o	o	o	o	o	o	o
j,6	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x	o	o	o	o	o	o	o	o	o	o	o	o
j,7	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x	x	o	o	o	o	o	o	o	o	o
j,8	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	o	o	o	o	o	o	o	o
j,9	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x	o	o	o	o	o	o
j,10	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x	o	o	o	o
j,11	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x	o	o
j,12	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x

Figure 10.1 Selling prices (’×’) and hypothetical selling prices (’o’) for an entire WOZ period.

We consider two methods to compute price indices in this situation. The first method considers a chain of selling price and hypothetical selling prices for each house over the entire WOZ period. For such a chain for each pair of months price indices can be computed in the form of price ratios. In the next step, an index for each pair of months is computed by averaging the indices of the individual houses for this period. The second method replaces the real selling price of a house by an estimate, using the same model as used for the hypothetical selling prices in the other months of the WOZ period. This simplifies the formula for the price index, while the numerical values will quite often be essentially the same.

We remark that it suffices to make the index comptutations for a subset of all possible pairs, say the consecutive months, or by using a fixed base month. That is because the index is transitive.

10.1.1 Using both real and hypothetical selling prices

We use selling prices and hypothetical selling prices for all houses in the complementary months in the WOZ period. Figure 10.1 schematically characterizes the situation. A price index is now obtained by comparing these totals for different pairs of months:

$$\pi^{\mu,\nu} = \frac{\sum_{i\in H_\nu^g} V_{\nu,i}^g + k_\nu \hat{\phi}_\nu^g \bar{W}_\nu^g}{\sum_{i\in H_\mu^g} V_{\mu,i}^g + k_\mu \hat{\phi}_\mu^g \bar{W}_\mu^g}, \tag{65}$$

where k_ν is the number of houses in the WOZ-period considered that are not sold in month ν and $\bar{W}_\nu^g$ is the average WOZ value of the houses in the WOZ-period, not sold in month ν. The index defined in (65) is transitive.

10.1.2 Using hypothetical selling prices only

In (65) we have a mix of real and hypothetical selling prices, where the latter would dominate the former ones in most practical cases. The expression in (65) simplifies if we work with hypothetical selling prices only. We only need to replace the real selling prices in (65) by the estimates of the corresponding hypothetical selling prices. We then obtain after some simplification:

$$\pi_H^{\mu,\nu} = \frac{\hat{\phi}_\nu^g}{\hat{\phi}_\mu^g}, \tag{66}$$

which is equal to (31). In practice (66) and (65) are likely to yield essentially the same results.

Houses in two subsequent WOZ-periods are shown in Figure 10.2. For the houses sold in one particular month (the ‘overlap month’) the old as well as the new WOZ values are known, apart from their respective selling prices. These houses are coloured dark blue. Therefore, for these houses hypothetical selling price in both WOZ periods can be estimated, using the appropriate WOZ valuations. These are indicated by the areas coloured with lighter shades of blue, one corresponding to the old WOZ-period and the other to the new one. For the houses in the dark blue area two hypothetical selling prices can be estimated in the overlap month, based on the two WOZ valuations For the houses sold in the non-overlapping months the selling prices of the remaining months in the corresponding WOZ-period can be estimated. These areas are coloured gray. They include the months in which the selling prices are actually known.

A drawback of the method discussed in the present section, based on the data of hypothetical selling prices of all houses in WOZ periods is that data of a lot of houses are typically involved. Another drawback is that the method can only be applied after two subsequent WOZ periods have been completed. That leads to a substantial - and unrealistic - delay in estimation and publication of the estimates.

We now consider the transition of WOZ periods more closely. The houses in the overlap month are part of two consecutive WOZ periods, and for each period there is a WOZ valuation available.

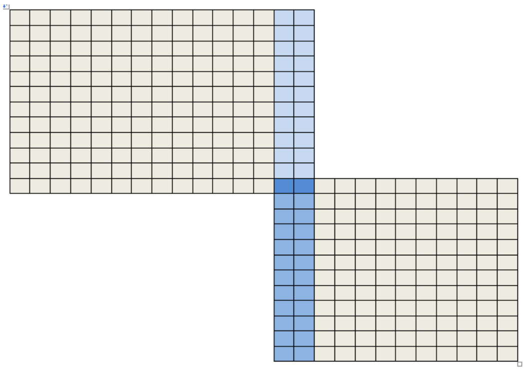


Figure 10.2 Overlapping WOZ periods.

For the older period one can use the appropriate WOZ valuation to compute the hypothetical selling prices for the previous months in that period. For the newer WOZ period the other WOZ valuation can be used. If we want to compute the average prices (real selling prices and hypothetical selling prices) for the overlap month there are contributions from both WOZ periods, namely from the other houses in both periods. If we average these prices, this is the same as taking a weighted average of the average per WOZ period, where the contribution of each period is weighted proportional to the number of houses involved, which seems a reasonable thing to do.

10.2 Adjacent months

In (65) and (31) the influence of the hypothetical prices is significant. We can try to reduce that influence by using a more localized model. Instead of the houses sold in an entire WOZ-period, we only consider houses sold in a given month, the month immediately preceeding and the month immediately succeeding, so the adjacent months for a reference month. This is depicted in Figure 10.3.

j,1	x	x	x	o	o																							
j,2	o	o	o	x	x	o	o	o																				
j,3				o	o	x	x	x	o	o	o	o																
j,4						o	o	o	x	x	x	x	o	o														
j,5									o	o	o	o	x	x	o	o												
j,6													o	o	x	x	o	o	o									
j,7															o	o	x	x	x	o								
j,8																	o	o	o	x	o	o						
j,9																				o	x	x	o	o				
j,10																					o	o	x	x	o	o		
j,11																							o	o	x	x	o	o
j,12																									o	o	x	x

Figure 10.3 Selling prices ('×') and hypothetical selling prices ('o') for adjacent months.

As in previous cases we can cope with this situation in two ways: first look vertically (per house) and compute ratios per house for available pairs of months, and then average these figures for all houses involved. Or, compute the total selling price (for real and hypothetical selling prices per month) and use thes total to compute total price ratios and use these as price indices. The resulting index is transitive.

The adjacent option requires one to delay the computation for a given month until the data for the next month are available. Such a delay is not very attractive in case one wants to publish current figures. The situation depectied in Figure 10.3 is similar but not the same as the one

j,1	x	x	x	o	o	o	o	o																				
j,2				x	x	o	o	o	o	o	o	o																
j,3						x	x	x	o	o	o	o	o	o														
j,4									x	x	x	x	o	o	o	o												
j,5													x	x	o	o	o	o										
j,6															x	x	o	o	o	o								
j,7																	x	x	x	o	o	o						
j,8																				x	o	o	o	o				
j,9																					x	x	o	o	o	o		
j,10																							x	x	o	o	o	o
j,11																									x	x	o	o
j,12																											x	x

Figure 10.4 Selling prices ('×') and hypothetical selling prices ('o') for the previous two months.

depicted in Figure 10.4, where for each month hypothetical selling prices are used for the previous two months.

10.3 Consecutive months

Another possibility is to even further restrict the number of months involved and look for each month only at the next or the previous month. We consider both cases in separate subsections. The results based on these cases are not likely to differ much. But the latter one is more practical, as it can be applied as soon as the data for the current month are available. In the former case one needs to wait one month, so there is a delay in the data provision.

In both cases we can compute price indices in two ways: either concentrate on the separate houses, compute price ratios and average the indices for month pairs. Or compute totals of selling prices (real and hypothetical) for each month and use these to compute price ratios to be used as indices. This is necessary only for a subset of all month pairs, as this price index is transitive.

10.3.1 Next month

The situation where hypothetical seling prices for the next month are used is schematically presented in Figure 10.5.

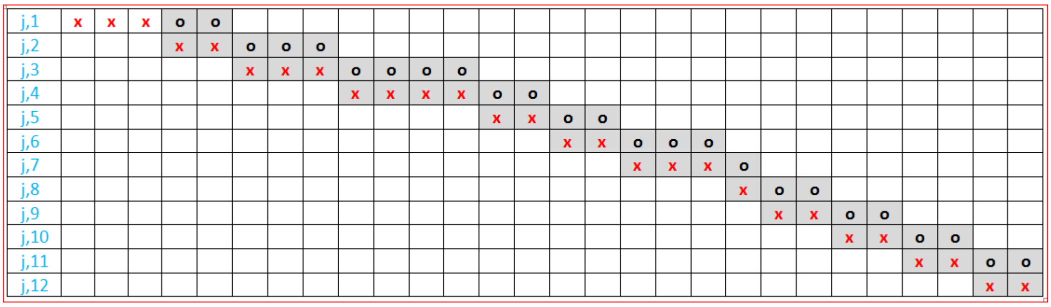
j,1	x	x	x																									
j,2	o	o	o	x	x																							
j,3				o	o	x	x	x																				
j,4						o	o	o	x	x	x	x																
j,5									o	o	o	o	x	x														
j,6													o	o	x	x												
j,7															o	o	x	x	x									
j,8																	o	o	o	x								
j,9																				o	x	x						
j,10																					o	o	x	x				
j,11																							o	o	x	x		
j,12																									o	o	x	x

Figure 10.5 Selling prices ('×') and hypothetical selling prices ('o') for the next month.

Figure 10.5 is an option that is of interest in theory, but not in practice, as it implies a publication delay of 1 month, each month.

10.3.2 Previous month

The case where the hypothetical selling prices of the previous month are used are shown in Figure 10.6.



j,1	x	x	x	o	o																							
j,2				x	x	o	o	o																				
j,3						x	x	x	o	o	o	o																
j,4									x	x	x	x	o	o														
j,5													x	x	o	o												
j,6															x	x	o	o	o									
j,7																	x	x	x	o								
j,8																				x	o	o						
j,9																					x	x	o	o				
j,10																							x	x	o	o		
j,11																									x	x	o	o
j,12																											x	x

Figure 10.6 Selling prices ('×') and hypothetical selling prices ('o') for the previous month.

The situation depicted in Figure 10.6 is more attractive in practice as there is no delay in publication. As soon as the data for a particular month have been processed one is able to update the index, without any delay.

10.4 Incremental approaches

We consider two variants of an incremental approach: a global and a local version. In the former version there is no limit to the months look back (within the WOZ period), and in the former there is.

In both cases we can compute price indices in two ways: either concentrate on the separate houses, compute price ratios and average the indices for the month pairs. Or compute totals of selling prices (real and hypothetical) for each month and use these to compute price ratios to be used as price indices. This is necessary only for a suitable subset of all month pairs (which should form a spanning tree for the vertices months in the WOZ period), as this price index is transitive.

