Some new perspectives on price aggregation and hedonic index methods:

Empirical application to rents of office and shop premises

Antti Suoperä

Statistics Finland

Prices and Wages

2006

1 INTRODUCTION

In simplified terms, two alternative methods can be used in index calculations for controlling change in quality. One is to analyse realised changes in the prices of qualitatively comparable observations, or the so-called matched pairs method (Bailey, Muth and Nourse, 1963; Case and Shiller, 1989; Quigley, 1995). The construction strategy of a hedonic index shows that in index calculations the method becomes reduced into a classic classification index (see Chapter 4, p. 8). The second method is not limited to the analysis of identical statistical units, but calculates price changes from representative cross-sectional samples of the population. The method combines relevant stratification, i.e. classification, of the studied topic on the one hand, and regression analysis of heterogeneous cross-sectional data, on the other. The index application of the method is based on the Oaxaca decomposition (Oaxaca, 1973), which breaks change in geometric mean prices down into quality adjustment factors and price change standardised for quality.

In this study we deduce a similar decomposition of relative change in arithmetic mean prices for semi-logarithmic price models. We develop two new aggregation solutions in which logarithmic prices at the observation level are aggregated to stratum level, i.e. micro class level, so that the (weighted or unweighted) arithmetic mean is received as the argument of logarithmic function. In the analysis we exploit the logarithmic mean (L. Törnqist, 1935, p. 35; L. Törnqist, P. Vartia and Y. Vartia, 1985, p. 44). The micro indices of the study are methodologically analogous with the decomposition presented by Koev but instead of geometric means the Oaxaca decomposition of the study in based on change in arithmetic means. The index application of the study is developed as a logarithm of Laspeyres' index formula (Y. Vartia, p. 128, 1976). In the empirical part of the study the analysis is applied to the statistics of KTI Property Information Ltd on the rents of office and shop premises. The study is a continuation of the hedonic quality standardising methods that are widely applied at Statistics Finland to index applications of dwelling prices.

The structure of the study is as follows: Chapter 2 describes the price models, stratification of the data and estimation of the price models. Chapter 3 presents the aggregation of observation level logarithmic prices to the micro classification level. It describes the inference of micro level price aggregates and the qualitative characteristics that control them by means of the logarithmic mean. In Chapter 4, the micro level indices are aggregated to less detailed classification levels so that the consistency of the Oaxaca decomposition of the classification index is retained. Chapter 6 presents the results of the study, followed by Chapter 7 containing conclusions.

2 ANALYSIS OF HETEROGENEOUS CROSS-SECTIONAL DATA

Stratification of observation data and statistical inferrence of price models form the theoretical basis of the study. The analysis combines classification and typical regression analysis. Because the study combines analysis of variance and typical regression analysis, the method is based on analysis of covariance. A similar statistical deduction method has also been applied by e.g. Vartia, Y. & Kurjenoja, J. (1992), Koev, E. (1996, 1997, 2003), Vartiainen, 2001, Koev, E. & Suoperä, A. (2002) Kouvonen, S. & Suoperä, A. (2000, 2002), Suoperä, A. (2002, 2003), Korkeamäki, O., Kyyrä, T. & Luukkonen, A. (2004).

2.1 Classification of observations

Let us examine the population of statistical units $A=\{a_1,a_2,a_3,...,a_n\}$ and its stratification into micro classes A_k, where sub-index $k=1,...,K$ represents the strata, or micro classes. Subpopulations A_k of the statistical units are separate and exclude each other, so $A_k \cap A_{k'} = \varnothing$,

$\forall k \neq k'$ and $A = \bigcup_{k=1}^{K} A_k$ holds true for their compound. Indicator $1(a_i; A_k)$ is specified for each micro class so that

$$1(a_i; A_k) = T(a \in A_k) = \begin{cases} 1, \text{if } a_i \in A_k \\ 0, \text{if } a_i \in \mathrm{A}_\mathrm{k}^\mathrm{c} \end{cases}$$

which unambiguously classifies all statistical units into strata, e.g. micro classes by region. Combining of the indicator variables forms the design matrix of the variance analysis model.

2.2 Specification of the price model

We shall examine the determination of prices for one stratum. We shall investigate observation *i* of stratum *k* in time period *t*. The logarithmic unit price $\log(p_{ikt})$ of this observation is modelled as follows:

(2.1) $$\log(p_{ikt}) = \alpha_{kt} + x'_{ikt}\, \beta_{kt} + \varepsilon_{ikt}$$

where α_{kt} represents the price effectof stratum A_k in time period *t*. Parameters α_{kt} can be specified so that $\alpha_{kt} = \mu_t + \mu_{kt}$, where μ_t represents average price effectand μ_{kt}, respectively, the deviation of the *k*th stratum from the common mean μ_t. It then holds true for the price effects μ_{kt} of the strata that $\sum_k w_{kt}\mu_{kt} = 0$, where $\sum_k w_{kt} = 1$ and represents the relative frequencies of the strata. Vector x'_{ikt} contains qualitative background variables at observation level, and parameter vector β_{kt} reaction parameters of the respective qualitative background variables in time period *t*. Variable ε_{ikt} of the model is a so-called random error term, with assumed expected value of zero and finite variance. The price model is specified as linear relative to the parameters, thus belonging to the 'family' of flexible function forms. The specification of the price model is not only flexible relative to its function form but also relative to its parameters - all its unknown parameters can vary in time by stratum.

The unknown parameters are estimated in two stages. Parameter vectors β_{kt} are estimated in the first stage by centring the statistical data relative to stratum A_k. With vector and matrix notation we obtain OLS estimator (C. Hsiao, 1986, pp. 29-32)

(2.2) $$\hat{\beta}_{kt} = (X'_{kt} M_{kt} X_{kt})^{-1} X'_{kt} M_{kt} \log(p_{kt}),$$

where M_{kt} is an idempotent, symmetrical transformation matrix which transformas individual observations as deviations from the corresponding stratum means. The price effects of the specific stratum A_k are estimated in the second stage as follows:

(2.3) $$\hat{\alpha}_{kt} = log(\bar{p}^G_{kt}) - \bar{x}'_{kt}\, \hat{\beta}_{kt}$$

where $\log(\bar{p}^G_{kt})$ represents the logarithm of the (unweighted) geometric mean price of stratum A_k, vector $\bar{x}'_{kt}$ the (unweighted) arithmetic means of the quality variables and parameter vector $\hat{\beta}_{kt}$ the OLS estimates of the same stratum. For an example of a similar method for estimating vector coefficients for heterogeneously behaving cross-sectional data, see Koev, E. & Suoperä, A. (2002), Koev, E. (2003).

