A Hedonic Price Index Number for New One-Family Detached Houses in Finland

Suoperä, Antti, Statistics Finland
Vuorio, Johanna, Statistics Finland
2019

Abstract

Statistics Finland has a relatively long experience in constructing price indices of old dwellings and old one-family detached houses using the hedonic approach where regression analysis and classification of flats are combined. The bilateral price-links cannot be used for new one-family detached houses because each house exists only once in data. So in outline, we use similar methods as in new blocks of flats and terraced houses and all statistical methods, price aggregation and index number methods which follow study of Vartia, Suoperä and Vuorio (2019). We show that quality adjustment is necessary and quality differences should be removed from index series based on different elementary aggregates of prices. In this study, we use some basic and excellent index number formulas. We show that differences between them is not so severe as in the case of new blocks of flats and terraced houses.

Our test data is collected from register maintained by tax authorities. The data includes statistical unit specific information on prices, quantities, values and some unit specific quality characteristics from 2015/I to 2019/II being quarter data. The quarter data includes about 300 – 600 observation per a quarter.

1 Introduction

Each observation appears only once in the data so the bilateral price links for new one-family detached house prices (matched pairs) cannot be applied. The price index for new one-family detached houses follows precisely the method of 'Hedonic Price Index Number for New Blocks of Flats and Terraced Houses in Finland'
(Vartia, Suoperä & Vuorio, 2019).

Data includes information of unit values (i.e. prices), quantities, values and some quality characteristics of detached houses from 2015/I to 2019/II being quarter data. The prices include all building costs but not cost of a building lot. The quarter data includes about 300 – 600 observation per a quarter. The data is very narrow and do not tolerate fine classification of observations.

2 Basic Concepts, Notation and the Index Number Problem

See, 'Hedonic Price Index Number for New Blocks of Flats and Terraced Houses in Finland'
(Vartia, Suoperä & Vuorio, 2019).

2.1 Basic Concepts and Notation

See, 'Hedonic Price Index Number for New Blocks of Flats and Terraced Houses in Finland'
(Vartia, Suoperä & Vuorio, 2019).

2.2 The Index Number Problem

We formulate 10 questions, which together specify what the index number problem looks like (Vartia, 1976, pp. 92-95). In this study we analyze new one-family detached house prices and because each observation appears only once in data, 10 questions are presented for some partition of one-family detached houses. Our partition is a cartesian product of NUTS 2 regions and construction-type of building; that is 4*3 = 12 stratums.

Intended use of the price index

1. Characterization of the set A of commodities?: One-family detached houses.
2. Definition of the economic agents?: Household.
3. Definition of the time periods?: Quarter of a year.

Preliminary specification of the relevant information

4. Definition of the elementary or stratum aggregates, EA?: Unweighted and weighted arithmetic and geometric averages, that is; uA, uG, wA and wG respectively.
5. Prices for elementary or stratum aggregates?: Unit prices for unweighted and weighted arithmetic and geometric averages. Observed unit prices for *i* in period *t* is $p_i^t = v_i^t/q_i^t$, where v_i^t is total value of construction type and q_i^t is total square meters build.
6. What kind of weights are used? Consistent weights for each elementary or stratum aggregates.

Technical choices

7. Index number formula?: Two set of formulas - basic and excellent (Vartia & Suoperä, 2017, 2018)
8. Construction strategy?: In this study we use the base strategy, where the base period is defined to be the previous year, which is normalized as an average quarter and the observation period is the quarter of current year. We use similar strategy as in Finnish CPI applied to a scanner-data complete micro data (see, Vartia, Suoperä, Nieminen & Montonen, 2018a, 2018b).
9. Treatment of quality changes?: Hedonic method, see, 'Hedonic Price Index Number for New Blocks of Flats and Terraced Houses in Finland' (Vartia, Suoperä & Vuorio, 2019).
10. Treatment of the new or vanishing commodities or one-sided nulls?: Does not emerge.

2.2.1 Quality Changes

See, 'Hedonic Price Index Number for New Blocks of Flats and Terraced Houses in Finland' (Vartia, Suoperä & Vuorio, 2019).

3 Analysis of Heterogeneous Cross-sectional Data

Our analysis combines classification and typical regression analysis. In statistical terms, the method combines analysis of variance and typical regression analysis and is called analysis of covariance model (se FE-model Hsiao, 1986 p.29-32).

3.1 Classification of Observations

We examine time periods $t = 0,1,\ldots,T$ and the finite set $A = \{a_1, a_2, \ldots, a_{n_t}\}$ of new one-family detached houses. Each observation appears only once in the data and bilateral price links for homogeneous one-family detached houses (matched pairs) cannot be applied. The solution to tackle this situation is to form classification of observations into micro classes or stratum A_k, $k = 1,\ldots,K$, so that $A_k \cap A_r = \emptyset,\ \forall\, k \neq r$ and $A = \bigcup_{k=1}^{K} A_k$.

In this study the partition is based on four separate NUTS 2 regions which are divided into three different types of constructions: 1. self-motivated, 2. modular house and 3. turnkey construction. So, the partition includes 12 separate micro class at all time period *t.* The idea is to aggregate the observations into strata level and then to calculate some appropriate indicator for price change for each stratum (See, 'Hedonic Price Index Number for New Blocks of Flats and Terraced Houses in Finland' (Vartia, Suoperä & Vuorio, 2019).

3.3 Estimation of the Price Model

The estimation of unknown parameters follows the ordinary-least-squares (OLS) method (see more details 'Hedonic Price Index Number for New Blocks of Flats and Terraced Houses in Finland' (Vartia, Suoperä & Vuorio, 2019).

3.4 Estimation Results of the Price Model

Our regression analysis for heterogeneously behaving cross-sections is standard statistical inference familiar to most statisticians. We estimate four NUTS 2 regional equations each having three explanatory variables and three dummy-variables according to three construction type, that is; 1. self-motivated, independent or 'do-it-yourself' building, 2. modular house construction and 3. turnkey construction. We specify the model as linear in respect to parameters

(1) $$\log(p_{irt}) = \alpha_{r1t} + \cdots + \alpha_{r3t} + \boldsymbol{x}'_{irt}\boldsymbol{\beta}_{rt} + \varepsilon_{irt}$$

where $\log(p_{irt})$ represents house *i* specific logarithmic unit value per square meter in region *r* in period *t*. The 3-dimensional vector of parameters $\boldsymbol{\beta}_{rt}$ may vary according to regional grouping and time. Parameters $\alpha_{r1t}, \dots \alpha_{r3t}$ present construction type effects in the region *r* and period *t*. The 3-dimensional vector $\boldsymbol{x}'_{irt}$ consists of exogenous independent variables (i.e. quality characteristics) that are presented in Table 1. Term ε_{irt} is random error term having $E\left(\varepsilon_{irt}\middle|\boldsymbol{x}'_{irt}\right) = 0$ and $Var\left(\varepsilon_{irt}\middle|\boldsymbol{x}'_{irt}\right) = \sigma^2_{rt} < \infty$.

Table 1: The exogenous independent variables used in the NUTS 2 regional price models

Explanatory variables	Description of variable
Building type dummies	Classify observations into three building types: self-motivated building, modular house and turnkey construction for each NUTS 2 region
x_1	Total number of square meters of a building
$x_2 = sqrt(x_1)$	Square root of total number of square meters of a building
x_3	Square meters of non-living area :total number of square meters of a building minus living area of a building –.

Table 2: The estimation results for the price models in years 2017 and 2018 (*t* = 0 refers the whole year and other *t* = 1, 2, 3, 4 quarters of corresponding year), see details Suoperä & Vartia, 2011.