10.4.1 Incremental approach - global version

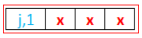
The average price of real and hypothetical selling prices is determined by using all the sold houses in the current WOZ period up to a given month (in a stratum). Information of houses to be sold in the future is not known, and therefore is not used (cannot be used). Houses sold in previous WOZ periods are not used at all. Houses sold in an overlap month use the appropriate WOZ value to estimate the hypothetical selling prices in other months of the corresponding WOZ period. Previously computed averages are not recomputed.

Also the houses in which the WOZ valuation is available but the selling price is missing can be used in the imputation/estimation process as well. But it is an option to use only houses with both these variables available.

Another variant of this model uses only hypothetical selling prices. So for the current months one uses the hypothetical selling prices on the basis of the corresponding WOZ valuations. In practice the results are likely to be very similar.

The method is illustrated below with figures to illustrate the situation for each month within a WOZ period (a year j) and the transition to the next WOZ period (year $j+1$).

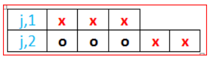
The average house price in month $(j, 1)$ is that of the houses sold in that period.



j,1	x	x	x

Figure 10.7 Incremental approach global version, month $(j, 1)$. Selling prices '×'.

For month $(j, 2)$ the average house price is that of the hypothetical selling prices of the houses sold in month $(j, 1)$ and the real selling prices of the houses sold in month $(j, 2)$.

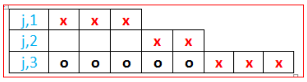


j,1	x	x	x		
j,2	o	o	o	x	x

Figure 10.8 Incremental approach global version, month $(j, 2)$. Selling prices '×' and hypothetical selling prices 'o'.

For month $(j, 3)$ the average house price is that of the hypothetical selling prices of the houses sold in month $(j, 1)$ and month $(j, 2)$ and the real selling prices of the houses sold in month $(j, 3)$.

For the remaining months of the WOZ period essentially the same pattern is followed.



j,1	x	x	x					
j,2				x	x			
j,3	o	o	o	o	o	x	x	x

Figure 10.9 Incremental approach global version, month $(j, 3)$. Selling prices '×' and hypothetical selling prices 'o'.

For the houses sold in month $(j + 1, 2)$ the WOZ valuations belonging to the new WOZ period are used to estimate the hypothetical selling prices in month $(j + 1, 1)$ and $(j + 1, 2)$.

For the houses sold in month $(j, 12)$ only the real selling prices are used. The WOZ valuations of these houses would only be used to compute hypothetical selling prices in previous months, but these are not needed in the present model.

For the houses sold in month $(j + 1, 1)$ the WOZ valuations belonging to the new WOZ period are used to estimate their hypothetical selling prices in month $(j + 1, 1)$.

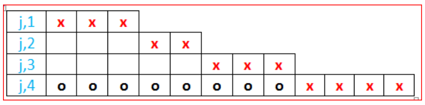
Fom hereon the pattern as shown before repeats itself (of course with not necessarily the same numbers of houses sold per month).

10.4.2 Incremental approach - local version

The average price of real and hypothetical selling prices is determined by using the sold houses in the current WOZ period which have been sold 1,2 or 3 months before a given month (in a stratum). Information of houses to be sold in the future is at that moment in time is not known, and is therefore not used (in fact, cannot be used). Houses sold in previous WOZ periods are not used at all. Houses sold in an overlap month use the appropriate WOZ value to estimate the hypothetical selling prices in other months of the corresponding WOZ period. Previously computed averages are not recomputed.

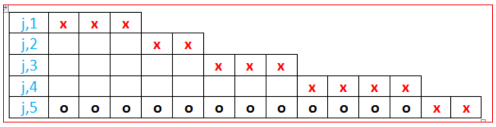
So this model is local in the sense that hypothetical selling prices only of houses sold in the recent past are used.

Also the houses were the WOZ valuation is available and the selling price is missing can be used in the imputation/estimation process as well. A variant of this model is one in which only hypothetical selling prices are used. In practice the results are likely be very similar.



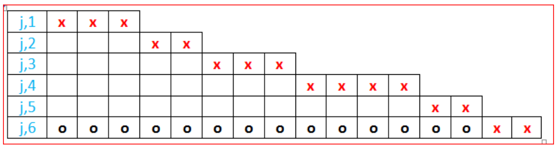
j,1	x	x	x									
j,2				x	x							
j,3						x	x	x				
j,4	o	o	o	o	o	o	o	o	x	x	x	x

Figure 10.10 Incremental approach global version, month $(j, 4)$. Selling prices '×' and hypothetical selling prices 'o'.



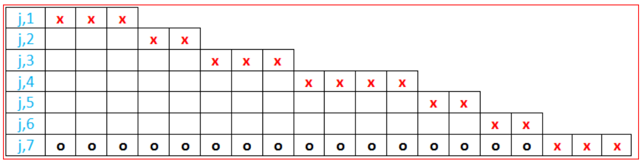
j,1	x	x	x											
j,2				x	x									
j,3						x	x	x						
j,4									x	x	x	x		
j,5	o	o	o	o	o	o	o	o	o	o	o	o	x	x

Figure 10.11 Incremental approach global version, month $(j, 5)$. Selling prices '×' and hypothetical selling prices 'o'.



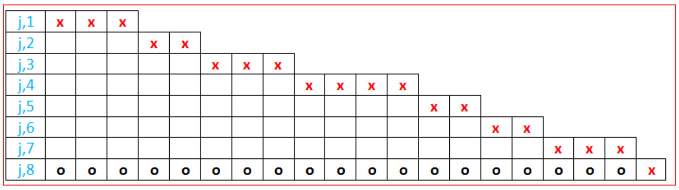
j,1	x	x	x													
j,2				x	x											
j,3						x	x	x								
j,4									x	x	x	x				
j,5													x	x		
j,6	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x

Figure 10.12 Incremental approach global version, month $(j, 6)$. Selling prices '×' and hypothetical selling prices 'o'.



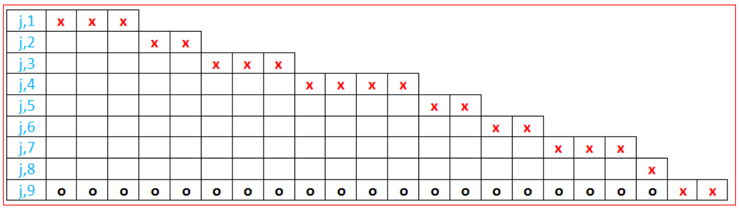
j,1	x	x	x															
j,2				x	x													
j,3						x	x	x										
j,4									x	x	x	x						
j,5													x	x				
j,6															x	x			
j,7	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x	x

Figure 10.13 Incremental approach global version, month $(j, 7)$. Selling prices '×' and hypothetical selling prices 'o'.



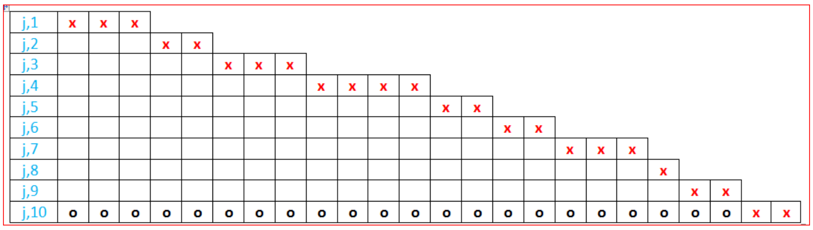
j,1	x	x	x																
j,2				x	x														
j,3						x	x	x											
j,4									x	x	x	x							
j,5													x	x					
j,6															x	x				
j,7																	x	x	x	
j,8	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x

Figure 10.14 Incremental approach global version, month $(j, 8)$. Selling prices '×' and hypothetical selling prices 'o'.



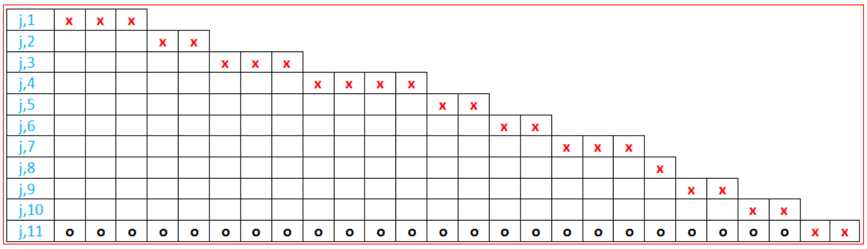
j,1	x	x	x																		
j,2				x	x																
j,3						x	x	x													
j,4									x	x	x	x									
j,5													x	x							
j,6															x	x					
j,7																	x	x	x			
j,8																				x		
j,9	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x

Figure 10.15 Incremental approach global version, month $(j, 9)$. Selling prices '×' and hypothetical selling prices 'o'.



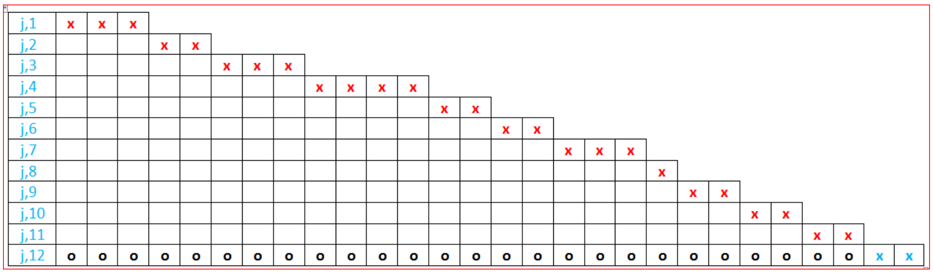
j,1	x	x	x																				
j,2				x	x																		
j,3						x	x	x															
j,4									x	x	x	x											
j,5													x	x									
j,6															x	x							
j,7																	x	x	x				
j,8																				x			
j,9																					x	x		
j,10	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x

Figure 10.16 Incremental approach global version, month $(j, 10)$. Selling prices '×' and hypothetical selling prices 'o'.



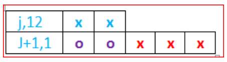
j,1	x	x	x																							
j,2				x	x																					
j,3						x	x	x																		
j,4									x	x	x	x														
j,5													x	x												
j,6															x	x										
j,7																	x	x	x							
j,8																				x						
j,9																					x	x				
j,10																							x	x		
j,11	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x

Figure 10.17 Incremental approach global version, month $(j, 11)$. Selling prices ’×’ and hypothetical selling prices ’o’.