3 AGGREGATION OF OBSERVATION PRICES IN STRATA

The aggregation of all variables of model (2.1) is guided by the three following mathematical characteristics: First, we demand that residuals for any individual stratum sum up to zero. Second, the hypersurface of the regression model must always go through the means of input and output variables. Third, the mean of the fitted values, i.e. predictions of the output variable, of the regression model must match precisely the arithmetic mean of the dependent variable. These three characteristics reveal that the dependent output variable is decomposed into two orthogonal components of which the first is expressed as a linear combination of exogenous input variables and the second as an error term (residual) which is orthogonal with the exogenous input variables of the model. These conditions are trivially met when we choose the same weight for all observations. Then the semi-logarithmic specification of model (2.1) leads to an unweighted geometric mean at the stratum level in the aggregation of the output variable, and input variables, in turn, become unweighted arithmetic means. This will naturally hold true if the β heterogeneity of model (2.1) is excluded by assuming that (2.1) represents homogeneous behaviour (se diverse variations in e.g. Oaxaca (1973), Mincer (1974), Willis (1986), Card (1999), Vartiainen (2001), Bayard, Hellerstein and Troske (2003), Korkeamäki and Kyyrä (2002), Korkeamäki and Kyyrä (2003), Korkeamäki, Kyyrä and Luukkonen, (2004)). The method is based on the typical aggregation of 'unit prices' in which all observations have the same weight and therefore contribute equally to stratum averages irrespective of their real quantitative categories. In the index application of the method, relative change of geometric means is measured at the stratum level, as best example of which see Koev (2003).

In this study we develop two new aggregation solutions for the observations of model specification (2.1), the first one of which leads to a logarithmic unweighted and the second to a logarithmic weighted arithmetic mean at the stratum level. Annex 5 proves that both in the case of strata and their arbitrary union they satisfy the above three basic Gaussian criteria. The new aggregation rules are deduced by means of the logarithmic mean (L. Törnqvist, 1935, p. 35; Y. Vartia, 1976; L. Törnqvist, P. Vartia and Y. Vartia, 1985, p. 44).

We will first examine what with out current knowledge can be said about differences between the geometric and arithmetic means. By means of Taylor series we obtain as the difference between the geometric and arithmetic means at the stratum level (L. Törnqvist, 1936; Y. Vartia and P. Vartia, 1984)

$$(3.1) \qquad 1/n\sum_{i}\log(p_{ikt}) = \log(\overline{p}_{kt}^{G}) \approx \log(\overline{p}_{kt}^{PC}) - \tfrac{1}{2}s_{kt}^{2},$$

where $\log\left(\overline{p}_{kt}^{G}\right)$ is a logarithmic transformation of the (unweighted) geometric mean and $\log\left(\overline{p}_{kt}^{PC}\right)$ respectively a logarithmic transformation of the (unweighted) arithmetic mean for stratum A_k. Term s_{kt}^{2} is the variance of the logarithmic variable in stratum A_k in time period *t*, so $\log\left(\overline{p}_{kt}^{G}\right) \le \log\left(\overline{p}_{kt}^{PC}\right)$ always holds true. The second-order Taylor approximation is accurate only if the logarithmic variable is normally distributed (see Van Dalen and Bode, 2004, p.2-3). This is mainly contemplation of theory, so nothing can be said for certain about the difference in the relative change between the arithmetic and geometric means, that is, remainder $\nabla\log(\overline{p}_{kt}^{G}) - \nabla\log(\overline{p}_{kt}^{PC})$ can be positive, negative or zero between the differences of the statistics. In practice, because the second-order Taylor approximation is 'adequately' accurate for regularly behaving variables, it follows from the stability of variances s_{kt}^{2}, $t = 0,1,\ldots$ that $\nabla\log(\overline{p}_{kt}^{G}) - \nabla\log(\overline{p}_{kt}^{PC}) \approx 0$. In small strata the variances can be varied significantly and the above approximation is not valid.

When we exclude the normality of a variable (that is, $log(Y)$), all we know about differences between the geometric and arithmetic means is approximation. The situation cannot be changed by endless development of Taylor expansions - we will still lack exact empirical measures about the difference between arithmetic and geometric means. We will now go on to rectify this shortcoming. In our study we develop for logarithmic observations two new aggregation rules with which differences between the arithmetic and geometric means can be exactly modelled mathematically. The method can also be applied precisely to the ratios of these statistics and to their remainder. The analysis exploits the expression of logarithmic mean in the aggregation of logarithmic observations, that is, e.g. the left side of equation (2.1). As a result we receive logarithmic stratum aggregates in which the arguments for the logarithmic function are arithmetic - as opposed to the conventional geometric - means. In fact, we develope accurate mathematical sollution to the problen of 'estimation biases in quality-adjusted hedonic price indices' (more spesificly see Annex 5) represented nicely in J. van Dalen and B. Bode (2004). In this solution the order of workings are important - first we estimate log-transformed multiplicative hedonic regressions and after that we agregate known micro level behaviour in to, say, $k=1,\ldots,K$ strata levels.

Logarithmic mean is determined for two positive figures x and y as follows (L. Törnqvist, 1935, p. 35; Y. Vartia, 1976; L. Törnqvist, P. Vartia and Y. Vartia, 1985, p. 44).

$$\text{(3.2)} \qquad L(x,y) = (y\text{-}x)/\log(y/x), \quad \text{if } x \neq y$$
$$= x, \qquad \text{if } x = y$$

The definition can also be expresses as $\log(y/x) = (y\text{-}x)/L(x,y)$, when it connotes that the log change from x to y is a relative change compared to the logarithmic mean. This indicator of relative change is a ratio that is symmetrical, additive and independent of measurement unit. Let us examine positive figures $\{y_{1kt},\ldots,y_{nkt}\}$ $\{x_{1kt},\ldots,x_{nkt}\}$ in stratum A_k in time period t and determine their logarithmic mean as follows

$$\text{(3.3)} \qquad L\left(\sum_i y_{ikt}, \sum_i x_{ikt}\right) = \frac{\sum_i y_{ikt} - \sum_i x_{ikt}}{\log\left(\sum_i y_{ikt} / \sum_i x_{ikt}\right)},$$

or congruently

$$\log\left(\sum_i y_{ikt} / \sum_i x_{ikt}\right) = \sum_i \frac{y_{ikt} - x_{ikt}}{L\left(\sum_i y_{ikt}, \sum_i x_{ikt}\right)},$$

whereby we receive (see definition of logarithmic mean)

$$\text{(3.4)} \qquad \log\left(\sum_i y_{ikt} / \sum_i x_{ikt}\right) = \sum_i \frac{L(y_{ikt}, x_{ikt})}{L\left(\sum_i y_{ikt}, \sum_i x_{ikt}\right)} \log(y_{ikt}/x_{ikt}),$$

The operational characteristics of expression (3.4) are so far difficult to observe. We will concretise the expression with price aggregation by examining the two arbitrarily divided variables x and y. We will first determine: $y_{ikt} = v_{ikt} = q_{ikt}p_{ikt}$ and $x_{ikt} = q_{ikt}$, where v_{ikt} connotes the value of observation i in stratum A_k in time period t and q_{ikt}, respectively, its quantities. By locating the variables in equation (3.4) we obtain

$$\text{(3.5a)} \qquad \log\left(\sum_i q_{ikt}p_{ikt} / \sum_i q_{ikt}\right) = \log(\overline{p}_{kt}^{Aw}) = \sum_i \frac{L(q_{ikt}p_{ikt}, q_{ikt})}{L\left(\sum_i q_{ikt}p_{ikt}, \sum_i q_{ikt}\right)} \log(p_{ikt})$$

where argument $\overline{p}_{kt}^{Aw}$ of the logarithmic function connotes the weighted arithmetic mean price in stratum A_k in time period *t*. The second price aggregation principle is developed by selecting $x_{ikt} = q_{ikt} = 1$ and $y_{ikt} = q_{ikt} p_{ikt} = 1 \cdot p_{ikt} = p_{ikt}$ when equation (3.4) can be expressed as