	2017	2017	2017	2017	2017	2018	2018	2018	2018	2018
t	1	2	3	4	0	1	2	3	4	0
n_t	536	569	408	532	2045	554	492	421	592	2059
Strata	12	12	12	12	12	12	12	12	12	12
Adjr2	0.7062	0.734499	0.7263	0.758342	0.725472	0.7413	0.760596	0.781103	0.737459	0.74327
RMSE	0.2599	0.262316	0.26777	0.223988	0.256667	0.229286	0.249363	0.215877	0.222604	0.23511
$\hat{\alpha}_t$	7.2638	8.437672	8.82991	8.125374	8.128393	7.687112	7.64421	8.469305	8.303864	8.07375
se($\hat{\alpha}_t$)	(0.4822)	(0.4246)	(0.4787)	(0.4546)	(0.2262)	(0.4883)	(0.4912)	(0.4347)	(0.4794)	(0.2402)
$\hat{\beta}_{1t}$	0.00131	0.007577	0.00889	0.006336	0.005777	0.002847	0.002645	0.006632	0.00647	0.00482
se($\hat{\beta}_{1t}$)	(0.0029)	(0.0025)	(0.0028)	(0.0027)	(0.0013)	(0.0029)	(0.0029)	(0.0025)	(0.0029)	(0.0014)
$\hat{\beta}_{2t}$	0.00021	-0.17907	-0.22164	-0.13101	-0.12733	-0.05764	-0.04921	-0.16967	-0.14997	-0.11219
se($\hat{\beta}_{2t}$)	(0.0748)	(0.0639)	(0.0726)	(0.0696)	(0.0345)	(0.0749)	(0.0746)	(0.0664)	(0.0744)	(0.0368)
$\hat{\beta}_{3t}$	-0.0061	-0.00637	-0.00617	-0.00713	-0.00631	-0.00535	-0.00482	-0.0058	-0.00627	-0.00571
se($\hat{\beta}_{3t}$)	(0.0008)	(0.0008)	(0.0009)	(0.0008)	(0.0004)	(0.0008)	(0.0009)	(0.0008)	(0.0007)	(0.0004)
γ_t	1	1	1	1	1	1	1	1	1	1
se(γ_t)	(0.0280)	(0.0259)	(0.0330)	(0.0260)	(0.0141)	(0.0254)	(0.0271)	(0.0299)	(0.0262)	(0.0141)
$\boldsymbol{\lambda}_t$	1	1	1	1	1	1	1	1	1	1
se($\boldsymbol{\lambda}_t$)	(0.035)	(0.0281)	(0.039)	(0.050)	(0.0172)	(0.0275)	(0.0429)	(0.0363)	(0.0284)	(0.0243)

For the estimation method for results in Table 2 see details in Suoperä & Vartia, 2011, and Vartia, Suoperä and Vuorio, 2019. The estimation results are summarized as:

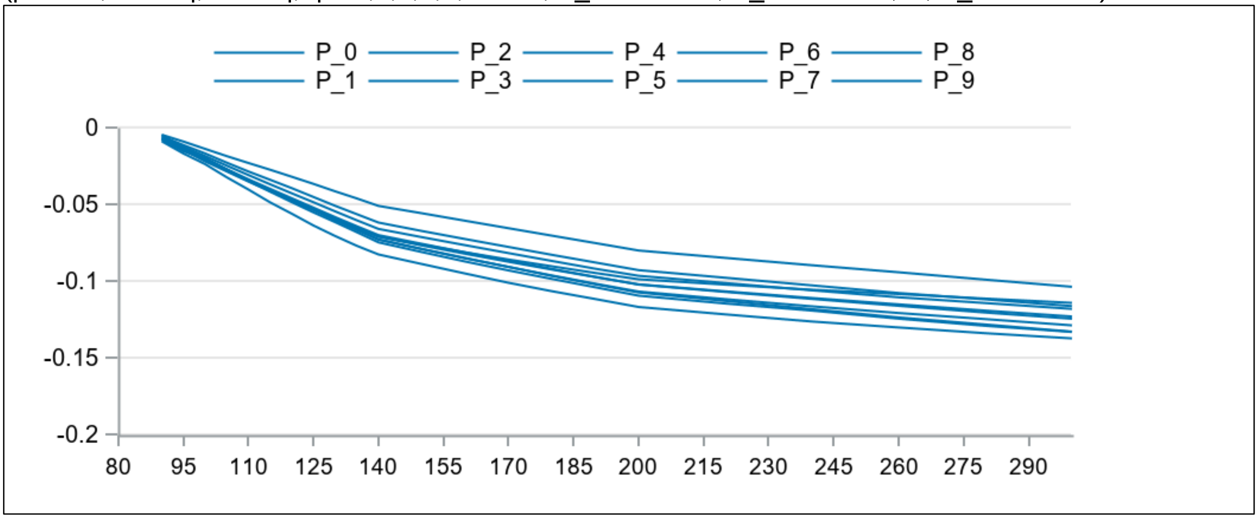
1. The most of estimates for representative behavior are statistically significant.
2. 'Construction-type' indicators (i.e. γ_t) for different NUTS 2 regions are strongly significant and must include into the price models
3. The data should be analyzed by heterogeneously behaving cross-section (i.e. time and regionally varying betas; in Table parameters $\boldsymbol{\lambda}_t$).
4. Quality characteristics (i.e. $x_1$and x_3) have negative effect on prices.

To interpret the estimation results of eq. (1) (see Vartia, Suoperä & Vuorio, Appendix 1, 2019) we take partial derivatives with respect to number of square meters, that is

$$\frac{\partial \log (p_{irt})}{\partial x_{i1rt}} = \hat{\beta}_{1rt} + \hat{\beta}_{3rt} + \hat{\beta}_{2rt}/sqrt(x_{i1rt}), \forall\, i \in A_r.$$

When we calculate cumulative sum from ordered sequence of partial derivates (i.e. ordered according to x_{i1} starting from smallest), we get Figure 1. In Figure 1 we see that square meter prices fall when number of squares increase. Square meter price of a 300 m^2 one-family detached house is about 10 – 15 log-% lower compared to a 90 m^2 house.

Figure 1: The price effect of size on the square meter price of new one-family detached house in Finland (periods, 2017.q, 2018.q, q = 0,1,2,3,4; that is, P_0 = 2017.0, P_0 = 2017.1,…, P_9 = 2018.4).



4 Price Aggregation from Observations into Stratum Aggregates

Price aggregation from observation into strata level follows methods derived in Vartia, Suoperä & Vuorio (2019). The starting point for price aggregation is semilogarithmic price models that are estimated by the OLS method. We derive four elementary aggregates for our partition: 1. unweighted geometric (*uG*), 2. unweighted arithmetic (*uA*), 3. weighted geometric (*wG*) and 4. weighted arithmetic average prices (*wA*). The study Vartia, Suoperä and Vuorio in Appendix 2 (2019) presents the derivation of these elementary aggregates. Elementary aggregates from 2 to 4 are new aggregation results for semilogarithmic price models. They all are based on reparameterization of the OLS estimates for construction-type effects and are also new results for price-ratio decomposition for different stratum aggregates.

Table 2 defines all that is necessary for construction of hedonic price index for stratum $A_k, k = 1, \dots, K$ (derivations are presented in Appendix 1 and 2 in study Vartia, Suoperä and Vuorio, 2019)

The stratum or elementary aggregates in Table 2 are used to construct the so called Oaxaca decomposition (Oaxaca, 1973). The decomposition leads to strata level indices for actual average price ratios, quality changes and quality adjusted price ratios.

Table 2: Price aggregation from observations into stratum level for a semilogarithmic model estimated by the OLS (note: $\overline{\boldsymbol{x}}'^{uG}_{kt}$ is noted here as equally weighted arithmetic average for quality characteristics).