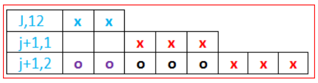
j,1	x	x	x																									
j,2				x	x																							
j,3						x	x	x																				
j,4									x	x	x	x																
j,5													x	x														
j,6															x	x												
j,7																	x	x	x									
j,8																				x								
j,9																					x	x						
j,10																							x	x				
j,11																									x	x		
j,12	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x

Figure 10.18 Incremental approach global version, month $(j, 12)$. Selling prices ’×’ and hypothetical selling prices ’o’.



j,12	x	x			
J+1,1	o	o	x	x	x

Figure 10.19 Incremental approach global version, month $(j + 1, 1)$. Selling prices ’×’ and hypothetical selling prices ’o’.



J,12	x	x						
j+1,1			x	x	x			
j+1,2	o	o	o	o	o	x	x	x

Figure 10.20 Incremental approach global version, month $(j + 1, 2)$. Selling prices ’×’ and hypothetical selling prices ’o’.

Now we illustrate this method with some figures depicting the situation.

The average house price in month $(j, 1)$ is that of the houses sold in that period.

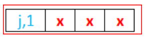


Figure 10.21 Incremental approach local version, month $(j, 1)$. Selling prices '×'.

For month $(j, 2)$ the average house price is that of the hypothetical selling prices of the houses sold in month $j, 1$ and the real selling prices of the houses sold in month $(j, 2)$.

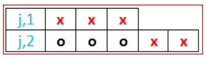


Figure 10.22 Incremental approach local version, month $(j, 2)$. Selling prices '×' and hypothetical selling prices 'o'.

For month $(j, 3)$ the average house price is that of the hypothetical selling prices of the houses sold in month$(j, 1)$ and month $(j, 2)$ and the real selling prices of the houses sold in month $(j, 3)$.

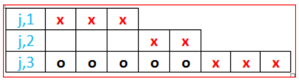


Figure 10.23 Incremental approach local version, month $(j, 3)$. Selling prices '×' and hypothetical selling prices 'o'.

For the remaining months of the WOZ period essentially the same pattern is followed, so that in each period roughly the same amount of data (houses) is used. Only the hypothetical selling prices for the houses sold in the 2 months immediately preceding the current month are used as well as the selling prices for the current month.

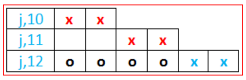
For the houses sold in month $(j, 12)$ only the real selling prices are used, as the WOZ valuations of these houses would only be used to compute hypothetical selling prices in previous months; but these are not needed in the present model.

For the houses in month $(j + 1, 1)$ the WOZ valuations belonging to the new WOZ period are used to estimate their hypothetical selling prices in month $(j + 1, 1)$.

For the houses in month $(j + 1, 2)$ the WOZ valuations belonging to the new WOZ period are used to estimate the hypothetical selling prices in month $(j + 1, 1)$ and $(j + 1, 2)$.

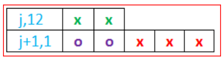
10.5 Pooled data

The final incremental model that we consider is one in which a finite window is considered, as in case of the local version of the incremental approach in Section 10.4.2. For the moment, we consider only complete records, that is, records that contain both WOZ valuation and selling price. We now pool the data from this reference period with the aim to estimate a single model relating selling prices and WOZ valuations. We use the estimated ϕ values for the reference period to estimate the WOZ valuation for the current month. In previous cases, the ϕ values were estimated using only data from a single month. In the pooled case we use data from a reference period, which consists of several, recent months. It is reasonable to assume that in the relatively short period of time of only a few months prices do not change drastically. The advantage is that pooling data of several months creates a bigger set of data and would produce more stable estimates of the ϕ's, and hence more stable estimates of the index based on it. On



j,10	x	x				
j,11			x	x		
j,12	o	o	o	o	x	x

Figure 10.24 Incremental approach local version, month $(j,12)$. Selling prices '$\times$' and hypothetical selling prices 'o'.



j,12	x	x			
j+1,1	o	o	x	x	x

Figure 10.25 Incremental approach local version, month $(j+1,1)$. Selling prices '$\times$' and hypothetical selling prices 'o'.

the other hand, the index should be able to pick up changes, as the 'pooling period' is only a few months. In our example we consider periods of three consecutive months.

The period used to pool the data for month $(j,m+1)$ has an overlap of 1 or 2 months with that of the previous period, resulting in a correlation between the estimates $\phi_{j,m}$ and $\phi_{j,m+1}$. By taking subsamples for each period one is able to control the correlation of the ϕ's. The smaller the overlap the less the correlation, but also the higher the variances for the ϕ's. These are variants of the method described below that have not been eloborated here.

We now can present the figures illustrating the method. For month $(j,1)$ only selling prices and WOZ valuations of the houses sold in that period are used.

For month $(j,2)$ only selling prices and WOZ valuations of the houses sold in month $(j,1)$ and in month $(j,2)$ are used.

For month $(j,3)$ selling prices and WOZ valuations of the houses sold in months $(j,1),(j,2),(j,3)$ are used to estimate $\phi_{j,3}$. This is the first time a full period consisting of three months is reached. For subsequent periods this is similar until month $(j,11)$ is reached.

Month $(j,12)$ is a special month, and therefore $\times$ symbols are coloured blue and not red: month $(j,12)$ is an overlap month with two WOZ valuations available, for year j and year $j+1$. For period $(j,12)$ the WOZ valuations for year j are used. Furthermore the data from months $(j,10),(j,11)$ are used to estimate $\phi_{j,12}$.

The period associated with month $(j+1,1)$ consists only of data from month $(j+1,1)$ and $(j,12)$, as no WOZ valuations for the new WOZ period $j+1$ are available. For the houses sold in month $(j,12)$ the WOZ values belonging to the new WOZ period, which we assume to be year $j+1$, are used to estimate $\phi_{j+1,1}$. To indicate this change we have given the $\times$ symbols for month $(j,12)$ a different colour (green).

For month $(j+1,2)$ the associated period consists of three months: $(j,12),(j+1,1),(j+1,2)$. For the houses in month $(j,12)$ the WOZ valuations belonging to the new WOZ period are used.

From hereon the pattern of the previous WOZ period, year j, repeats itself.

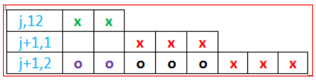


Figure 10.26 **Incremental approach local version, month $(j+1,2)$. Selling prices ’×’ and hypothetical selling prices ’o’.**

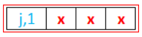


Figure 10.27 **Incremental approach using pooled data, month $(j,1)$. Selling prices ’×’ and hypothetical selling prices ’o’.**

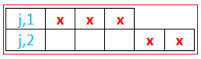


Figure 10.28 **Incremental approach using pooled data, month $(j,2)$. Selling prices ’×’ and hypothetical selling prices ’o’.**

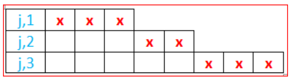


Figure 10.29 **Incremental approach using pooled data, month $(j,3)$. Selling prices ’×’ and hypothetical selling prices ’o’.**

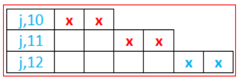


Figure 10.30 **Incremental approach using pooled data, month $(j,12)$. Selling prices ’×’ and hypothetical selling prices ’o’.**

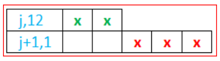


Figure 10.31 **Incremental approach using pooled data, month $(j+1,1)$. Selling prices ’×’ and hypothetical selling prices ’o’.**

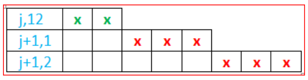


Figure 10.32 **Incremental approach using pooled data, month $(j+1,2)$. Selling prices ’×’ and hypothetical selling prices ’o’.**

11 Results

We have applied some of the estimators introduced above to some real housing data, in fact the same data that were used in our SPAR bootstrap study (cf. Willenborg and Scholtus (2018)). These data cover the period 1995-2017 for two regions (municipalities) in The Netherlands (344 = Utrecht and 518 = The Hague) and 2 types of houses (A = apartments and T = terraced houses).

In Table 11.1 an overview is given of the indices used in the figures below linked to the appropriate formulas or subsections.

Table 11.1 Indices used in the figures in the present section.

Index	Location in text
SPAR	Formula (1)
no WOZ	Formula (8)
HSP (lm)	Formulas (27) and (25)
phi ratio	Formula (31)
NCHSP full	Section 10.1 and (66)
NCHSP adjacent	Section 10.2 and (66)
NCHSP adj_next	Section 10.3.1 and (66)
NCHSP adj_prev	Section 10.3.2 and (66)
NCHSP incr	Section 10.4.1 and (66)
NCHSP incr_max_3	Section 10.4.2 and (66)
NCHSP pool_max_2	Section 10.5 and (66)

The indices mentioned in Table 11.1 are only a subset of the indices mentioned in the text. The first half of the indices mentioned in Table 11.1 are chained indices, whereas the latter half are nonchained ones (the prefix 'NC' indicates this). The SPAR index is the one currently used by the department of house prices at CBS, and is therefore taken as a reference index for the other indices. This does not necessarily imply that it is 'the best', or the most preferable index of all indices considered in the present paper, let alone those mentioned in Table 11.1. But as it is the index currently adopted by the Department of house price statistics it is not an arbitrary index but one with a certain status.

We start our presentation of some of the results with a picture of the development of the WOZ valuations, in the form of a long index with 1995 as base year. This long 'WOZ-index' was obtained by chaining short indices per WOZ period. In Subsection 4.4 the development of the WOZ valuation (within a WOZ period) was discussed.

Figure 11.1 shows short index series of WOZ valuations per stratum (Type of house × Region) and for each of the WOZ periods. Inspecting it immediately shows how the developments of the average monthly WOZ valuations per stratum differ. Also it shows that for entire WOZ periods, the index is smaller or bigger than 1. For region 518 in particular the development seems to be far from random. Note how strongly the short indices for T518 fluctuate compared to the other WOZ indices in Figure 11.1.

In Figure 11.2 the correlation between the WOZ valuations and the selling prices of houses is shown, for the two strata (apartments and terraced houses) and two regions over the entire period studied. As the tables show the correlation is usually quite high. There is only one sharp dip for the apartments in region 344.