$$(3.5b) \qquad \log\left(\sum_i p_{ikt} \Big/ \sum_i q_{ikt}\right) = \log\left(n_{kt} \cdot \overline{p}_{kt}^{A} / n_{kt}\right) = \log\left(\overline{p}_{kt}^{A}\right) = \sum_i \frac{L(p_{ikt}, 1)}{L\left(\sum_i p_{ikt}, n_{kt}\right)} \log(p_{ikt})$$

where n_{kt} is the number of observations and $\overline{p}_{kt}^{A}$ the unweighted arithmetic mean price, i.e. the mean of unit prices (that is, p_{ikt}) in stratum A_k in time period *t*. In the mean of unit prices, the weight of each price at the observation level is $1/n_{kt}$ (see left side of equation (3.5b)), whereas in the mean $\overline{p}_{kt}^{A}$ of equation (3.5a) the quantities at the observation level contribute to the weighted arithmetic mean price, as they also should do. The difference between the arithmetic and geometric means obtained with term (3.5a) is

$$(3.6) \qquad \log(\overline{p}_{kt}^{Aw}) - \log(\overline{p}_{kt}^{G}) \equiv \sum_i \left(w_{ikt}^{Aw} - n_{\mathrm{kt}}^{-1}\right) \log(p_{ikt}),$$

where weights w_{ikt}^{Aw} have been defined with equation (3.5a) (or alternatively with (3.5b)), $\log(\overline{p}_{kt}^{Aw})$ is calculated with (3.5a) (or alternatively with (3.5b)) and $\log(\overline{p}_{kt}^{G}) = \sum_i n_{kt}^{-1} \log(p_{ikt})$. In deviation from the Taylor series expansion, the above difference between the geometric and arithmetic means is identically valid for all positive figures q_{ikt} and p_{ikt}. In fact, the clause holds true for the arbitrarily divided *p* and *q* variables for which a logarithmic transformation has been defined. In addition, if $\sum_i w_{ikt}^{Aw} = 1$, then the difference between the arithmetic and geometric means is $\sum_i \left(w_{ikt}^{Aw} - n_{kt}^{-1}\right) \log(p_{ikt}) \equiv n_{kt} \mathrm{cov}(w_{kt}^{Aw}, \log(p_{kt}))$ [1]. The difference between weighted arithmetic and geometric means is calculated with (3.6) by replacing n_{kt}^{-1} with volume weights $q_{ikt} \Big/ \sum_i q_{ikt}$.

The obtained difference between the arithmetic (here weighted arithmetic) and geometric means is: $\nabla\log(\overline{p}_{kt}^{G}) - \nabla\log(\overline{p}_{kt}^{Aw}) \equiv \nabla\left\{\sum_i \left(w_{ikt}^{Aw} - n_{\mathrm{kt}}^{-1}\right)\log(p_{ikt})\right\}$. See Annex 5. Differences in the micro indices depend on the samples of each given point in time - the differences can grow, remain unchanged or diminish over time.

Table 3.1 examines the accuracy of second-order Taylor approximation in the estimation of difference between unweighted arithmetic and geometric means with data of the KT Institute on rents of office and shop premises. The data are from the time period 1995/2 - 2005/1 and their stratification in respect of office and shop premises has been described in Annexes 1 and 2. The real difference between the arithmetic and geometric means is calculated for each stratum at each point in time with equation (3.6), (weights w_{ikt} have been defined in equations (3.5b)) and second order Taylor approximation, respectively, with formula (3.1). The table presents deviation of the Taylor approximation relative to the real difference (3.6). The distribution of errors is described at 5, 25, 75 and 95 percentage points of the distribution, and with the mean and the median.

---

[1] $cov(x, y) = E\{(x - E(x))\,(y - E(y))\} = E\{(x - E(x))\, y\} = E\{(y - E(y))\, x\}$ (Y. Vartia, 1979).

Table 3.1: Relative errors in second order Taylor approximation of difference between the arithmetic and geometric means (eq. (3.6)) at 5, 25, 75, 95 percentage points, means and medians for rents of office and shop premises for time period 1995/2-1995/1 (log %)

	Strata	$p_{0.05}$	$p_{0.25}$	Mean	Median	$p_{0.75}$	$p_{0.95}$
Shops	1,302	-16.46	-7.92	-3.30	-3.33	1.46	9.59
Offices	1,174	-6.78	-0.80	3.17	3.05	7.39	13.30

Second order Taylor approximation calculates the difference between the arithmetic and geometric means with a systematic bias - in rents of shop premises it underestimates and in the case of office premises it overestimates the real difference between the means. The usability of the method is weakened by the fact that at the stratum level the approximation errors can be randomly of different signs at different points in time.

In modern microeconomic theory, demand and cost functions belong to the same function family as specification (2.1). Above-described 'price aggregation' problems do not arise in micro-aggregation-macro analyses of economic theory. Price aggregation is trivially precise because in an optimal economy unit prices of individual commodities or production inputs do not vary over micro agents. Thus, $\log(\overline{p}_{kt}^{Aw}) \equiv \log(\overline{p}_{kt}^{G})$ always applies trivially (see e.g. Muellbauer, 1975, 1976; Christensen, Jorgenson & Lau, 1971; Sargan, 1971; Deaton & Muelbauer, 1980; Diewert, 1982). Because the number of commodities and production inputs is extremely large, their stratification and aggregation to the stratum level are often necessary in an empirical analysis. For instance, in modern demand theory the situation could be describes as follows: Let us examine commodity stratum A_k - e.g. group 'food' in the SNA commodity classification. The unit prices (of e.g. different kinds of fruit) vary within the group. Then the aggregation of logarithmic unit prices to the group level - if performed with observation weights other than those defined in equation (3.5a) or (3.5b) - inevitably leads to the geometric mean price. By replacing the geometric mean price with the arithmetic mean price, in the same way as Deaon and Muellbauer (1980, p. 318), the aggregation of unit prices generates an approximation error of precisely the same magnitude as equation (3.6), the impacts of which on estimation results are not known.

This Chapter examined price aggregation as a separate research problem. Annex 5 presents the mathematics of aggregation clause (3.5a) (or analogically (3.5b)) by applying it to individual stratum A_k in the case of estimated model specification (2.1).

4 MICRO INDICES, STANDARDISATION FOR QUALITY AND INDEX DECOMPOSITION

This Chapter examines the analysis of heterogeneous cross-sectional data in index calculation. The statistical deduction of price models (2.1) is performed in basic text book manner and the actual interest focuses on the aggregation of estimated models, their differentiation and decomposition. In the aggregation of output and input variables by weights derived in equation (3.5a) (or (3.5b)) leads to re-parametrization of the model at the stratum level (See Annex 5). After this, the 'elementary aggregate' models of the strata are differentiated and decomposed with the Oaxaca decomposition. In the decomposition, the change in mean prices is divided on the one hand into quality adjustments for qualitative variables controlled in the index calculation and into price change adjusted for quality. At the stratum level the method is analogous with Koev's (2003) analysis but is performed in the study with logarithmic arithmetic means. The actual index calculation is performed to the components of the decomposition as logarithmic presentation of Laspeyres' index formula (Y. Vartia, 1976, p. 128).