Statistics	Mathematical formula for weights w_{ikt}, $\forall i \in A_k$	Stratum aggregate
Unweighted arithmetic average (*uA*)	$w_{ikt}^{uA} = \frac{L(p_{ikt}, 1)}{L(\sum_i p_{ikt}, n_t)}$	$\log(\bar{p}_{kt}^{uA}) = \hat{\alpha}_{kt}^{uA} + \overline{\boldsymbol{x}}'^{uA}_{kt}\widehat{\boldsymbol{\beta}}_{kt}$, where $\hat{\alpha}_{kt}^{uA} = log(\bar{p}_{kt}^{uA}) - \overline{\boldsymbol{x}}'^{uA}_{kt}\widehat{\boldsymbol{\beta}}_{kt}$
Weighted arithmetic average (*wA*)	$w_{ikt}^{wA} = \frac{L(v_{ikt}, q_{ikt})}{L(\sum_i v_{ikt}, \sum_i q_{ikt})}$	$\log(\bar{p}_{kt}^{wA}) = \hat{\alpha}_{kt}^{wA} + \overline{\boldsymbol{x}}'^{wA}_{kt}\widehat{\boldsymbol{\beta}}_{kt}$, where $\hat{\alpha}_{kt}^{wA} = log(\bar{p}_{kt}^{wA}) - \overline{\boldsymbol{x}}'^{wA}_{kt}\widehat{\boldsymbol{\beta}}_{kt}$
Unweighted geometric average (*uG*)	$w_{ikt}^{uG} = \frac{1}{n_t}$	$\log(\bar{p}_{kt}^{uG}) = \hat{\alpha}_{kt}^{uG} + \overline{\boldsymbol{x}}'^{uG}_{kt}\widehat{\boldsymbol{\beta}}_{kt}$, where $\hat{\alpha}_{kt} = log(\bar{p}_{kt}^{uG}) - \overline{\boldsymbol{x}}'^{uG}_{kt}\widehat{\boldsymbol{\beta}}_{kt}$
Weighted geometric average (*wG*)	$w_{ikt}^{wG} = \frac{q_{ikt}}{\sum_i q_{ikt}}$	$\log(\bar{p}_{kt}^{wG}) = \hat{\alpha}_{kt}^{wG} + \overline{\boldsymbol{x}}'^{wG}_{kt}\widehat{\boldsymbol{\beta}}_{kt}$, where $\hat{\alpha}_{kt}^{wG} = log(\bar{p}_{kt}^{wG}) - \overline{\boldsymbol{x}}'^{wG}_{kt}\widehat{\boldsymbol{\beta}}_{kt}$

5 Hedonic Price Index Numbers and its Decomposition

The key problem of the chain type indices (i.e. strategy) is the *chain error* (or chain drift) that tends to grow when chaining is applied frequently – typically on a monthly or quarterly basis. So, because chain error is data contingent and realized only for a chain-type strategy, we abandon the chain and favor the base strategy. In this study we use the base strategy, where the base period is defined to be the previous year, which is normalized as an average quarter. In fact, similar strategy is used in Finnish CPI applied to a scanner type complete micro data (see, Vartia, Suoperä, Nieminen & Montonen, 2018a, 2018b; Vartia, Suoperä, Nieminen & Markkanen, 2019).

No index number formulas are needed when decomposing average price ratios in strata level into quality corrections and quality adjusted price changes. When these stratum statistics are aggregated into crude categories, the use of index number formulas is necessary. We analyze two sets of index number formulas. The first set is based on formulas using old or new weights and are called as *basic set of index numbers*. Laspeyres (*L*) and Log-Laspeyres (*l*) uses base period weights (i.e. old weights) and Log-Paasche (*p*) and Paasche (*P*) instead uses observation period weights (i.e. new weights). The second set of index numbers include four formula: Montgomery-Vartia (*MV*), Törnqvist (*T*), Fisher (*F*) and Sato-Vartia (*SV*). We call these index number formulas as *excellent formula*. For the fundamental analysis of these index number formula, see Vartia & Suoperä, 2018.

5.1 Within Stratum Hedonic Quality Adjustment and its Decomposition

For simplicity, we have two time periods, the base (t = 0, previous year) and the observation quarter of current year (t = 1) and only one stratum A_k. We calculate the difference between these two price model (0, t) separately for each stratum aggregate indexed by sub index $m = uA, wA, uG, wG$ (see Table 2), that is

$$\log(\bar{p}_{k1}^{m}/\bar{p}_{k0}^{m}) = \hat{\alpha}_{kt}^{m} + \bar{\boldsymbol{x}}'^{m}_{kt}\hat{\boldsymbol{\beta}}_{kt} - \hat{\alpha}_{k0}^{E} - \bar{\boldsymbol{x}}'^{m}_{k0}\hat{\boldsymbol{\beta}}_{k0}$$

Defining first Oaxaca decomposition (1973) and then exp-transformation, we get

$$\bar{p}_{k1}^{m}/\bar{p}_{k0}^{m} = exp\{(\bar{\boldsymbol{x}}'^{m}_{k1} - \bar{\boldsymbol{x}}'^{m}_{k0})\hat{\boldsymbol{\beta}}_{k1}\} \cdot\ exp\{\hat{\alpha}_{k1}^{m} - \hat{\alpha}_{k0}^{m} + \bar{\boldsymbol{x}}'^{m}_{k0}(\hat{\boldsymbol{\beta}}_{kt} - \hat{\boldsymbol{\beta}}_{k0})\} \quad (2)$$

The left side in equation (2) is simply the price ratio for a given stratum aggregate m. We call this price ratio as '*actual or true price change*' (= A) for aggregate m (see Table 2). The first term on right side is a '*price change due to quality difference*' (= QC) of the sample mix at current year observation period valuation of the characteristics. The second term on right is a '*quality adjusted price change*' (= QA) evaluated at *standard point of quality*, that is $\bar{\boldsymbol{x}}'^{m}_{k0}$.

Index number formula is not needed for the compilation of price change for stratum A_k– the direct price link from period 0 to period t is based purely on 'average statistics' for separate stratum aggregates. The left side of equation (2) simply tells that average prices in base 0 and observation period t depends on corresponding averages of quality characteristics. When these vectors of averages $(\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_t)$ are not equal then the price ratio is not based on 'commodities' that are comparable in quality. The price ratio of averages of 'actual or true prices', that is $\bar{p}_A^{t/0} = \bar{p}^t(\bar{\boldsymbol{x}}_t)/\bar{p}^0(\bar{\boldsymbol{x}}_0)$ is not a satisfactory choice for official price change statistics – the quality differences should be removed from it. We have two main choice for the standard quality point of quality characteristics – either $\bar{\boldsymbol{x}}_0$ or $\bar{\boldsymbol{x}}_t$. The base period standard quality point $\bar{\boldsymbol{x}}_0$ is estimated from previous year and $\bar{\boldsymbol{x}}_t$ from quarter of current year. Previous year (i.e. period 0) includes about a fourfold of observations compared to observation quarter and so statistical properties of $\bar{\boldsymbol{x}}_0$ favor selecting it as standard quality point. The following table collect together all information from eq. (2) in terms of prices.

Table 3: Components of equation (2) in price terms for any stratum aggregate separately.