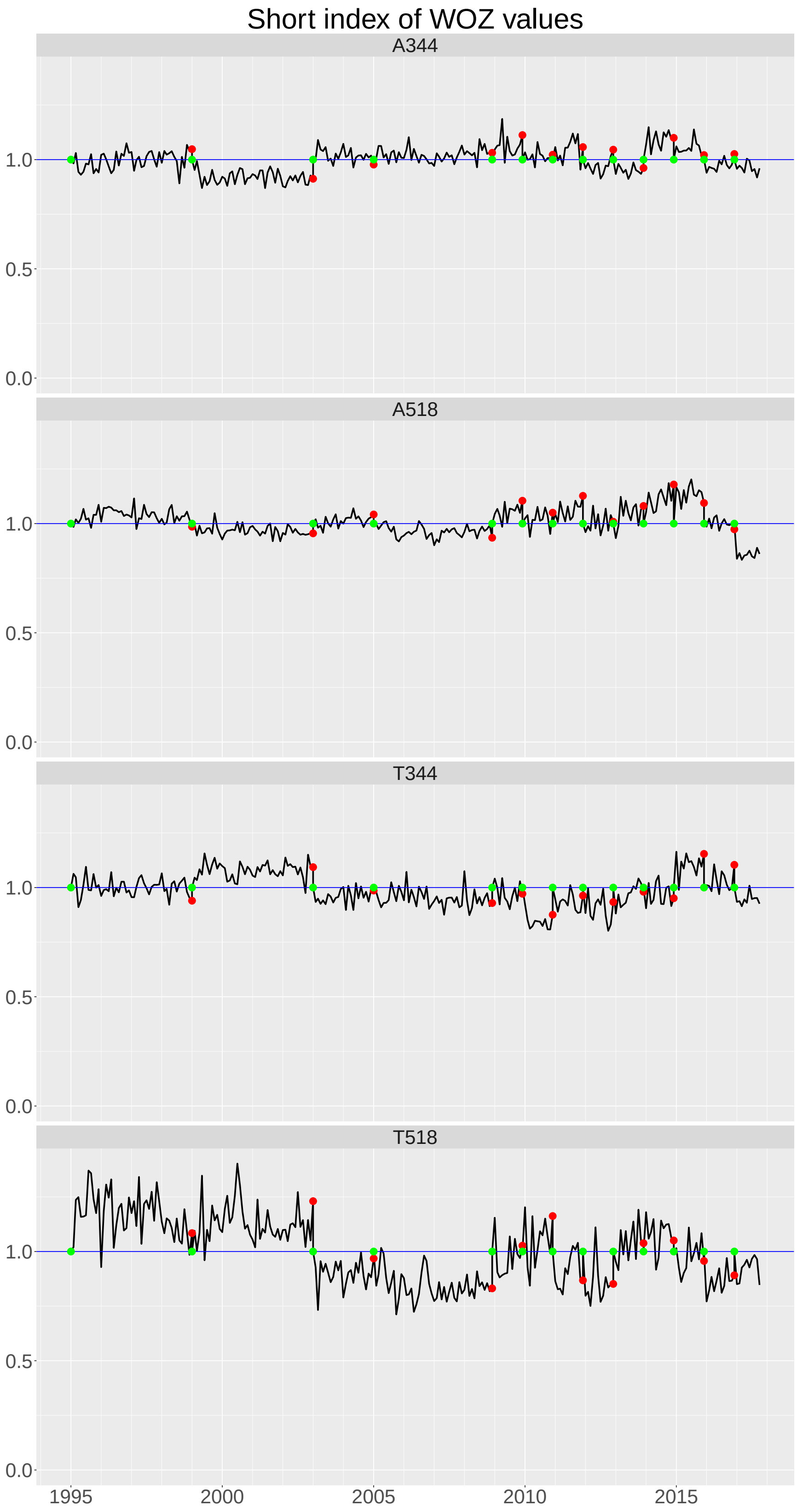


Figure 11.1 Short indices of average monthly WOZ valuations per stratum (type of house × region). Each short index correspondes to a WOZ period. Start of each short index series is indicated with a green dot and its end with a red dot.

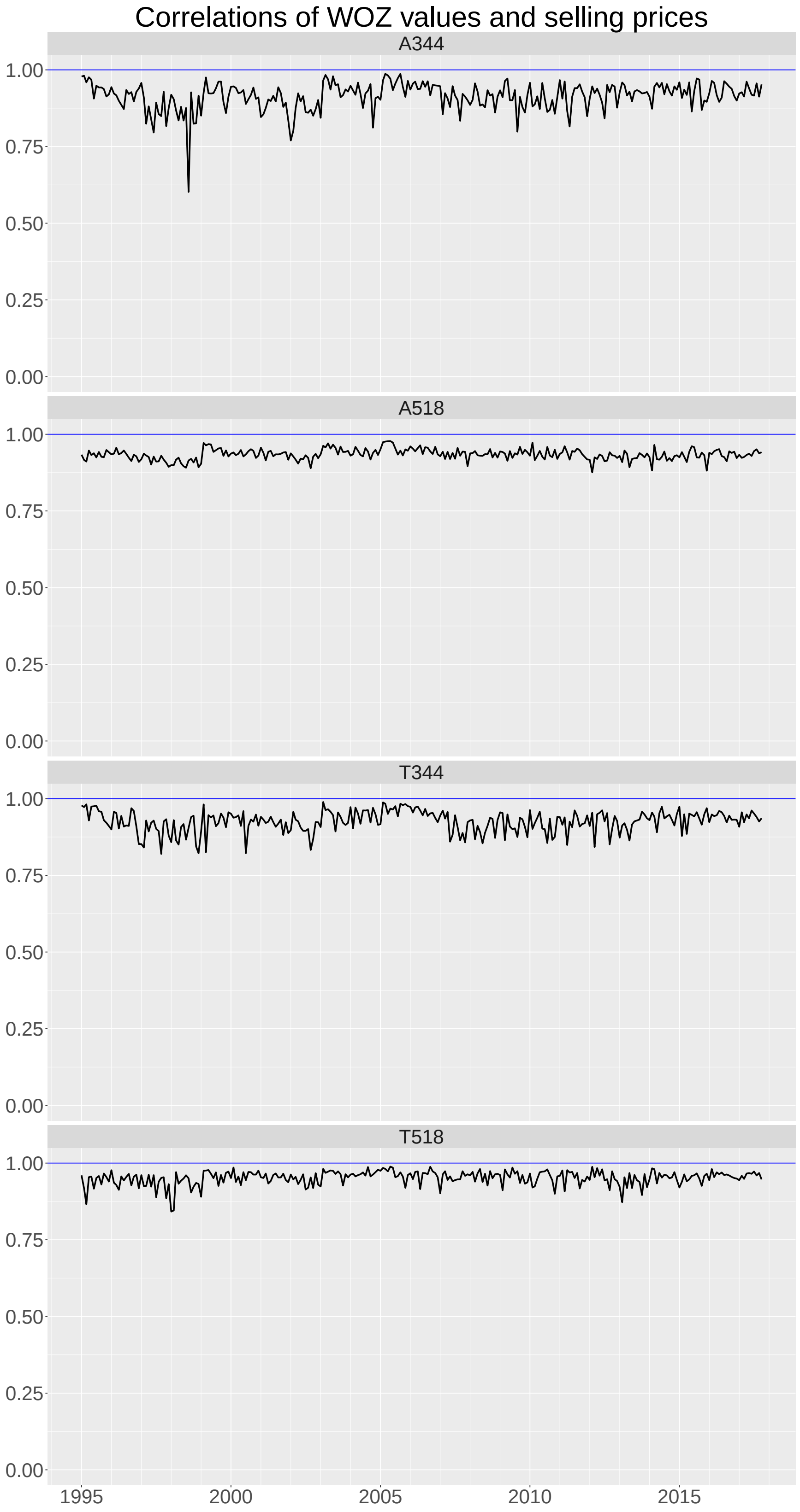


Figure 11.2 Correlations between WOZ valuations and selling prices.

More detailed information about the relationship between WOZ valuations and selling prices of houses can be obtained from Figure 11.3. Here four scatterplots show the relationship, per stratum and municipality, for one particular month (March 2011). The pictures indicate that the assumption to assume a constant ratio of selling price and WOZ valuation, per stratum and per month, is generally pretty good, at least at first sight, although there are obviously exceptions. However, it should not be forgotten that this is a result of the outlier criterion applied. Maybe this criterion should be tightened in order to remove some of these mavericks.

We now start looking at the behaviour of various indices, first the chained versions. The SPAR index is the prime example of such indices. The no WOZ index is included in some figures, not as a serious index, but as a simple, crude index, just to see how it compares to more serious rivals.

Figure 11.4 shows the results of several chained indices applied to the apartments in region 344 for the entire period under study. The SPAR index (in black) serves as the guiding index. Roughly all the indices follow the same line of development. Note how closely HSP (lm) follows the SPAR index. At some point in time the phi ratio index is higher than the SPAR index and the HSP (lm) index. Remarkable is that even the simplest index of this lot - no WOZ - provides the same trend as all the other indices, except that it is ’noisier’. From this picture it seems natural to think that all these indices should be smoothed to give satisfactory results, with the noise removed or dampened.

Figure 11.5 shows the same set of indices as Figure 11.4 but now for apartments in region 518. The HSP (lm) index is still close to the SPAR index and the phi ratio index is still larger than both the SPAR index and the HSP (lm) index. However, the no WOZ index deviates in certain periods a bit more from the SPAR index than in case of Figure 11.4. But overall, over the entire time period, the indices show more or less the same kind of general behaviour.

Figure 11.6 shows the results of the same price indices as in Figure 11.4, but this time for terraced houses in region 344. Here the SPAR index, the HSP (lm) index and the phi ratio index are in good agreement. Only the no WOZ index shows results that on average seem to be a bit lower from a certain moment in time (starting around 2005). And, as before, this index is much ’noisier’ than the other indices considered.

The next figure, Figure 11.7, shows the results for terraced houses in region 518. In this case the results diverge more markedly than in the three other cases considered before. Only the SPAR index and the HSP (lm) are in good agreement. The phi ratio index gives results that are lower than those of the SPAR index and the HSP (lm) index, from about 2001 onwards. Most obviously, the no WOZ index is well above the other three indices for essentially the entire period.

So far we have considered chained indices. Now we start looking at nonchained indices. Because the number of nonchained indices that we consider is too big to plot the results in a single plot (per stratum, region combination) we have split them in two sets. The split was such that indices with the same behaviour were grouped together. To provide a reference index, the SPAR index is included in all figures, although it should be borne in mind that it is a chained index.