4.1 Division of the classification index into its elements

The analysis of the Chapter follows the mathematics of Annex 5, in which the estimation of price models and it's aggregation from the observation level are kept separate. For instance, equation (2.1), aggregated with equation (3.5a) weights from observation level to stratum level receives expression $\log\left(\overline{p}_{kt}^{Aw}\right)=\hat{\alpha}_{kt}^{Aw}+\overline{x}_{kt}'^{Aw}\hat{\beta}_{kt}$ (see Annex 5, eq. (8)) in which the difference at the base and comparison points of time (τ, t) is determined for stratum A_k as follows

$$\log\left(\frac{\overline{p}_{kt}^{Aw}}{\overline{p}_{k\tau}^{Aw}}\right)=\hat{\alpha}_{kt}^{Aw}+\overline{x}_{kt}'^{Aw}\hat{\beta}_{kt}-\hat{\alpha}_{k\tau}^{Aw}-\overline{x}_{k\tau}'^{Aw}\hat{\beta}_{k\tau}$$

The variables and parameters of the model are presented in Annex 5. Aggregation clause (3.5a) developed in Chapter 3 is applied to the input and output variables of the price model in the aggregation of observations. In the aggregation of observations semi-logarithmic models are re-parametrisized so that the arithmetic mean is received as the argument of the logarithmic function. As the unfortunate mathematics of Annex 5 show, the 'average models' of the strata are precise because they do not contain approximative residual terms. The stratum level equation presents the log change of arithmetic means between the base and comparison periods. By defining for the equation first an Oaxaca decomposition (1973) and then an exponent change, micro index for stratum k can be expressed as

$$(4.1)\qquad \frac{\overline{p}_{kt}^{Aw}}{\overline{p}_{k\tau}^{Aw}}=\exp\left\{\left(\overline{x}'^{Aw}_{kt}-\overline{x}'^{Aw}_{k\tau}\right)\hat{\beta}_{k\tau}\right\}\exp\left\{\hat{\alpha}_{kt}^{Aw}-\hat{\alpha}_{k\tau}^{Aw}+\overline{x}'^{Aw}_{kt}\left(\hat{\beta}_{kt}-\hat{\beta}_{k\tau}\right)\right\}$$

which is a well-known method in, for example, dwelling price indices and studies concerning pay discrimination (Oaxaca, 1973; Vartiainen, 2001; Koev, 2003; O. Korkeamäki, T. Kyyrä, & A. Luukkonen, 2004). However, decomposition (4.1) deviates from studies on pay discrimination in two respects: 1) As in Koev's (2003) index application, decomposition (4.1) is developed for heterogeneously behaving cross-sectional data and 2) Decomposition (4.1) is based on arithmetic, and not on geometric, means like other decompositions.

In simplified terms, the price index of micro class A_k is explained by two factors: The first term, $\exp\left\{\left(\overline{x}'^{Aw}_{kt}-\overline{x}'^{Aw}_{k\tau}\right)\hat{\beta}_{k\tau}\right\}$, reveals which part of the change in arithmetic mean prices is explained by changes in qualitative factors between the base and comparison periods. The second term, presented as $\exp\left\{\hat{\alpha}_{kt}^{Aw}-\hat{\alpha}_{k\tau}^{Aw}+\overline{x}_{kt}'^{Aw}\left(\hat{\beta}_{kt}-\hat{\beta}_{k\tau}\right)\right\}$ = $\exp\left\{\left(\hat{\alpha}_{kt}^{Aw}+\overline{x}_{kt}'^{Aw}\hat{\beta}_{kt}\right)-\left(\hat{\alpha}_{k\tau}^{Aw}+\overline{x}_{kt}'^{Aw}\hat{\beta}_{k\tau}\right)\right\}$ = $\exp\left\{\log\left(\overline{p}_{kt}^{Aw}/\tilde{\overline{p}}_{k\tau}^{Aw}\right)\right\}$, in turn, expresses change in prices standardised for quality (for analogy see Koev, 2003, p. 23) when qualitative factors equal the situation at the comparison period. A simple brain exercise is in place here: Let us assume that qualitative properties are alike $\overline{x}'^{Aw}_{k\tau}\equiv\overline{x}'^{Aw}_{kt}$, whereby $\exp\left\{\left(\overline{x}'^{Aw}_{kt}-\overline{x}'^{Aw}_{k\tau}\right)\hat{\beta}_{k\tau}\right\}=\exp(0)=1$ (i.e. no change in quality). When we substitute $\overline{x}'^{Aw}_{k\tau}\equiv\overline{x}'^{Aw}_{kt}$ with the index standardised for quality, we obtain $\exp\left\{\hat{\alpha}_{kt}^{Aw}-\hat{\alpha}_{k\tau}^{Aw}+\overline{x}_{kt}'^{Aw}\left(\hat{\beta}_{kt}-\hat{\beta}_{k\tau}\right)\right\}=\exp\left\{\left(\hat{\alpha}_{kt}^{Aw}+\overline{x}_{kt}'^{Aw}\hat{\beta}_{kt}\right)-\left(\hat{\alpha}_{k\tau}^{Aw}+\overline{x}_{k\tau}'^{Aw}\hat{\beta}_{k\tau}\right)\right\}$ = $\overline{p}_{kt}^{Aw}/\overline{p}_{k\tau}^{Aw}$. In other words, if no change in quality is present, decomposition (4.1) of the index reduces it to a classic classification index (see repeat-sales and hybrid models; Bailey, Muth and Nourse, 1963; Case and Shiller, 1989; Quigley, 1995).

The model for 'commodities' due for repeat sales assumes that $\overline{x}'^{Aw}_{k\tau}\equiv\overline{x}'^{Aw}_{kt}$ and the index calculation becomes reduced to a classic classification method. However, if the age of the commodity that is due for repeat sales can be equated to its 'physical depreciation' and it has a significant price impact, $\exp\left\{\left(\overline{x}'^{Aw}_{kt}-\overline{x}'^{Aw}_{k\tau}\right)\hat{\beta}_{k\tau}\right\}=\exp(0)=1$ does not follow even if we assume that $\overline{x}'^{Aw}_{k\tau}\equiv\overline{x}'^{Aw}_{kt}$ and

$\hat{\beta}_{k\tau} = \hat{\beta}_{kt}$. In other words, if $\bar{x}'^{Aw}_{k\tau} \equiv \bar{x}'^{Aw}_{kt}$ does not hold true, change in arithmetic mean prices is partly explained by qualitative changes in characteristics and the remainder of the price change represents price change when quality is standardised. Analytically this means that the change in prices is analysed by decomposing decomposition (4.1).

4.2 Aggregation from stratum level to total data level

The index application of decomposition (4.1) is developed for Laspeyres' index formula - Paasche's index formula can be deduced analogously with Laspeyres' analysis. Fisher, in turn, is obtained as the geometric mean of Laspeyres' and Paasche's. The mathematics of Paasches' and Fisher's formulae are left to the reader.