	Price ratio	Explicit formula
Actual or true price change, $p_A^{t/0}$	$\bar{p}^t(\bar{\boldsymbol{x}}_t)/\bar{p}^0(\bar{\boldsymbol{x}}_0)$	$exp\{(\hat{\alpha}_t + \bar{\boldsymbol{x}}'_t\hat{\boldsymbol{\beta}}_t) - (\hat{\alpha}_0 + \bar{\boldsymbol{x}}'_0\hat{\boldsymbol{\beta}}_0)\}$
Quality adjusted price change, $p_{QA}^{t/0}$	$\bar{p}^t(\bar{\boldsymbol{x}}_0)/\bar{p}^0(\bar{\boldsymbol{x}}_0)$	$exp\{(\hat{\alpha}_t + \bar{\boldsymbol{x}}'_0\hat{\boldsymbol{\beta}}_t) - (\hat{\alpha}_0 + \bar{\boldsymbol{x}}'_0\hat{\boldsymbol{\beta}}_0)\}$
Price change for quality correction of square meter, $p_{QC,x_1}^{t/0}$	$\bar{p}_{x_1}^t(\bar{x}_{1t}, \bar{x}_{2t})/\bar{p}_{x_1}^0(\bar{x}_{10}, \bar{x}_{20})$	$exp\{(\bar{x}_{1t}\hat{\beta}_{1t} + \bar{x}_{2t}\hat{\beta}_{2t}) - (\bar{x}_{10}\hat{\beta}_{1t} + \bar{x}_{20}\hat{\beta}_{2t})\}$
Price change for quality correction of non-living area (square meter), $p_{QC,x_3}^{t/0}$	$\bar{p}_{x_3}^t(\bar{x}_{3t})/\bar{p}_{x_3}^0(\bar{x}_{30})$	$exp\{(\bar{x}_{3t}\hat{\beta}_{3t} - \bar{x}_{30}\hat{\beta}_{3t})\}$

Now the equation (2) may expressed by price ratios, that is

$$(3) \qquad p_A^{t/0} \equiv p_{QC,x_1}^{t/0} \cdot p_{QC,x_3}^{t/0} \cdot p_{QA}^{t/0}$$

Each price ratio in (3) is estimated separately for a given price aggregate (i.e. $m = uA, wA, uG, wG$). Each x-variable - size of dwelling in square meters, distance and share of owner of a building lot - have negative effect on prices. When these x-variables are smaller than their standard quality point (i.e. $\bar{\boldsymbol{x}}_t - \bar{\boldsymbol{x}}_0 < \mathbf{0}$) quality corrections $p_{QC,x_1}^{t/0}$, $p_{QC,x_3}^{t/0}$ exceeds one. This means that we need to adjust actual price change $p_A^{t/0}$ just amount of quality corrections $p_{QC,x}^{t/0} = p_{QC,x_1}^{t/0} \cdot p_{QC,x_3}^{t/0}$ downward to get quality adjusted price change $p_{QA}^{t/0}$ that is standardized of quality. Decomposition (3) is calculated separately for all four stratum or elementary aggregates.

5.2 Hedonic Price Index Numbers

The equation (3) simply says that 'actual or true price change' (i.e. sub index A) of a given aggregate is divided in our case into three components – two of them are due to quality change (i.e. sub index QC) and last one is 'quality adjusted price change' (= QA). Equations (2) and (3) are identities for any aggregate in question. For simplicity we have defined the price-link from period 0 to period 1 in eq (2). Replacing period 1 by t we get our base strategy familiar to our CPI analysis based on scanner-type complete micro data (see for example Vartia, Suoperä, Nieminen & Montonen, 2018a, 2018b). In this strategy the base period is the previous year normalized as average quarter and observation period t is a quarter of current year. This is the most natural choice, because it is free of chain error (or drift).

Index number theory begin by aggregation of decomposition of equation (3). We analyze two sets of index number formulas. The first set is based on formulas using old (Laspeyres L and Log-Laspeyres l) or new weights (Log-Paasche p and Paasche P). These index number formulas are called as *basic formulas*. These formulas are data contingently biased and should not be used. The second set of index numbers include four formulas: Montgomery-Vartia, *MV*, Törnqvist, *T*, Fisher, *F* and Sato-Vartia, *SV*). We call these index number formulas as *excellent formulas*. The fundamental analysis of these index number formulas, see Vartia & Suoperä (2018).

In Table 4, we collect together all information that is necessary for calculation of price indices. All index number formulas are presented in multiplicative form, including Laspeyres and Paasche, which are derived from their logarithmic representations (see Vartia, 1976, p.128). Practically this means, that aggregation of price changes in (3) is done always much simpler in logarithmic form (i.e. additive form) and then transformed back as indices. As we have stressed, all formulas are evaluated separately for each component of (3) and separately for aggregate in question.

These index number formulas are used when strata decompositions are aggregated into crude aggregates like 'construction-type' -aggregates in Finland etc. Aggregation of decomposition (3) are done in logarithmic form (i.e. in additive form) keeping stratum aggregate and index number formula fixed.

Table 4: Necessary information for calculation of hedonic price indices for different formulas and aggregates (Vartia & Suoperä, 2017, 2018).

Basic formula: Contingently biased index numbers		
Symbol and name of formula	$P^{1/0}$	Weights of the formula, w_i
L Laspeyres	$\prod(p_k^1/p_k^0)^{w_k^0}$	$w_k^0=\frac{L(p_k^1 q_k^0, p_k^0 q_k^0)}{L(p^1 q^0, p^0 q^0)}$
l log- Laspeyres	$\prod(p_k^1/p_k^0)^{w_k^0}$	$w_k^0=\frac{v_k^0}{V^0}$
p log-Paasche	$\prod(p_k^1/p_k^0)^{w_k^1}$	$w_k^1=\frac{v_k^1}{V^1}$
P Paasche	$\prod(p_k^1/p_k^0)^{w_k^1}$	$w_k^1=\frac{L(p_k^1 q_k^1, p_k^0 q_k^1)}{L(p^1 q^1, p^0 q^1)}$
Excellent formulas		
T Törnqvist	$\prod(p_k^1/p_k^0)^{\bar{w}_k}$	$\bar{w}_k=0.5*(w_k^0+w_k^1)$
SV Sato-Vartia	$\prod(p_k^1/p_k^0)^{\bar{w}_k}$	$\bar{w}_k=\frac{L(w_k^1, w_k^0)}{\sum L(w_k^1, w_k^0)}$
MV Montgomery-Vartia	$\prod(p_k^1/p_k^0)^{\bar{w}_k}$	$\bar{w}_k=\frac{L(v_k^1, v_k^0)}{L(V^1, V^0)}$
F Fisher	$(L^{1/0}*P^{1/0})^{1/2}$	

6 Empirical Results

6.1 Comparation of Stratum Statistics and Their Decomposition of 'True or Actual Price Changes'

Partition of new one-family detached houses consists of 12 stratums (4 NUTS 2 regions, which are divided into three construction-type). For each stratum we calculate four types of aggregates (see Table 2), for which holds

1. unweighted arithmetic average, $uA \geq uG$, unweighted geometric average
2. weighted arithmetic average, $wA \geq wG$, weighted geometric average

As an example, the following figures shows for stratum 'Nuts 2 region Uusimaa, turnkey construction' how significantly these aggregates deviate.

The Figure 2 presents how *uA*, *uG* and *wG* deviate from weighted arithmetic average, that is D*uA*=log(*uA*/*wA*), D*uG*=log(*uG*/*wA*) and D*wG*=log(*wG*/*wA*) in log-%. Here the largest deviation is about 10 log-%, but in some other strata it may be over 10 log-%. The Figure 3 presents how much index series of weighted arithmetic average deviates from other aggregates in log-scale. Here deviations are not so large as for comparation of averages – largest deviations are about 5 log-%. Similar comparation of index series show that for some strata deviations may exceed 10 log-%.

Figure 2: Ratio of averages with respect to respect weighted arithmetic average in ‘NUTS 2 region Uu- region Uu- simaa, turnkey construction’ (deviations log-%).