The first two figures in the nonchained series are Figure 11.8 and 11.9 showing the results for apartments in region 344. The results in Figure 11.8 seem to be in good agreement with those of the SPAR index. Closer inspection reveals some (minor) temporary deviations, but the overall picture shows considerable agreement as to the general development of the indices.

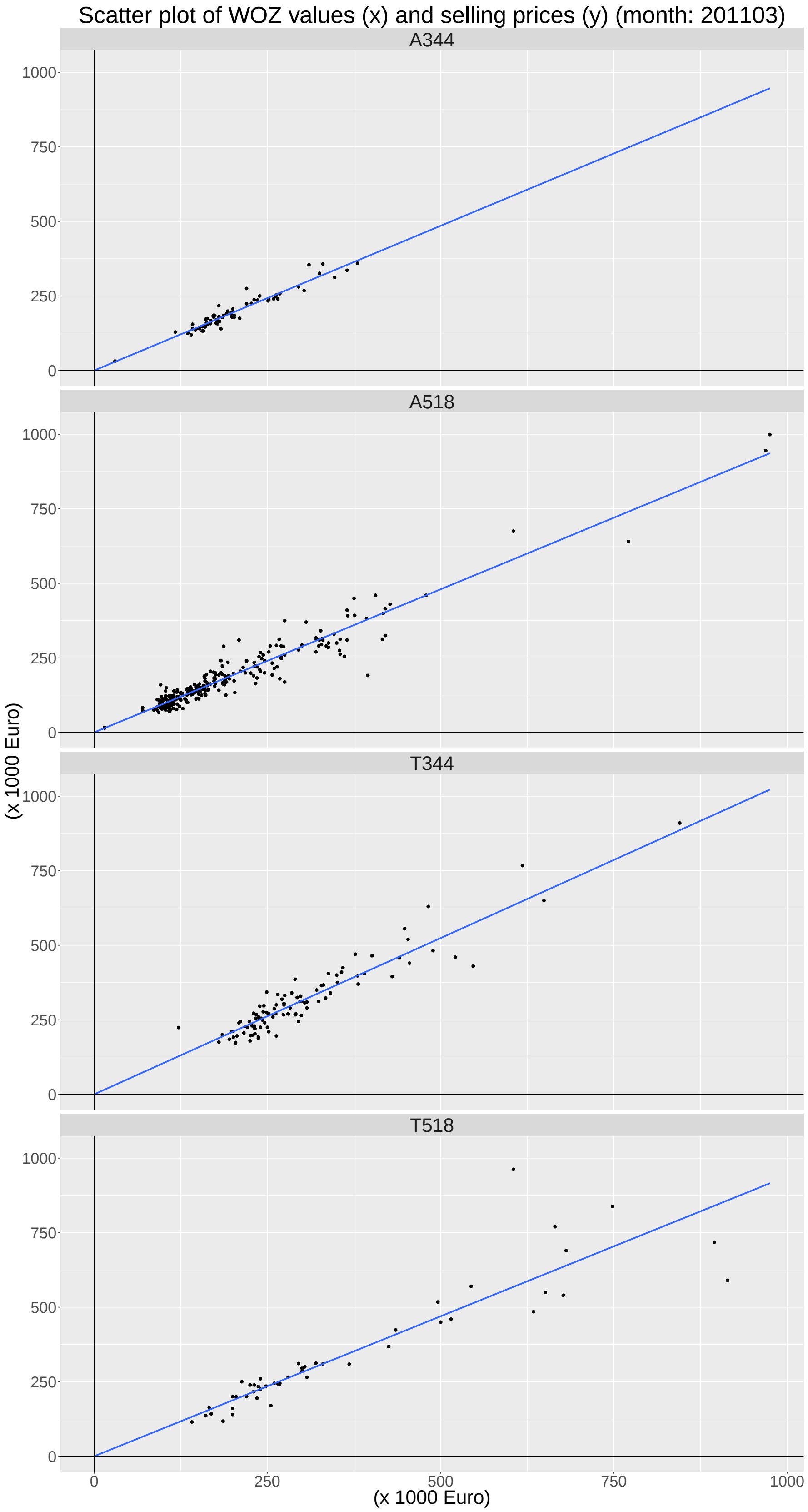


Figure 11.3 Scatterplot of WOZ valuations and house prices in March 2011.

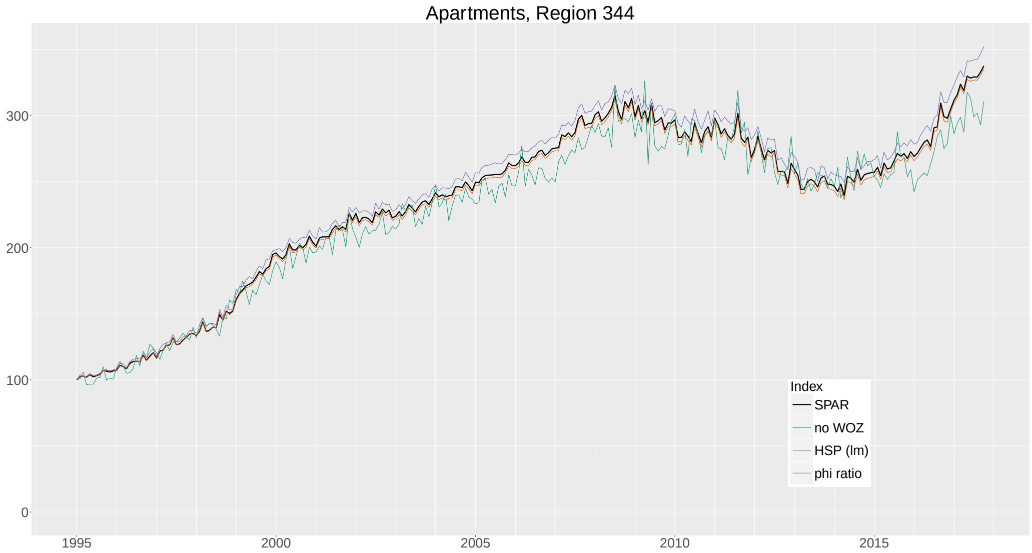


Figure 11.4 Chained price indices for apartments in Region 344.

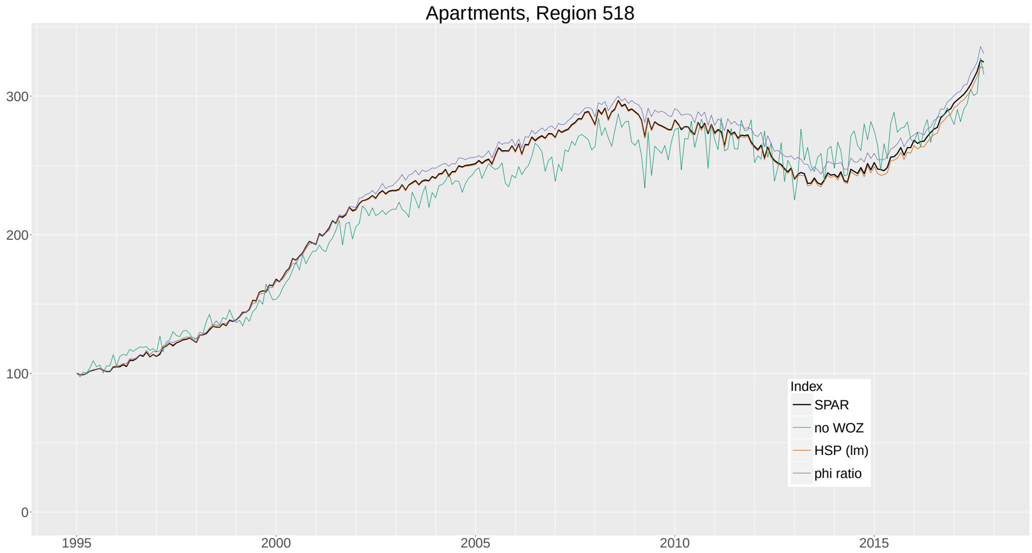


Figure 11.5 Chained price indices for apartments in Region 518.

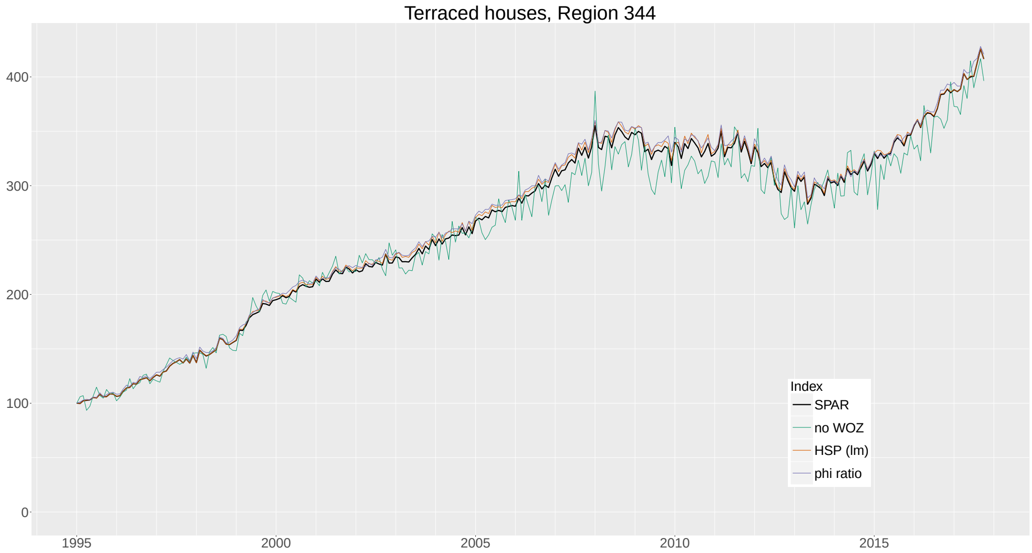


Figure 11.6 Chained price indices for terraced houses in Region 344.