It is generally known that Laspeyres's price index can be written in logarithmic form (Y. Vartia, 1976, p. 128), which as far as is known has never been applied to index solutions in practice. As a point of departure for an index application this sounds suspicious - the choice of a new perspective is perplexing. The logarithm of Laspeyres' index formula is defined as

$$\text{(4.2)} \qquad \log\left(\frac{\sum_k q_{k\tau}\bar{p}^{Aw}_{kt}}{\sum_k q_{k\tau}\bar{p}^{Aw}_{k\tau}}\right) = \sum_k w_{k\tau}\log\left(\frac{\bar{p}^{Aw}_{kt}}{\bar{p}^{Aw}_{k\tau}}\right), \text{ where } w_{k\tau} = \frac{L\left(q_{k\tau}\bar{p}^{Aw}_{kt}, q_{k\tau}\bar{p}^{Aw}_{k\tau}\right)}{L\left(\sum_k q_{k\tau}\bar{p}^{Aw}_{kt}, \sum_k q_{k\tau}\bar{p}^{Aw}_{k\tau}\right)}.$$

Variable $q_{k\tau}$ measures the quantity of stratum A_k at base period τ and variables $\bar{p}^{Aw}_{k\tau}, \bar{p}^{Aw}_{kt}$ arithmetic mean prices in the respective stratum at points in time (τ, t).

If we substitute logarithmic (arithmetic) mean prices in equation (4.2) with equation (4.1) and proceed as per Annex 5, we receive (Annex 5 eq. (14))

$$\text{(4.3)} \qquad \sum_k w^L_{k0}\log\left(\frac{\bar{p}^{Aw}_{kt}}{\bar{p}^{Aw}_{k\tau}}\right) = \log\left(\frac{\bar{p}^{Aw}_{t}}{\bar{p}^{Aw}_{\tau}}\right) = \left\{\left(\bar{x}'^{Aw}_t - \bar{x}'^{Aw}_\tau\right)\hat{\bar{\beta}}^{Aw}_\tau + \bar{x}'^{Aw}_t\left(\hat{\bar{\beta}}^{Aw}_t - \hat{\bar{\beta}}^{Aw}_\tau\right) + \left(\hat{\bar{\alpha}}^{Aw}_t - \hat{\bar{\alpha}}^{Aw}_\tau\right)\right\}$$

which holds true identically if we select as weights $w^L_{k\tau}$

$$w^L_{k\tau} = \frac{L\left(q_{k\tau}\bar{p}^{Aw}_{kt}, q_{k\tau}\bar{p}^{Aw}_{k\tau}\right)}{L\left(\sum_k q_{k\tau}\bar{p}^{Aw}_{kt}, \sum_k q_{k\tau}\bar{p}^{Aw}_{k\tau}\right)}$$

Finally, by locating equation (4.3) to the right of equation (4.2) and taking an exponential transformation of it on both sides, we receive (see Annex 5, equation (15))

$$\text{(4.4)} \qquad \frac{\sum_k q_{k\tau}\bar{p}^{Aw}_{kt}}{\sum_k q_{k\tau}\bar{p}^{Aw}_{k\tau}} = \exp\left\{\left(\bar{x}'^{Aw}_t - \bar{x}'^{Aw}_\tau\right)\hat{\bar{\beta}}^{Aw}_\tau\right\}\exp\left\{\left(\hat{\bar{\alpha}}^{Aw}_t + \bar{x}'^{Aw}_t\hat{\bar{\beta}}^{Aw}_t\right) - \left(\hat{\bar{\alpha}}^{Aw}_\tau + \bar{x}'^{Aw}_t\hat{\bar{\beta}}^{Aw}_\tau\right)\right\}$$

where the left side of the equation is a typical Laspeyres classification index. The first right-hand term contains quality standardisations of the quality factors controlled in the index calculation. The second term, in turn, represents Laspeyres price index standardised for quality.

From the perspective of a statistical expert, the mathematics of Chapters 2, 3 and 4, and Annex 5, which lead to price index (4.4) are at best a macabre joke - they contain statistical science, latest price aggregation principles, re-parameterization of price models, appliance of the basic aggregation clause to the aggregation of heterogeneous price models (Vartia, 1979) and eventually deviant

mathematics on index figures. They simply do not seem to relate in any way to statistics in practice. The conclusion is natural because as far as we know the classic classification index (here Laspeyres') has never before been presented in a parametric form at the level of whole economy. Nevertheless, for instance the Consumer Price Index can be derived from equation (4.4) as follows: The commodities in the Consumer Price Index are of equal quality (i.e. $\bar{x}_{kt}' \equiv \bar{x}_{k\tau}'$), whereby quality adjustment factors disappear in individual strata and compound strata. Therefore, basing on the analyses of Chapter 3 and Annex 5, (4.4) becomes reduced to Laspeyres' price index (4.2) (in other words, its exp transformation). The analysis corresponds precisely with the so-called matched pairs method (See e.g. Bailey, Muth and Nourse, 1963; Case and Shiller, 1989; Quigley, 1995; Koev, 2003).
Annex 5 presents the analogy of (4.4) for quarterly statistics on dwelling prices (Koev, 2003) where the log-Laspeyres receives the following parametric expression (See Annex 5 eq. (11) and (12))

$$(4.5) \qquad \sum_k w_{k\tau}^{Koev} \log\left(\frac{\bar{p}_{kt}^G}{\bar{p}_{k\tau}^G}\right) = \frac{\bar{p}_t^G}{\bar{p}_\tau^G} = \exp\left\{(\bar{x}_t' - \bar{x}_\tau')\hat{\hat{\beta}}_0\right\}\exp\left\{\left(\hat{\hat{\alpha}}_t + \bar{x}_t'\hat{\hat{\beta}}_t\right) - \left(\hat{\hat{\alpha}}_\tau + \bar{x}_t'\hat{\hat{\beta}}_\tau\right)\right\}$$

At the level of the whole country, the quarterly statistics on dwelling prices correspond precisely with log-Laspeyres price index (4.5) in parametric form.

As price indices (4.4) and (4.5) show, the core finding of the study can be generalised as follows: Index calculations performed with heterogeneous statistical data, inclusive of hedonic price indices, become reduced to calculations of parameters and means of input and output variables. This study, too, ends up in calculating means - yet, not haphazardly - but by controlling it with coherent and mathematically consistent analysing methods.

5 EMPIRICAL EXAMPLE WITH RENTS OF OFFICE AND SHOP PREMISES

The study is limited to the rents of office and shop premises only. The data on both office and shop premises are divided into new and old tenancy agreements. An inquiry about rents covers all tenancy agreements once a year. In addition, a separate inquiry in February/March covers new tenancy agreements. Indices complying with the theoretic analysis are constructed for both new and old tenancy agreements in the study. The mathematics of the study are programmed into an index application with the SAS, SAS/STAT, SAS/IML and SAS/AF software packages. The performer of the study is responsible for the implementation of the programming while Seppo Suomalainen designs and implements the user interface. The tables of the results and the statistics on mean prices generated in the index calculation are automatically converted in the application into Excel files of directly publishable format.

5.1 Definition of the statistical data

The creation of the statistical data in the SAS application is automated from the ACCES database into SAS files. The validation of the statistical data is controlled in the SAS application inside the data at the reading stage as far as possible. The study uses the definitions applied by KTI: Regional rent levels are controlled within the limits of Tables 5.1a and b. In addition, the size of a rented office premises must be at least nine square metres. No limits are set for the floor area of shop premises.

Table 5.1a: Limits for rents levels of office premises by area (same as those of KTI).