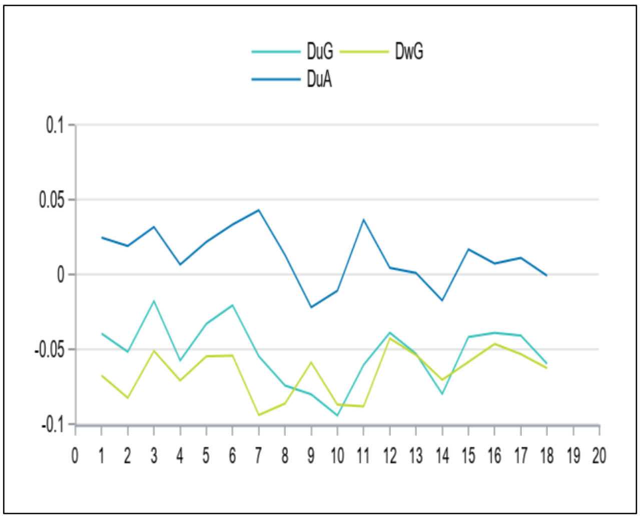
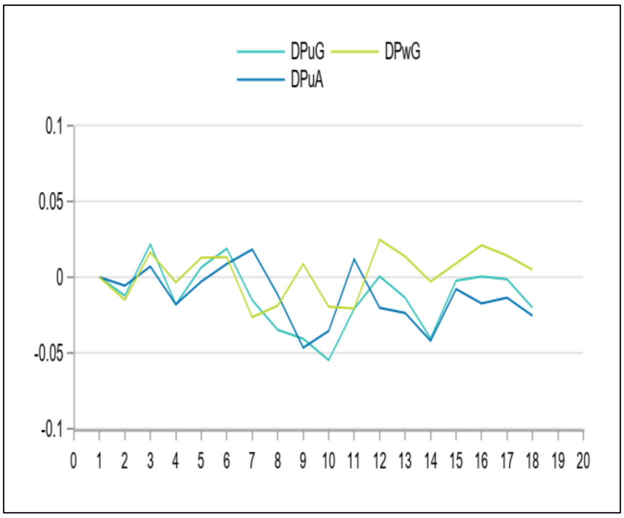


Figure 3: Ratio of index series of averages with to weighted arithmetic average in ‘NUTS 2 simaa, turnkey construction’ (deviations log-%).



For the construction of the official average price statistics difference between these four aggregates are unexpected large. Normally officially published average statistics are based on weighted arithmetic averages and we think it is the most natural choice also for new one-family detached house prices. So, our benchmark statistics is the weighted arithmetic average for which other aggregates are compared.

6.2 Decomposition of Actual Price Change for Stratum

The base strategy is free of chain error, so we choose it as the benchmark strategy for the construction of index series. In practice this means that we are interested in direct price-links, that is $0 \rightarrow t$, where base period 0 is a previous year normalized as average quarter and period t is a quarter of a current year.

Figure 4: Actual (A) and quality adjusted (QA) price corrections changes (weighted arithmetic average) in NUTS 2 region Uusimaa, ‘Turnkey construction’ 2015 = 1

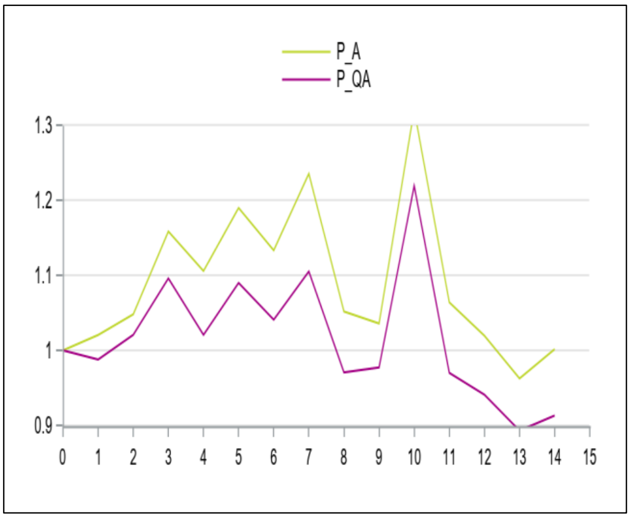


Figure 5: Corresponding indices for quality ($QC_{x_1}, QC_{x_3}, QC_x = all\ together$) in NUTS 2 region Uusimaa, ‘Turnkey construction’ 2015 = 1

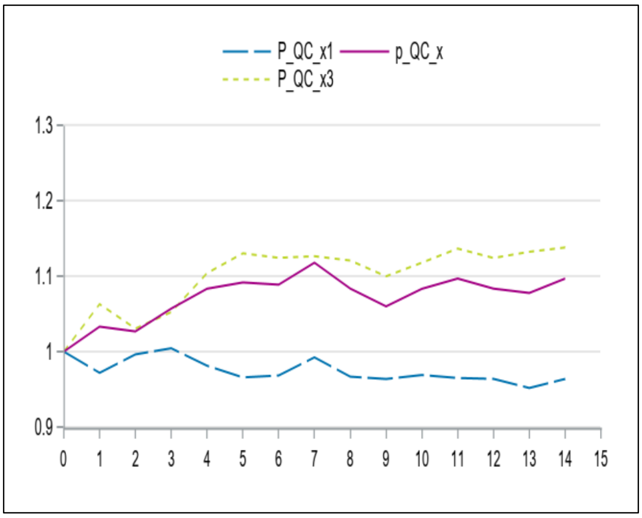


Figure 6: Actual (*A*) and quality adjusted (*QA*) price changes (weighted arithmetic average) in NUTS 2 region Uusimaa, ‘Modular house’, 2015 = 1

Figure 7: Corresponding indices for quality corrections ($QC_{x_1}, QC_{x_3}, QC_x = all\ together$) in NUTS 2 region Uusimaa, ‘Modular house’, 2015 = 1

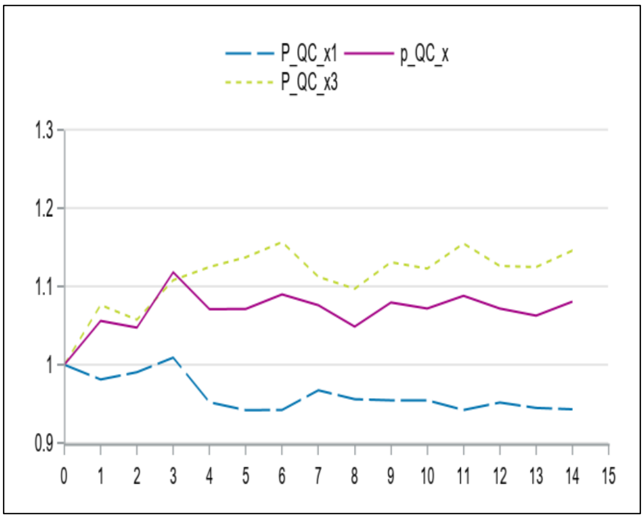


Figure 8: Actual (*A*) and quality adjusted (*QA*) price changes (weighted arithmetic average) in NUTS 2 region Uusimaa, ‘Do-it-yourself building’, 2015 = 1