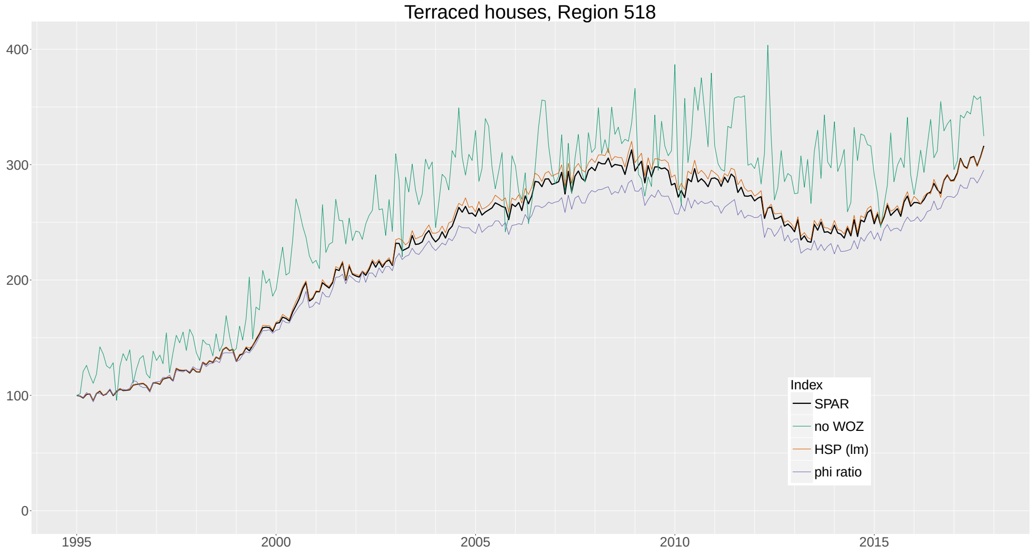


Figure 11.7 Chained price indices for terraced houses in Region 518.

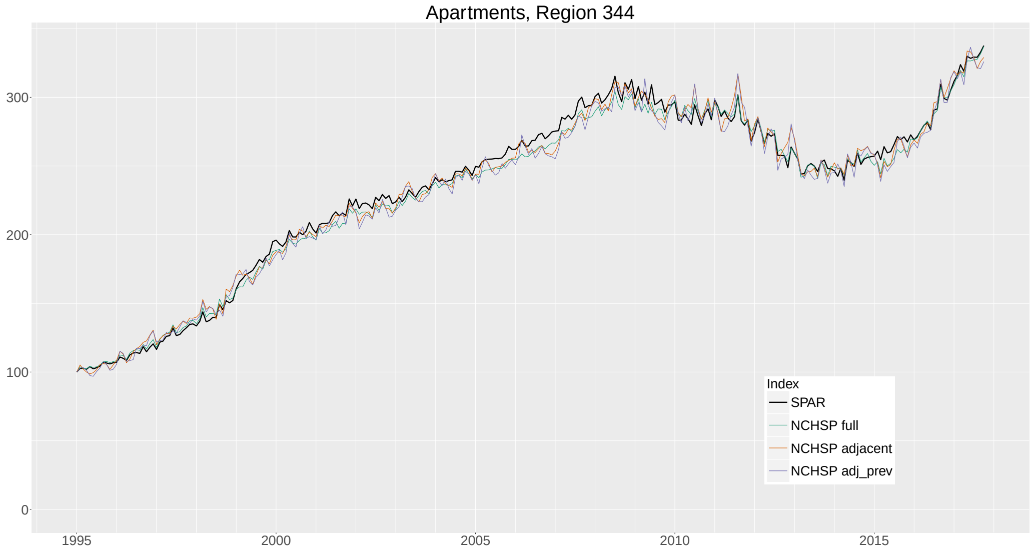


Figure 11.8 Nonchained price indices for apartments in Region 344.

In case of Figure 11.9 the SPAR index on the one hand and the three other indices at the other seem to behave similarly. Only from 2005 until 2008 the three indices - NCHSP incr, NCHSP_pool_max2 and NCHSP_incr_max3 - are slightly below the SPAR index.

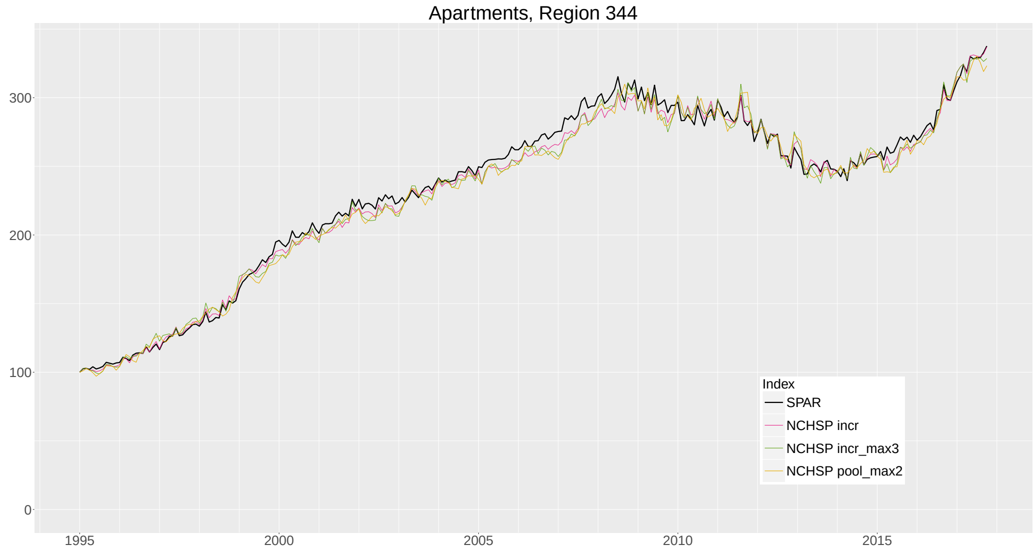


Figure 11.9 More nonchained price indices for apartments in Region 344.

In Figures 11.10 and 11.11 nonchained indices for apartments in region 518 are shown. Again there were too many indices to plot in a single picture, so they have been split into two groups that are homogeneous in terms of the development of the indices in this stratum. As a reference index, the SPAR index was plotted in both figures. If we look at Figure 11.10 we see that the four indices develop very similarly until about 2008. From then on the SPAR index on the one hand and the other three indices - NCHSP full, NCHSP adjacent and NCHSP adj_prev - part ways and develop differently. The values for the three indices are all higher than those of the SPAR index. Of the three NCHSP indices, NCHSP full, generally has the lowest values from about 1999 onwards.

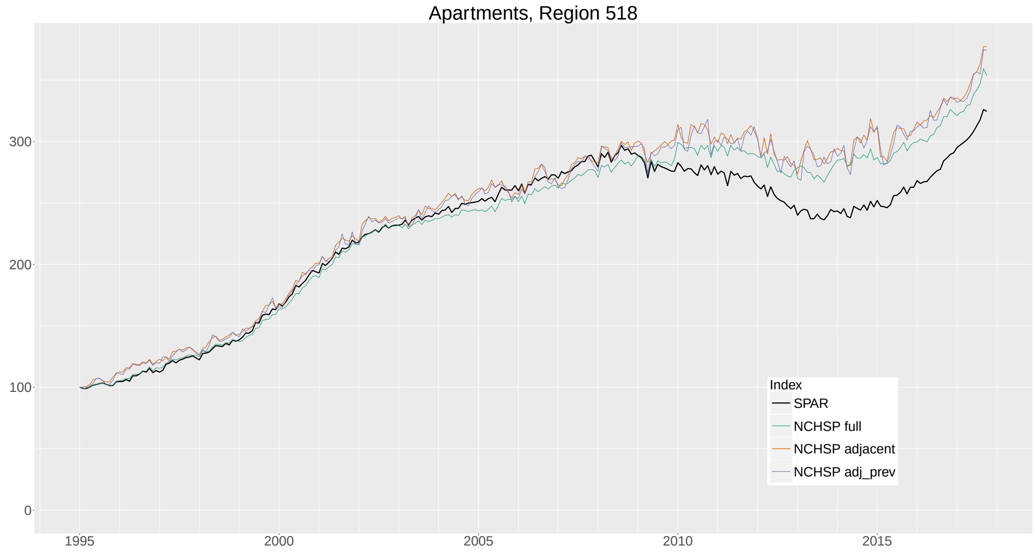


Figure 11.10 Nonchained price indices for apartments in Region 518.

In Figure 11.11 we again see that the nonchained indices - NCHSP incr, and NCHSP incr_max3, NCHSP pool_max2 - collectively behave differently from the SPAR index from 2008 onwards.

From this time onwards their graphs are well above that of the SPAR index. Over the entire period the three NCHSP indices behave similarly. They also behave similarly to the NCHSP indices in Figure 11.10. So the SPAR index is the odd one out here. Again, we did not have enough time to explain the different behaviour of the NCHSP indices on the one hand and the SPAR index on the other.

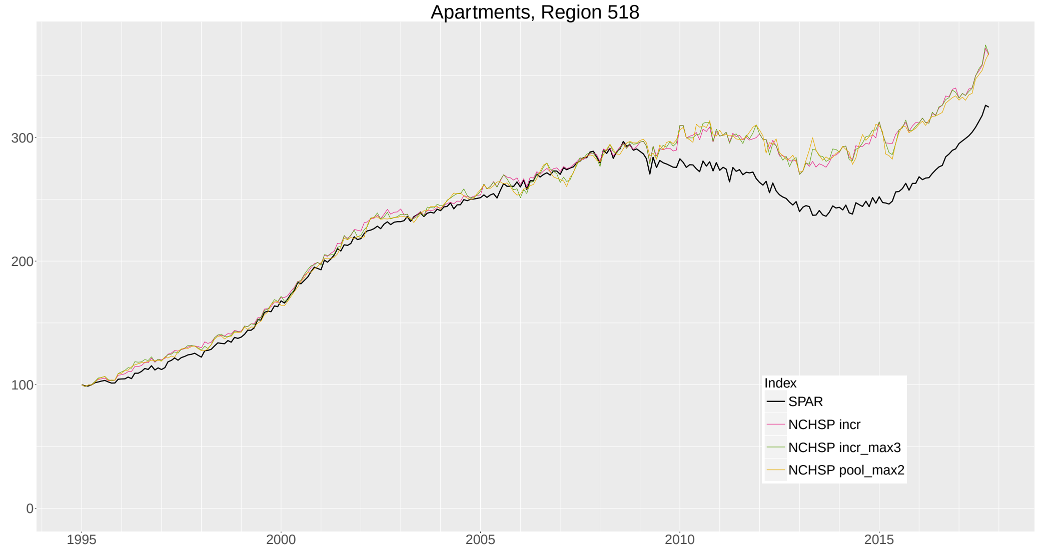


Figure 11.11 More nonchained price indices for apartments in Region 518.