Municipality	Lower limit of rents (EUR/m^2)	Upper limit of rents (EUR/m^2)
Helsinki	4	30
Espoo, Vantaa and Kauniainen	4	22
Tampere and Turku	3.5	17
Jyväskylä, Kuopio, Lahti, Oulu	3	15
Other municipalities	3	13

Table 5.1b: Limits for rents levels of shop premises by area (same as those of KTI).

Municipality	Lower limit of rents (EUR/m^2)	Upper limit of rents (EUR/m^2)
Helsinki	5	120
Espoo, Vantaa and Kauniainen	4	80
Tampere and Turku	4	70
Jyväskylä, Kuopio, Lahti, Oulu	4	65
Other municipalities	3	50

5.2 Stratifications of the statistical data

Because there are large regional differences in the levels of rents, the data are divided into five estimation categories by area. This estimation classification is consistent for both new and old tenancy agreements. Because the locations of office and shop premises vary considerably, internal variation of rent levels within the estimation categories is controlled with an additional classification based on municipal sub-areas (so-called village level classification). The partitioning of the estimation categories into smaller sub-areas will hereafter be referred to as micro classification. The micro classifications of office and shop premises are not congruent.

Table 5.2: Numbers of strata of office and shop premises by estimation areas (same stratification is applied to both new and old agreements - data on new agreements missing from rent levels and qualitative characteristics are substituted with mean for the micro class)

Estimation areas	**Micro classes of office premises**	**Micro classes of shop premises**
Helsinki	15	15
Espoo, Vantaa and Kauniainen	12	10
Tampere and Turku	10	6
Jyväskylä, Kuopio, Lahti, Oulu	8	8
Other municipalities	11	23
Micro classes, total	56	62

More precise division of estimation areas into strata is shown in Annexes 1 and 2.

5.3 Price model estimation results

The study analyses the rents per square metre of office and shop premises for new and old tenancy agreements. All tenancy agreements are analysed annually in five estimation categories in accordance with Table 5.2. Each estimation category divides into smaller sub-areas, such as municipality and its sub-areas, so the number of independently estimated parameters easily grows large over time. Their detailed presentation is simply not purposeful. The co-efficient estimates of rent equations, their t values (dispersions), coefficients of determination (R^2) and prediction errors are presented in the study by aggregating the estimation results of the equations from equation level to aggregate level. The method is described in the margin (footnote) of the next page.

Table 5.3 presents the means and key figures for the coefficient estimates of rent equations for office premises for the 2002-2004 time period. Other estimation results are presented in Annex 3. At the observation level, standard errors in the price models for office premises amount to approximately 20-30 log per cent, so the standard errors in the mean (standard errors/ $\sqrt{N_t}$) amount to around 0.3-0.4 log per cent. The estimation results of office premises deviate from other dwelling price statistics (see Koev, E. & Suoperä, A, 2002, Koev, E., 2003) in respect of the impact of the floor area of the target of rental on price - unlike in other statistics on the prices of real estate/dwellings the size of the floor area of the rental target does not seem to have a significant impact on rent levels at all points in time. By contrast, age is a very central explanatory variable in equations for rents of office premises.

Table 5.4 contains the main estimation results for shop premises for the 2002-2004 time period. Estimation results for other points in time are presented in Annex 4. The results principally differ from those for office premises as follows: Standard errors at observation level are larger than for office premises at approximately 50 log per cent. Standard errors of the mean, in turn, are of the magnitude of around 0.5 log per cent. The coefficients of determination of the model are lower than for office premises, although as a rule the explanatory variables of the model are even more significant than in the rent equation for office premises, on the average.

Table 5.3: Estimation results for office premises[2] in 2002/1-2004/2 (standard errors in brackets)

Key figures/Point in time	2002/1	2002/2	2003/1	2003/2	2004/1	2004/2
N	8877	7094	7832	7113	7880	6547
Micro classes	110	112	112	110	110	112
Adj. R^2	0.624317	0.633493	0.636713	0.635698	0.635692	0.64335
RMSE	0.272951	0.268587	0.267779	0.267813	0.267419	0.258001
Constant	2.590432	2.487331	2.515395	2.623503	2.611727	2.725481
	(0.017933)	(0.023453)	(0.02245)	(0.025972)	(0.023983)	(0.027302)
Floor area	-0.00001	-0.00002	-0.00002	-0.00003	-0.00003	-0.00002
	(4.793E-6)	(3.94E-6)	(3.858E-6)	(5.452E-6)	(5.117E-6)	(4.806E-6)
Floor area$^{1/2}$	0.001913	0.003858	0.003603	0.004457	0.004003	0.003177
	(0.000442)	(0.000424)	(0.000406)	(0.000486)	(0.00046)	(0.000451)
Age at basic renovation	0.007799	0.003077	0.003984	0.010987	0.007827	0.008695
	(0.001644)	(0.001994)	(0.001924)	(0.002153)	(0.002017)	(0.002274)
Age at basic renovation$^{1/2}$	-0.1231	-0.07704	-0.08492	-0.13193	-0.11207	-0.13431
	(0.010307)	(0.013166)	(0.012712)	(0.014544)	(0.01353)	(0.015531)
He(α)	1	1	1	1	1	1
	(0.010444)	(0.011478)	(0.010844)	(0.010406)	(0.010119)	(0.01115)
He($x\beta$)	1	1	1	1	1	1
	(0.032926)	(0.023127)	(0.022711)	(0.017877)	(0.020005)	(0.019891)
$\sum$*He*($x\beta$)/ *N* (*log-%)*	0.010752	0.018431	0.018395	0.013979	0.012149	0.015519

Table 5.4: Estimation results for commercial premises in 2002/1-2004/2 (standard errors in brackets)

Key figures/Point in time	2002/1	2002/2	2003/1	2003/2	2004/1	2004/2
N	9864	7678	8260	7190	7863	6863
Micro classes	119	122	122	119	124	120
Adj. R^2	0.445272	0.443523	0.448503	0.444661	0.445233	0.462205
RMSE	0.515011	0.498803	0.497131	0.502478	0.505206	0.50583
Constant	3.240823	3.590222	3.601913	3.568135	3.554372	3.488971
	(0.053446)	(0.053708)	(0.051843)	(0.055716)	(0.053343)	(0.069356)
Floor area	0.000153	0.000159	0.000164	0.000184	0.000196	0.000191
	(0.000013)	(0.000013)	(0.000013)	(0.000015)	(0.000015)	(0.000013)
Floor area$^{1/2}$	-0.01618	-0.01827	-0.01865	-0.01959	-0.02064	-0.0224
	(0.001029)	(0.001085)	(0.001056)	(0.001166)	(0.001134)	(0.001095)
Age at basic renovation	0.01487	0.034699	0.033875	0.018736	0.015044	-0.00772
	(0.004318)	(0.004534)	(0.00444)	(0.004691)	(0.004509)	(0.005678)
Age at basic renovation$^{1/2}$	-0.19901	-0.3419	-0.34011	-0.25687	-0.23554	-0.101
	(0.029789)	(0.030403)	(0.029611)	(0.031472)	(0.030229)	(0.039235)
He(α)	1	1	1	1	1	1
	(0.011576)	(0.013891)	(0.013393)	(0.013697)	(0.013057)	(0.014435)
He($x\beta$)	1	1	1	1	1	1
	(0.037185)	(0.020331)	(0.018234)	(0.019361)	(0.02214)	(0.019361)
$\sum$*He*($x\beta$)/ *N* (*log-%)*	0.001557	0.001501	0.001736	0.003325	0.003885	0.008363