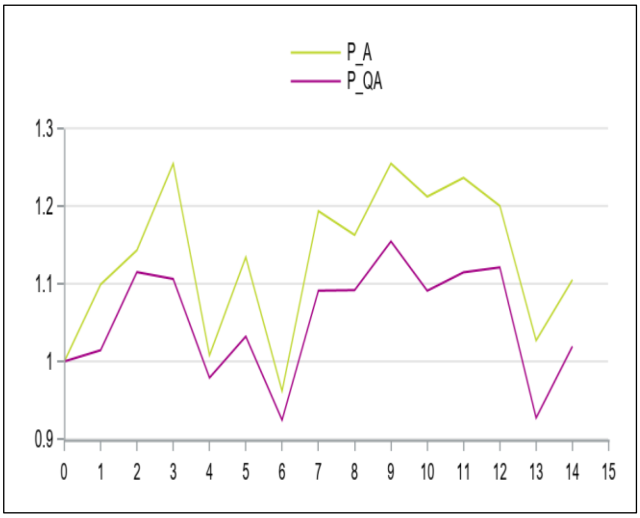
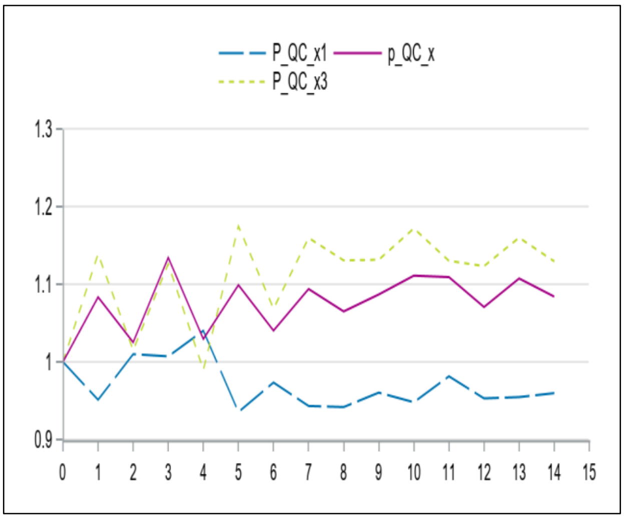


Figure 9: Corresponding indices for quality corrections ($QC_{x_1}, QC_{x_3}, QC_x = all\ together$) in NUTS 2 region Uusimaa, ‘Do-it-yourself building’, 2015 = 1



The increase of square meters have negative effect on average prices – when values of $\overline{\boldsymbol{x}}_t$ exceed (go under) their standard quality point $\overline{\boldsymbol{x}}_0$ we should correct actual prices upward (downward) in period *t* to quality adjusted level evaluated at standard quality point $\overline{\boldsymbol{x}}_0$.

Figures 4 to 9 show how strongly prices vary from quarter to quarter including sometimes quite clear seasonality. For ‘Turnkey construction’ in Uusimaa region biggest price increases happens at summer seasons - for ‘Modular house’ and ‘Do-it-yourself building’ at winter times, but not as systematically compared to ‘Turnkey construction’. This happens quite similarly in all Nuts 2 regions. The left side figures show that the quality adjusted price changes are always below the price change of actual prices. This means that quality of houses for all construction types has ‘increased’ (i.e. are smaller). Quality corrections in right side figures behave quite similarly for all construction types in Uusimaa region being significant role in determination of quality adjusted price changes. Quality corrections together (red lines in right side figures) increases quite rapidly from base year 2015 = 1 to 8 - 10 log-% for all construction types. Quality corrections are not so large for other Nuts 2 regions, but practically this means that that quality correction is necessary.

6.3 Decomposition for Price Index Numbers

Index number formulas defined in Table 4 are used when strata decompositions are aggregated into crude aggregates - categories like 'Turnkey construction in Finland', 'Modular house in Finland' etc. This will be done by following steps:

1. Take logarithm of equation (2) (additive form such that all price ratios in (3) are in logarithmic form).
2. Use index number formula (Table 4) in logarithmic form.
3. Calculate the log price change for each component of (3) separately using same index number formula.
4. Take the exp-transformation of each log price ratios.
5. Do the steps 1 to 4 separately for each average $m = uA, wA, uG, wG$.

Steps 1 – 5 gives us the hedonic price index number decomposition

(3) $$P_A^{t/0} \equiv p_{QC,x_1}^{t/0} \cdot p_{QC,x_3}^{t/0} \cdot P_{QA}^{t/0}$$

It is important to keep index number formula *P* and average *m* fixed for steps one to four.

First, we compare basic index numbers (*L* = Laspeyres and *P* = Paasche) to excellent ones (*MV* = Montgomery-Vartia, *T* = Törnqvist and *F* = Fisher). Figures 10 and 11 show that basic index number formulas are also here *data contingently biased* (see Vartia & Suoperä, 2018), but the bias is not severe. These figures show that excellent index number formulas are very closely related. As a rule, basic index number formulas should never be used, when excellent ones are available.

Figure 10: Quality adjusted (*QA*) price index series (*m* =weighted arithmetic average) for construction type type 'Do-it-yourself building' in Finland, 2015 = 1.

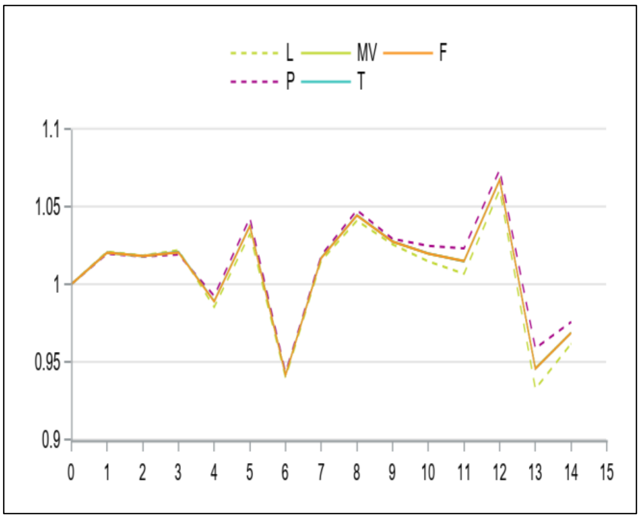
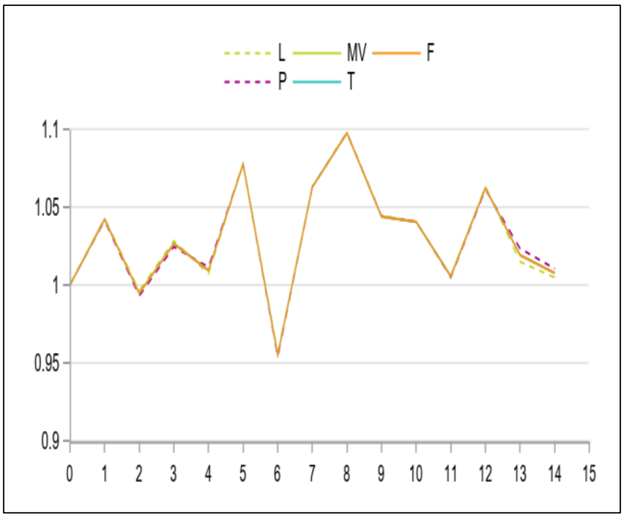


Figure 11: Quality adjusted (*QA*) price index series (*m* = weighted arithmetic average) for construction 'Modular house' in Finland, 2015=1.



Excellent index number formulas (see Vartia & Suoperä, 2018) are very closely related and any of them may be used. We demonstrate our results by Törnqvist formula. For 'Do-it-yourself' construction type in Finland the quality corrections are significant – from the year 2015 square meters decline and then stays quite stable. Index series constructed by actual price changes (yellow line) exceed index series constructed for prices being comparable in quality (red line). The quality corrections together increase sometimes up to 5 log-% being most often around 2 - 4 log-% (red lines in Figure 13, 15 and 17).

Figure 12: Actual (A) and quality adjusted (QA) price corrections changes (weighted arithmetic average) for 'Do-it-yourself' yourself' in Finland, 2015 = 1, (P = Törnqvist).