With Figures 11.12 and 11.13 we return to the terraced houses in region 344, and this time we see the results of the nonchained indices. Again we have divided the nonchained indices into two homogeneous groups, and the SPAR index in both cases as the (chained) reference index.

In Figure 11.12 we see that the three NCHSP indices - NCHSP full, NCHSP adjacent and HCSP adj_prev - from 2009 onwards have graphs that tend to be lower than that of the SPAR index. From 1995 until about 2005 these three NCHSP indices and the SPAR index behave rather similarly.

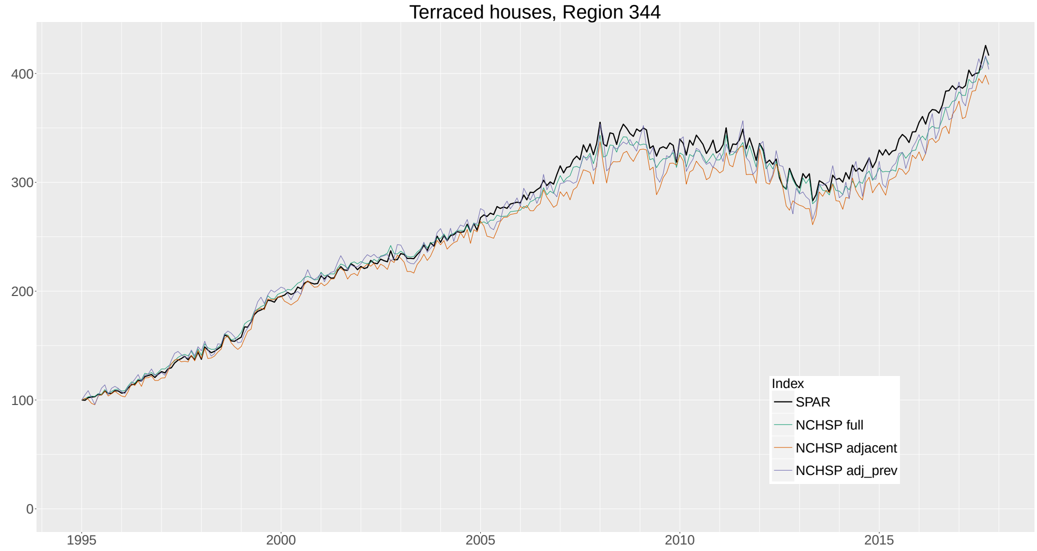


Figure 11.12 Nonchained price indices for terraced houses in Region 344.

If we look at the three other NCHSP indices in Figure 11.13 we see that the three NCHSP indices -

NCHSP incr, NCHSP incr_max3, NCHSP pool_max2 - behave similarly as the three NCHSP indices in Figure 11.12. But roughly all indices in Figures 11.12 and 11.13 have a similar behaviour.

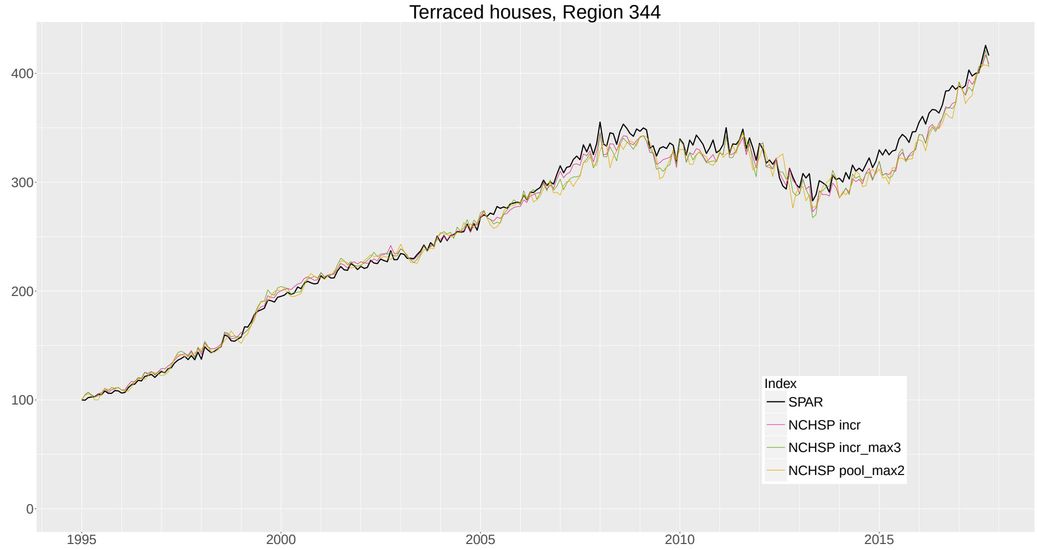


Figure 11.13 More nonchained price indices for terraced houses in Region 344.

With the final two pictures of our series of examples, i.e. Figure 11.14 and Figure 11.15, we see the results for terraced houses in region 518. Figure 11.14 is the most remarkable picture of the entire series that is presented in the present section, as it shows the most diverse results. This time the NCHSP indices do not act in unison as much as in previous cases. Two NCHSP indices yield results that are very similar, namely NCHSP adjacent and NCHSP adj_prev. Until about 2013 NCHSP full developed very similarly to the SPAR index, but rather different from the two NCHSP indices just mentioned. In 2005 the development of NCHSP full drops significantly below the level of the SPAR index. From 2011 onwards the development of NCHSP full is again more in agreement with that of the SPAR index, although it tends to give lower values. Why all this is the case is unclear at the moment and would require a study of the underlying data.

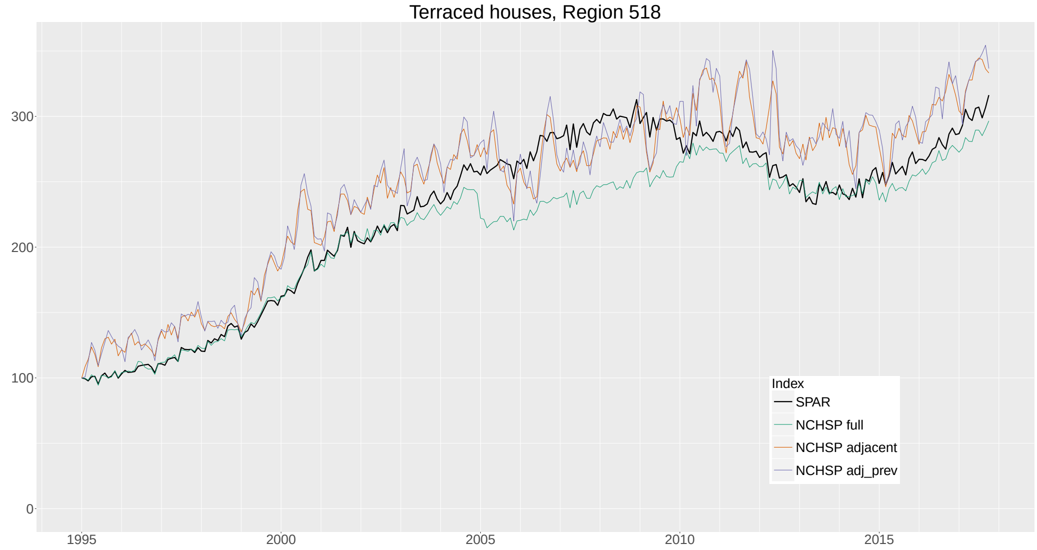


Figure 11.14 Nonchained price indices for terraced houses in Region 518.

In Figure 11.15 we see that the three NCHSP indices - NCHSP incr, NCHSP incr_max3 and NCHSP pool_max2 - develop very similarly over the entire period, but in a way that is different from the

SPAR index. As before, to solve the 'mystery' why this is the case, a deeper study of the underlying data is necessary.

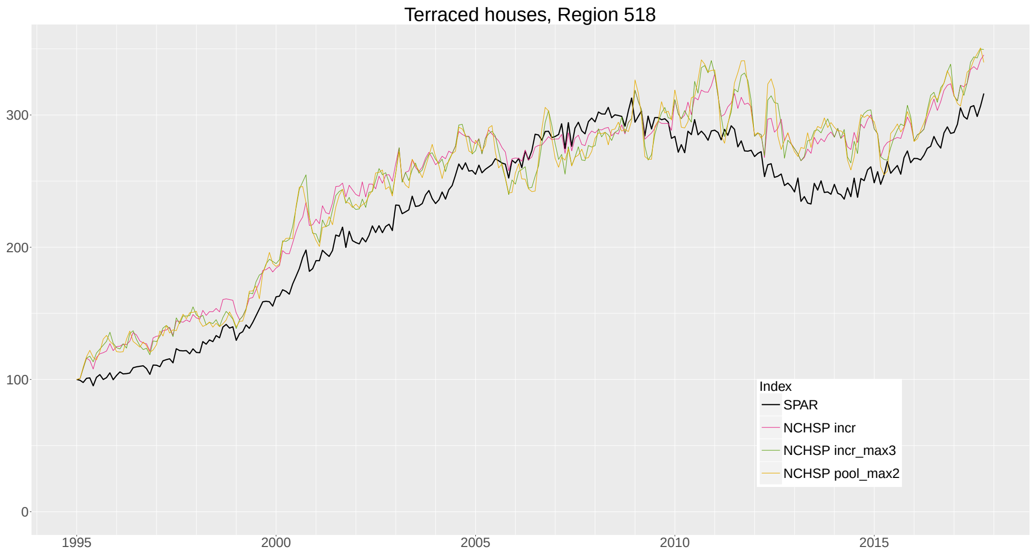


Figure 11.15 More nonchained price indices for terraced houses in Region 518.

The results in Figures 11.14 and 11.15 show rather wild fluctuations. It should be borne in mind that stratum T518 has the smallest number of houses sold of all the four strata considered in the present paper. The SPAR index in this stratum therefore has the relatively largest variance of all the four strata (cf. Willenborg and Scholtus (2018)). This remark probably also applies to the other indices considered. This possibly explains why there are such big differences between the indices in this stratum. But further investigation is needed to confirm (or reject) this hypothesis.

Of the non-chained indices, NCHSP full shows the least amount of fluctuation in these figures. As this index uses information from all houses that are sold in a WOZ period, it is based on a relatively large sample size. It is therefore expected that this index has a relatively small variance. For practical purposes, this index is not useful as it can be computed only at the end of a WOZ period. The incremental approaches do not have this drawback.