---

2 *Covariation variables* (*He*(α) and *He*($x\beta$)) can be calculated with formula $c_{ikt} = (\hat{\alpha}_{kt} - \hat{\hat{\alpha}}_t) + x'_{ikt}(\hat{\beta}_{kt} - \hat{\hat{\beta}}_t)$ where 'population parameters' $\hat{\hat{\alpha}}_t, \hat{\hat{\beta}}_t$ are weighted averages of strata level OLS estimates (that is, weights are relative frequencies of the micro classification). Thus the model used in parameter estimation in Tables 5.3 and 5.4 can congruently be expressed as $\log(p_{ikt}) = \hat{\hat{\alpha}}_t + x'_{ikt}\hat{\hat{\beta}}_t + c_{ikt}1 + \hat{\varepsilon}_{ikt}$. This is rewriting of the original heterogeneous models as a 'one equation model' which reproduces precisely the original fits, residuals and mean parameter estimates. This manner of presentation is *synthesis* of heterogeneous models performed for macro analysis, which decomposes the original models into *a common element and heterogeneity effects* (covariations). Basing on the OLS definition, we can prove even the following stronger result: If we calculate covariation variables from the first stage and in the second stage estimate model $\log(p_{ikt}) = \alpha_t + x'_{ikt}\beta_t + c_{ikt}\gamma_t + \varepsilon_{ikt}$ with OLS, we receive precisely the aforementioned model, including e.g. its estimates, i.e. $est(\alpha_t, \beta_t, \gamma_t) = (\hat{\hat{\alpha}}_t, \hat{\hat{\beta}}_t, 1)$. This is because the information allows optimum OLS solutions by equation. As an important additional result we obtain standard errors of all mean parameters and accuracies of estimates as their inverse figures. The exactitudes of the mean parameters are naturally multifold compared to the exactitudes of the parameters of respective heterogeneous sector models.

The mathematics of the 'covariation variable' shown in the last row of the tables is explained in the margin of the page. This variable collates the systematic statistical information originating from the heterogeneous behaviour of the stratification. The estimation results prove that it is necessary to take into account the heterogeneity of the micro classes in the modelling of the determination of rent levels.

Figure 5.1 presents the average 'age-effect' on the rent levels of office and shop premises. Age has a stronger impact on the rent levels of shop premises than on those of office premises. The rent levels of both office and shop premises are generally lower for older than for newer real estate. Because the index calculation requires from an 'index commodity' qualitative comparability between the base and comparison periods, growth of the average age of real estate over time requires upwards correction of the index and vice versa (See Koev, 2003).

Figure 5.1: Average price impact of age at basic renovation on rents of office and shop premises in log per cent in the 1995-2004 period

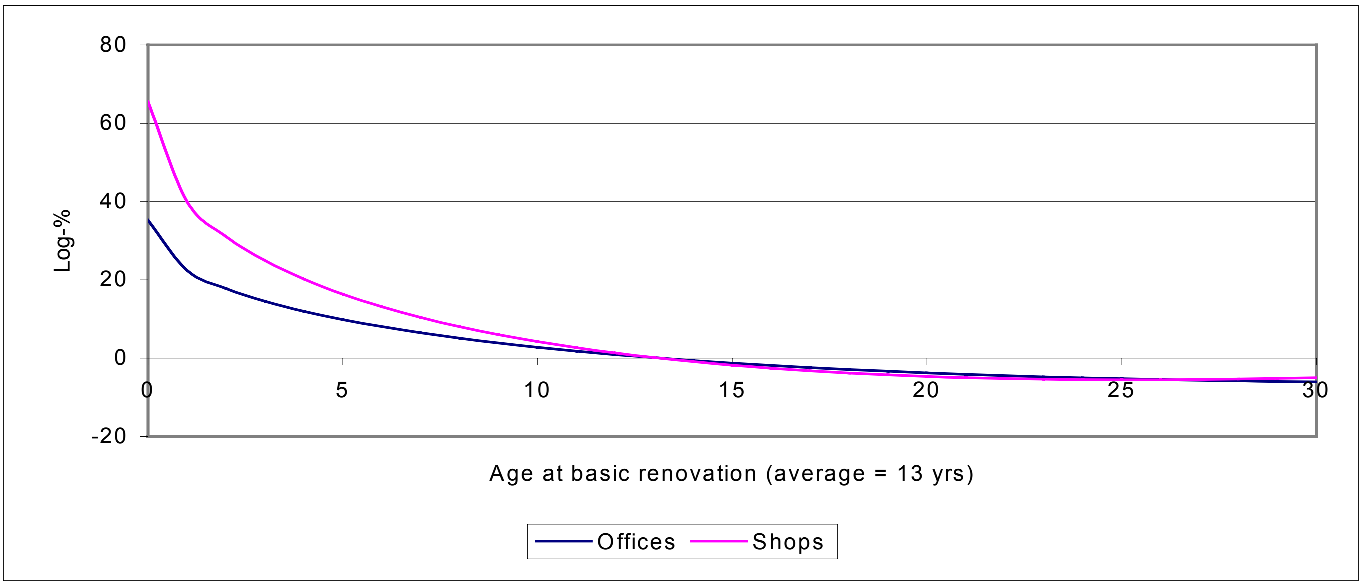

Figure 5.2: Average price impact of floor area of target of rental on rents of office and shop premises in log per cent in the 1995-2004 period