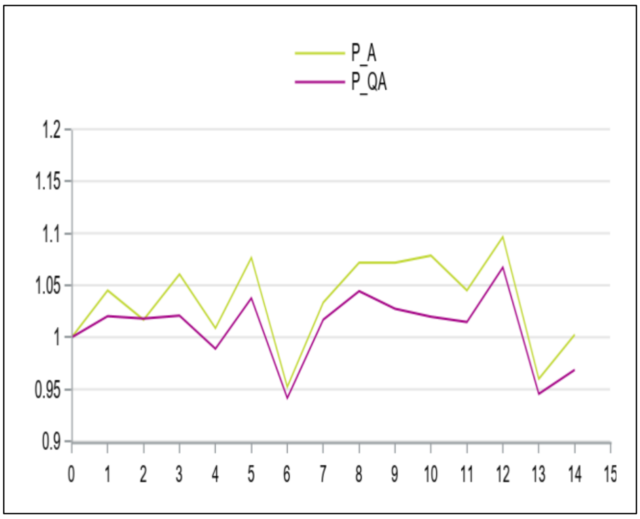


Figure 13: Corresponding indices for quality ($QC_{x_1}, QC_{x_3}, QC_x = all\ together$) for 'Do-it-yourself' in Finland, 2015 = 1, (P = Törnqvist).

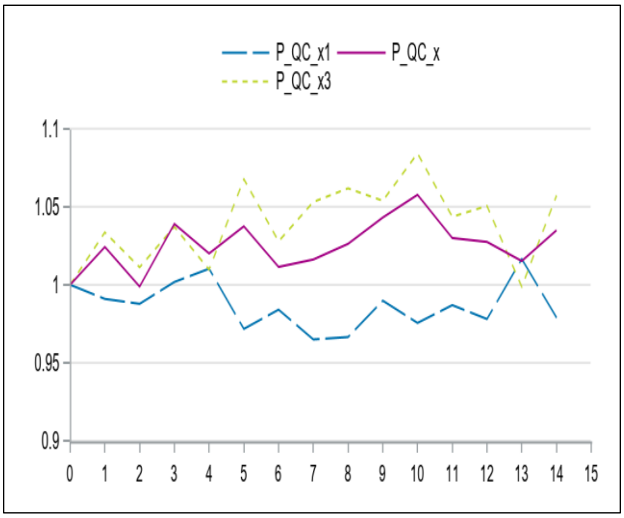


Figure 14: Actual (A) and quality adjusted (QA) price corrections changes (weighted arithmetic average) for 'Modular house' house' in Finland, 2015 = 1, (P = Törnqvist).

Figure 15: Corresponding indices for quality ($QC_{x_1}, QC_{x_3}, QC_x = all\ together$) for 'Modular house' in Finland, 2015 = 1, (P = Törnqvist).

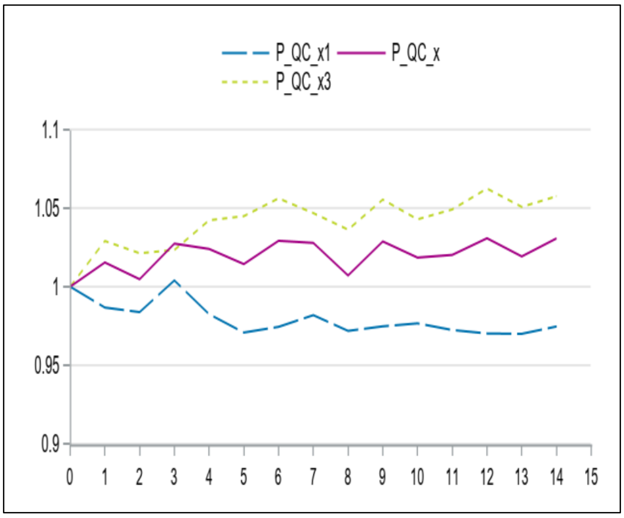
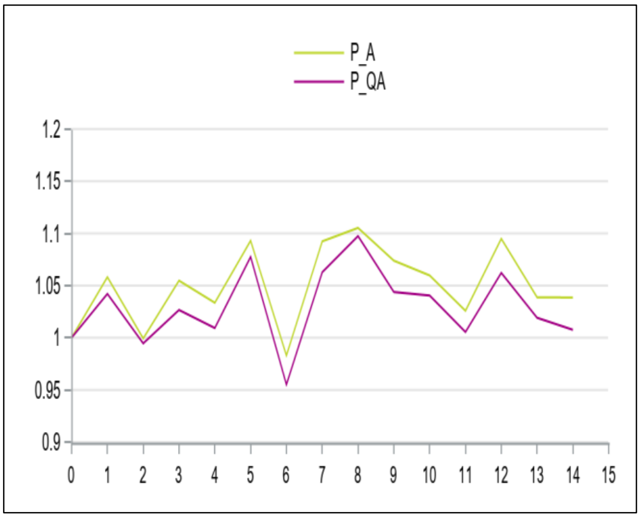


Figure 16: Actual (A) and quality adjusted (QA) price corrections
changes (weighted arithmetic average) for 'Turnkey construction' in Finland, 2015 = 1, (P = Törnqvist).

Figure 17: Corresponding indices for quality ($QC_{x_1}, QC_{x_3}, QC_x = all\ together$) for 'Turnkey construction' in Finland, 2015 = 1, (P = Törnqvist).



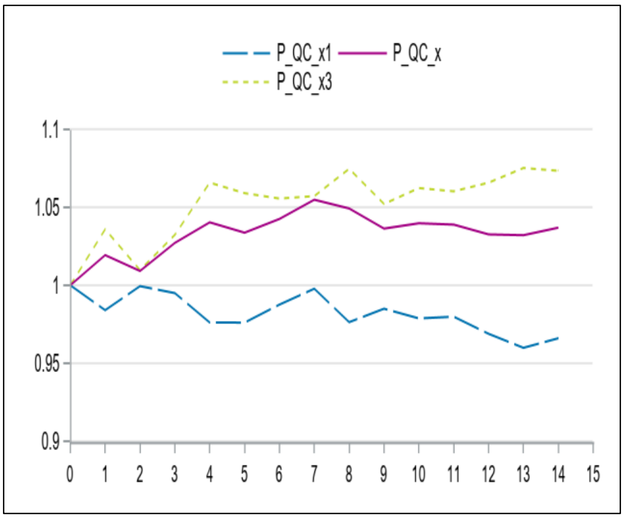
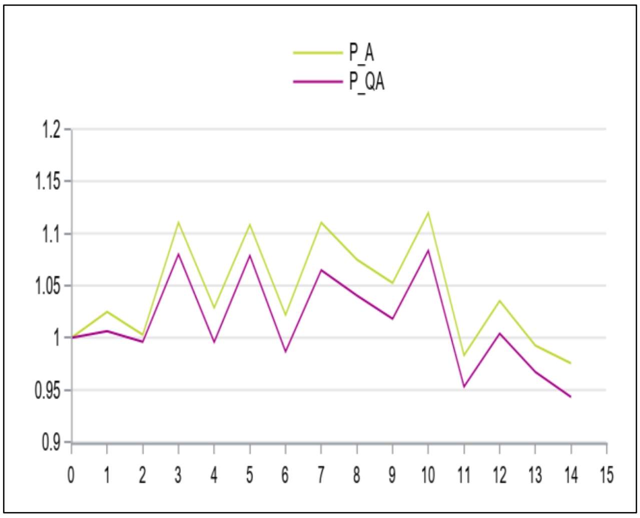
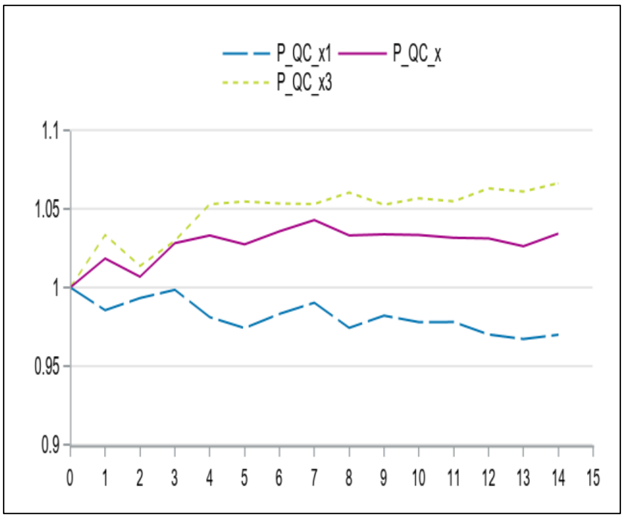


Figure 18: Actual (A) and quality adjusted (QA) price corrections
changes (weighted arithmetic average) in Finland, 1 2015 = 1, (P = Törnqvist).