12 Discussion

The SPAR index was the catalyst for a quest for a general principle as the basis for price indices to quantify the development of house prices. As a guiding principle we used the idea to compare the selling price of a house with a hypothetical price for the same house in January of the same year, as a reference period for a short, yearly series. This of course raises the question how to estimate such a hypothetical house price. The WOZ valuations of houses are crucial variables for this estimation problem. As is shown hypothetical selling prices can be estimated in various ways, some of which are considered in the present paper. The SPAR index indirectly follows from this approach as well. Using hypothetical prices gives a foundation to a whole class of price indices to quantify house price developments. It seems worthwhile to explore these (and similar) indices in more depth, using real data.

When producing a long index from short indices there is the problem that the further the long series advance, the bigger the influence of the matching factors becomes, in the latter part of the series. In contrast, the methods of Section 10 based on hypothetical selling prices depend only on two consecutive WOZ periods. No cumulation of errors does occur here.

The SPAR index method copes with incomplete data. They complicate matters. So it is a natural idea to consider imputation. Several such models are considered, all exploiting the close relation of selling prices and WOZ valuations of houses sold in the same period and within the same stratum. Using imputation will not only simplify price index computations. Also variance computations by using the bootstrap methods will probably be easier.

The SPAR index first produces short, yearly index series. These short series are chained into a long series. To do this, overlapping months are used in which index numbers are computed with reference to two consecutive WOZ valuations of houses. Two such methods were considered that turned out to produce the same matching factors that are used to accomplish this chaining. As they are critical elements of the long series they should be given special attention. A matching factor that is introduced at a given moment to extend the long series with another year’s data will be part of the long series from that moment onwards. If it is not very good it will ruin the long series from that moment on. So the message we take from this is that the procedure of chaining short index series to a long one should be given careful consideration. Research into alternative methods of chaining seems to be an endeavor to consider.

A possibility is to create a longer period of overlap. This can be achieved by recording the WOZ valuations for houses in other months as well, that is not only for the overlap month December. It would be attractive to store as many WOZ valuations of existing houses as possible. It would then be possible to compute hypothetical selling prices for more periods than two.

Another topic considered in the paper is transitivity. It is shown how transitive closure can be used to compute a transitive price index from the long index series, for each pair of months in a given time window. The application is pretty straightforward. There are no issues here. The reason to include this topic in the present paper is because the concept of transitive closure does not seem to be well-known in ’price index circles’. The notion of transitivity is general and beyond the application to the SPAR index or any other housing price index; cf. Willenborg (2018).

Deviating from the SPAR index in another way, we also consider index methods that directly produce a long series. These methods use the WOZ valuations as auxiliary information to estimate hypothetical selling prices. These prices, along with the selling prices are used to compute indices, without using the WOZ periods. Several models are suggested that fall into this category. In a sense they can also be viewed as modifications of the method that only uses selling prices and no WOZ valuations.

Compared to the CPI, there is another difference when computing indices for house price indices. For the CPI turnover is used (if available) to weigh price information. When one translates this situation to house prices, one should weigh each house with its turnover, which equals $\text{house price} \times \text{quantity} = \text{house price} \times 1$ which is the same value as the house price. This would imply that more expensive houses have a higher weight. If this idea would be applied in a regression context, fitting house prices and WOZ valuation in a stratum and for a particular period (say a month), one would have to resort to weighted least squares, instead of ordinary least squares. We have not investigated this idea in the present paper, but it would be interesting to investigate. It produces at least interesting material for comparison with standard practice.

Having considered quite a number of indices for housing prices, it seems hard to select the best one. Instead of looking for a single index, a more attractive strategy is perhaps to pick a ’lead index’ and a few ’reasonable’ indices and compare the results they yield when applied to real data. If the results are close this should give confidence that the price development of house prices is captured well. If they disagree, it should be an incentive to look at the data more closely in order to find an explanation of the differences found.

In fact there is room for optimizing the choice of the periods considered for some methods. We have chosen the length of a period (for instance in case of an index based on pooled data), but we did not consider an optimal choice of the period length. By chosing a longer period one would get more stable - less noisy - results, but possibly biased estimates. Another extension would be achieved by weighting the months, such that more distant months are weighted less than those that are closer in time. Yet another step that could be considered is to the use of time series models, for instance state space models with a trend component, a seasonal component and a noise component.

The bootstrap method applied in Willenborg and Scholtus (2018) can be used to test to what extent the indices differ significantly from each other. It is very well possible that this would reveal that many of the differences observed in Section 11 are in fact not significant at all if the variances of the indices are taken into account. With the bootstrap method this can be easily investigated.[32] This would be an interesting and important topic for future research.

References

De Haan, J. and R. Hendriks (2013). An appraisal-based generalized regression estimator of house price change. Survey Methodology, 39, pp. 395 - 418.

Devaney, S. and R. Martinez Diaz (2011). Transaction based indices for the UK commercial real estate market: an exploration using IPD transaction data. Journal of Property Research, 28, pp. 269 - 289.

Eurostat (2013). Handbook on residential property price indices (RPPIs). Document that can be downloaded from https://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF.

Eurostat (2017). Technical manual on owner-occupied housing and house price indices. Document that can be downloaded from https://ec.europa.eu/eurostat/documents/7590317/0/Technical-Manual-OOH-HPI-2017/ .

Fisher, J., D. Geltner, and H. Pollakowski (2007). A quarterly transaction-based index (TBI) of institutional real estate investment performance and movements in supply and demand. Journal of Real Estate and Financial Economics, 34, pp. 5 - 33.

Le, Q. (2014). Prijsindex bestaande koopwoning, functionele beschrijving (= Price index existing owner-occupied properties, functional description). Internal report, CBS The Hague.

[32] By determining a 95% confidence interval of the difference of two indices, and checking whether it contains 0.

Willenborg, L. (2018). Transitivity of price indices. Discussion paper, CBS The Hague.

Willenborg, L. and S. Scholtus (2018). Bootstrapping the SPAR index. Discussion paper, CBS The Hague.

Appendix

A Terminology and notation

In this appendix we have collected the specific terminology and notation used in the present paper.

The following terminology is used. Acronyms and abbreviations are presented in parentheses.

- Hypothetical selling price (HSP) : The price of a house if it had been sold in a different month, based on a model relating selling prices per stratum to WOZ valuations of the houses involved.
- SPAR : Sale Price Appraisal Ratio.
- WOZ period : A period in which the WOZ valuations are kept the same. This was four years in the past, but currently it is one year.

The following notation is used in the present paper.

- g: stratum indicator. Stratification is based on type of house and municipality;
- period (j, m): month m of year j. These are the indicators for the periods when houses were sold;
- i: indicator for a house sold;
- V: selling price of a house;
- $V^g_{j,m,i}$: the selling price of house i in stratum g sold in period (j, m). These are values of the variable V;
- $V^g_{j,13,i}$: the selling price of house i in stratum g sold in period $(j + 1, 1)$. This value is used in combination with the WOZ valuation in the previous year; it only applies in case the house already existed then;
- W: WOZ valuation of a house in the year in which it was sold;
- $W^g_{j,m,i}$: WOZ valuation of house i in stratum g sold in period (j, m). This valuation is issued in January of each year for all houses that exist then;
- $H^g_{j,m}$: set of houses i in stratum g sold in period (j, m);
- $HV^g_{j,m}$: set of houses i in stratum g sold in period (j, m) and where the selling price is present (and assumed correct);
- $HW^g_{j,m}$: set of houses i in stratum g sold in period (j, m) and where the WOZ valuation is present (and assumed correct);
- $HVW^g_{j,m} \triangleq HV^g_{j,m} \cap HW^g_{j,m}$: set of houses i in stratum g sold in in period (j, m) and with nonmissing selling price and WOZ valuation;
- $N^g_{j,m} \triangleq |H^g_{j,m}|$: the size of $H^g_{j,m}$;
- $NV^g_{j,m} \triangleq |HV^g_{j,m}|$: the size of $HV^g_{j,m}$;
- $NW^g_{j,m} \triangleq |HW^g_{j,m}|$: the size of $HW^g_{j,m}$;
- $NVW^g_{j,m} \triangleq |HVW^g_{j,m}|$: the size of $HVW^g_{j,m}$;
- $V^g_{j,m} \triangleq \sum_{i \in HV^g_{j,m}} V^g_{j,m,i}$;
- $\overline{V^g_{j,m}} \triangleq \frac{1}{NV^g_{j,m}} \sum_{i \in HV^g_{j,m}} V^g_{j,m,i}$; see (2);
- $W^g_{j,m} \triangleq \sum_{i \in HW^g_{j,m}} W^g_{j,m,i}$;
- $\overline{W^g_{j,m}} \triangleq \frac{1}{NW^g_{j,m}} \sum_{i \in HW^g_{j,m}} W^g_{j,m,i}$; see (3);

- $\phi^g_{j,m} = \frac{1}{NVW^g_{j,m}} \sum_{i \in HVW^g_{j,m}} \frac{V^g_{j,m,i}}{W^g_{j,1,i}}$; see (17);
- $\widehat{W}^g_{j,m,i} = \frac{V^g_{j,m,i}}{\phi^g_{j,m}}$; see (18);
- $\widehat{V}^g_{j,m,i} = \phi^g_{j,m} W^g_{j,1,i}$; see (19);
- $s^g_{j,m} = \frac{\overline{V^g_{j,m}}}{\overline{W^g_{j,m}}} / \frac{\overline{V^g_{j,1}}}{\overline{W^g_{j,1}}}$; see (1);
- $s^g_{j,13} \triangleq \frac{\overline{V^g_{j+1,1}}}{\overline{W^g_{j+1,1}}} / \frac{\overline{V^g_{j,1}}}{\overline{W^g_{j,1}}}$; see (44);
- $\bar{W}^g_\nu$: the average WOZ value of the houses in the WOZ-period, not sold in month ν;
- k_ν: the number of houses in the WOZ-period considered that are not sold in month ν;
- $\pi^{\mu,\nu} = \frac{\sum_{i \in H^g_\nu} V^g_{\nu,i} + k_\nu \hat{\phi}^g_\nu \bar{W}^g_\nu}{\sum_{i \in H^g_\mu} V^g_{\mu,i} + k_\mu \hat{\phi}^g_\mu \bar{W}^g_\mu}$; see (65);
- $\pi^{\mu,\nu}_H = \frac{\hat{\phi}^g_\nu}{\hat{\phi}^g_\mu}$; see (66).
- □: denotes the end of a remark.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information

© Statistics Netherlands, The Hague/Heerlen/Bonaire 2018.
Reproduction is permitted, provided Statistics Netherlands is quoted as the source