Figure 19: Corresponding indices for quality ($QC_{x_1}, QC_{x_3}, QC_x = all\ together$) in Finland, 2015 = (P = Törnqvist).





7 Conclusion

Because new one-family detached houses emerges only once in data, index number theory based on bilateral or multilateral methods cannot be used (i.e. no matched pairs are available). Observations should be analyzed within an appropriate partition, which in this study is based on four Nuts2 regions each divided into three construction -type – 'Do-it- yourself', 'Modular house' and 'Turnkey construction'. Our partition includes 12 strata. We apply regression analysis to our data with this given partition. Our regression analysis is based on heterogeneously behaving cross sections. We average our price models into strata level for each time period separately and define difference simply saying between periods 0 and t. Whatever average of prices we use, they are not comparable in quality for price-link $0 \rightarrow t$ – hence the quality differences should be removed from actual or true price changes to get quality adjusted price change evaluated in standardized point of quality. For that we use so called hedonic or Oaxaca-decomposition.

Normally the decomposition is based on a standard textbook solution, where the price ratio of unweighted geometric average prices is decomposed into price change due to quality difference and quality adjusted price change evaluated at standard point of quality. This solution is not satisfactory for officially published average statistics, because unweighted geometric average prices deviate from weighted arithmetic average prices about 5 – 10 log-% and sometimes more than 10 log-%. So, we develop three new theorem of price aggregation for semilogarithmic price models. We apply these theorems for estimated OLS solution and get three new hedonic decompositions based on weighted and unweighted arithmetic average and weighted geometric average prices. We derive the hedonic price decomposition for all four averages one of which is a standard textbook solution for unweighted geometric average prices.

In strata level no index number formula is needed – the hedonic price decomposition divides the average price change into quality corrections and quality adjusted price change for all averages separately. We show that quality adjustment is necessary. Instead of unweighted geometric average price decomposition, our benchmark decomposition is based on weighted arithmetic average prices – a new hedonic index solution.

We use index numbers for the first time, when strata level decompositions are aggregated into crude aggregation level. We use basic and excellent index number formulas. Our construction strategy of index series is base strategy, where the base period is the previous year normalized as average quarter. Index series constructed by this strategy are free of chain error. As conclusion, basic index number formulas are data contingently biased and should not be used as official statistic. Instead, excellent index number formulas are very closely related – any of them may be selected to official production. Similarly, as in ‘Hedonic Price Index Number for New Blocks of Flats and Terraced Houses in Finland’ (Vartia, Suoperä & Vuorio, 2019) we select Törnqvist formula.

References

Hsiao, C. ‘Analysis of Panel Data’., Cambridge University Press, 1986.

Nieminen, K. and Montonen, S, “The foundation of index calculation”, (http://www.stat.fi/meta/menetelmakehitystyo/index_en.html), 2018.

Koev, E. ‘Combining Classification and Hedonic Quality Adjustment in Constructing a House Price Index’, Licentiate thesis, Helsinki, 2003.

Koev, E. & Suoperä A. ’Pientalokiinteistöjen (omakotitalojen ja rakentamattomien pientalotonttien) hintaindeksit 1985=100’, Helsinki, 2002. (in Finnish, Statistics Finland).

Oaxaca, R. ‘Male-Female Wage Differentials in Urban Labour Markets’, International Economic Review, 14, pp. 693 - 709, 1973.

Suoperä, A. ’Kirkon ja muiden sektoreiden suhteelliset palkkaerot: Niiden analyysi ja tilastointi’, 2002. (study subject to a charge, Church labour markets/Statistics Finland).

Suoperä, A. ’Some new perspectives on price aggregation and hedonic index methods: Empirical application to rents of office and shop premises’, 2006 (unpublished, Statistics Finland).

Suoperä, A. ’Naisten ja miesten palkkaerot valtiosektorilla ja yleisillä työmarkkinoilla’, 2010a (study subject to a charge, Ministry of Finance/Statistics Finland).

Suoperä, A. ‘Kunnan ja yleisten työmarkkinoiden palkkaerot: Hedoninen menetelmä’, 2010b (study subject to a charge, Municipal labour markets/Statistics Finland).

Suoperä, A. & Vartia, Y. ‘Analysis and Synthesis of Wage Determination in Heterogeneous Cross-sections’, Discussion Paper No. 331, 2011.

Vartia, Y. ‘Relative Changes and Index Numbers’, Ser. A4, Helsinki, Research Institute of Finnish Economy, 1976.

Vartia, Y. ‘Ideal Log-Change Index Numbers’, Scandinavian Journal of Statistics., 3, pp. 121 . 126,1976.

Vartia, Y. ’Kvadraattisten mikroyhtälöiden aggregoinnista’, ETLA, Discussion topics no. 25,1979.

Vartia, Y. ’Good choices in index number production’, 2018 (http://www.stat.fi/static/media/uploads/meta_en/menetelmakehitystyo/over_identification_new_vartia2018.pdf).

Vartia, Y. & Suoperä, A. “Index number theory and construction of CPI for complete micro data”, 2017. (http://www.stat.fi/meta/menetelmakehitystyo/index_en.html).

Vartia, Y. & Suoperä, A. “Contingently biased, permanently biased and excellent index numbers for complete micro data”, 2018. (http://www.stat.fi/static/media/uploads/meta_en/menetelmakehitystyo/contingently_biased_vartia_suopera_updated.pdf)

Vartia, Y., Suoperä, A., Nieminen, K. & Montonen, S. “Circular Error in Price Index Numbers Based on Scanner Data. Preliminary Interpretations”, 2018a. (http://www.stat.fi/meta/menetelmakehitystyo/index_en.html)

Vartia, Y., Suoperä, A., Nieminen, K. & Montonen, S. "The Algebra of GEKS and Its Chain Error", 2018b. (http://www.stat.fi/meta/menetelmakehitystyo/index_en.html).

Vartia, Y., Suoperä, A., & Vuorio, J. 'Hedonic Price Index Number for New Blocks of Flats and Terraced Houses in Finland' (2019).

Vartia, Y. and Vartia, P. 'Descriptive Index Number Theory and the Bank of Finland Currency Index', Scandinavian Journal of Economics, vol. 3, pp. 352 . 364, 1985.

Törnqvist, L. 'A Memorandum Concerning the Calculation of Bank of Finland Consumption Price Index', unpublished memo, Bank of Finland, 1935.

Törnqvist, L. 'Levnadskostnadsindexerna i Finland och Sverige, Deras Tillförlitlighet och Jämförbarhet', Ekonomiska Samfundets Tidskrift, vol. 37, 1-35, 1936. (in Swedish)

Törnqvist, L. & Vartia, P. & Vartia, Y. 'How Should Relative Changes be Measured'? The American Statistician, Vol. 39, No. 1. pp. 43 - 46, 1985.