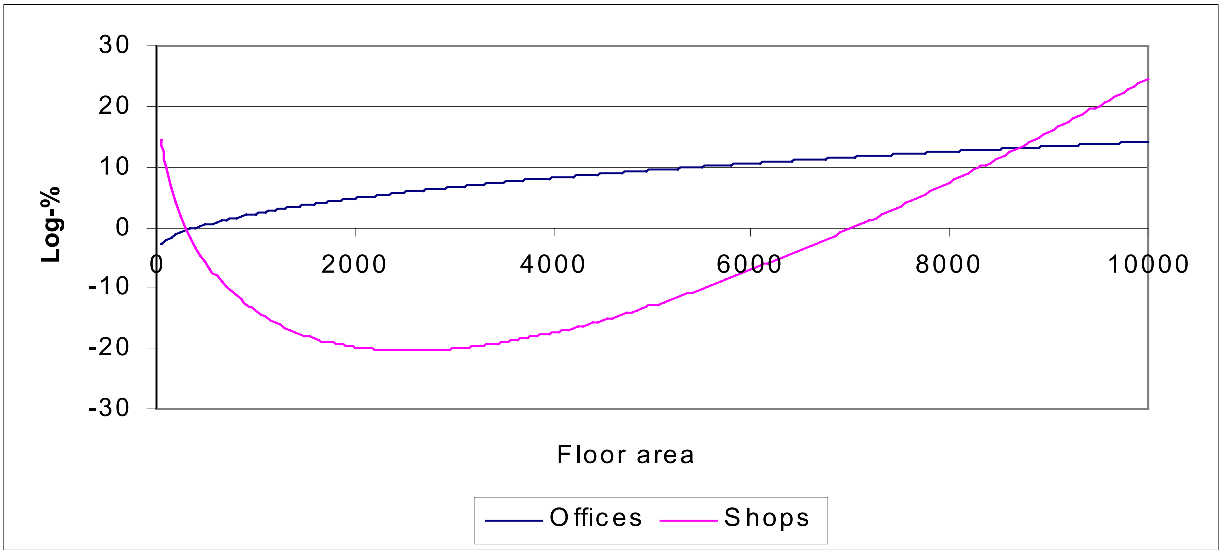

Figure 5.2 shows the impact of the size of the target of rental on the rents per square meter of office and shop premises. With office premises the average price impact is somewhat ascending, which depicts stronger concentration of demand on larger than average rental targets. Systematic growth of floor area of office premises over time means that the index must be corrected downwards, unlike in other statistics on dwelling prices (Koev, 2002). With shop premises the growth of floor area initially lowers the rent level but in the case of rental targets whose floor area exceeds 2,000 square metres the growth of the floor area has a positive impact on the rent level. As in the case of age at basic renovation, growth of the average floor area of shop premises over time means that the index must be corrected upwards in the descending part of the profile, and vice versa in its ascending part.

The profiles for the price impacts of age and size in Figures 5.1 and 5.2 have been estimated for the entire examined time period and they represent the price impacts of age at basic renovation and floor area relative to an average rental target. These profiles may deviate significantly by region from the situation shown here.

5.4 Index calculation with rents of office premises

Micro indices for new and old agreements at the stratum level are obtained direct by differentiating the estimated stratum level price models. Price changes are aggregated together with Laspeyres' logarithmic clause. The study develops an indicator to describe how the prices of new and old agreements change relative to the whole stock. Index series are calculated with both unweighted and weighted arithmetic means for the strata in Annexes 1 and 2. Altogether over 2,000 stratum level micro indices are calculated in the empirical analysis. Chained Laspeyres' indices are developed for each index series to describe price development over time. The presentation of all index series is unnecessary in this context, so this Chapter focuses on the change of rents per square meter in one micro index area and in the whole country during the 1995/1-2005/1 time period.

The results correspond precisely with the decompositions of either equation (4.1.) or (4.4): Quality changes in the variables controlled in the index calculation, change in rents per square metre standardised for quality and the classic Laspeyres' classification index are each estimated in the ways described in detail in Chapter 4. The combined impact of quality adjustments is obtained as a product of individual quality adjustment factors. The following holds true for the decomposition: The product of separately estimated quality adjustment factors and change in prices standardised for quality corresponds with change in arithmetic mean prices. The situation is described by Tables 5.5a and 5.5b, which show the index point estimates and chain indices for the rents of office premises in Kluuvi, Helsinki, during the 1005/2-2005/1 time period.

Table 5.5a: Changes in rents per square metre of office premises in Kluuvi calculated with old agreements for the 1995/2-2005/2 time period (aggregation of observations (3.5b) p. 6))

Point in time	Number of agreements	Arithmetic mean price $\overline{p}_t^{Aw}$	Laspeyres' index	Qality adjusted Laspeyres's index	Adjustments for quality	
					Floor area	Age at basic renovation
1995/2	257	15,71104	100	100	1	1
1996/1	329	15,89585	101,1763	100,3642	0,99529	1,012862
1996/2	328	14,93857	93,97783	94,82938	1,012053	0,979218
1997/1	392	14.5896	97.66393	97.70707	0.998245	1.001316
1997/2	359	15.42585	105.7318	107.1175	0.988768	0.998277
1998/1	418	15.46423	100.2488	100.4845	0.997801	0.999853
1998/2	402	16.40113	106.0585	106.0947	1.002872	0.996796
1999/1	448	16.53867	100.8386	100.8958	0.998349	1.001086
1999/2	388	17.48472	105.7202	105.5853	1.000257	1.001021
2000/1	457	18.63326	106.5688	105.4831	1.012587	0.997733
2000/2	423	19.752	106.004	107.5703	0.983297	1.002179
2001/1	490	20.00034	101.2573	101.5537	0.998238	0.998841
2001/2	444	21.06741	105.3352	102.9046	0.99943	1.024203
2002/1	482	21.35766	101.3777	101.3517	1.001108	0.99915
2002/2	399	22.26379	104.2427	104.4967	1.00672	0.99091
2003/1	436	22.30416	100.1813	100.4754	0.997827	0.999245
2003/2	369	23.2988	104.4594	102.2486	1.009835	1.011673
2004/1	428	22.99362	98.69012	98.39744	0.998593	1.004388
2004/2	365	23.31567	101.4006	101.8301	1.000815	0.994971
2005/1	401	23.22632	99.61678	99.7373	0.998357	1.000435

Table 5.5b: Chain index 1995/2=100 of rents per square metre of office premises in Kluuvi calculated with old agreements for the 1995/2-2005/2 time period (aggregation of observations (3.5a) p. 6))

Point in time	Number of agreements	Arithmetic mean price $\overline{p}_t^{Aw}$	Laspeyres' index	Qality adjusted Laspeyres's index	Adjustments for quality	
					Floor area	Age at basic renovation
1995/2	257	15.71104	100	100	1	1
1996/1	329	15.89585	101.1763	100.3642	0.99529	1.012862
1996/2	328	14.93857	95.08326	95.17475	1.007286	0.991812
1997/1	392	14.5896	92.86204	92.99246	1.005518	0.993117
1997/2	359	15.42585	98.18472	99.61117	0.994224	0.991406
1998/1	418	15.46423	98.42903	100.0938	0.992038	0.99126
1998/2	402	16.40113	104.3923	106.1942	0.994888	0.988083
1999/1	448	16.53867	105.2678	107.1456	0.993245	0.989156
1999/2	388	17.48472	111.2894	113.1299	0.9935	0.990167
2000/1	457	18.63326	118.5997	119.333	1.006006	0.987922
2000/2	423	19.752	125.7205	128.3669	0.989202	0.990075
2001/1	490	20.00034	127.3011	130.3613	0.98746	0.988927
2001/2	444	21.06741	134.093	134.1478	0.986897	1.012862
2002/1	482	21.35766	135.9404	135.961	0.987991	1.012001
2002/2	399	22.26379	141.7079	142.0748	0.994631	1.002802
2003/1	436	22.30416	141.9649	142.7502	0.992469	1.002045
2003/2	369	23.2988	148.2957	145.96	1.00223	1.013741
2004/1	428	22.99362	146.3532	143.6209	1.00082	1.01819
2004/2	365	23.31567	148.4031	146.2493	1.001636	1.013069
2005/1	401	23.22632	147.8344	145.8651	0.99999	1.01351

The differences between the classic classification index and the index standardised for quality are minor. When chained, the quality adjustment for floor area is close to 1.35 per cent. Because the floor areas of the rental targets grow and some of the price increase is explained by the growth in the floor area, the requirement for equal quality means that the classification index must be corrected downwards by the same 1.35 per cent (For justification see Figure 5.2).

Table 5.6a: Development of (unweighted) mean rents of office premises per square metre, quality adjustment factors and change in mean prices standardised for quality 1995/2=100 in whole Finland (aggregation of observations (3.5b) p. 6))

Point in time	Laspeyres' index	Qality adjusted Laspeyres's index	Adjustments for quality	
			Floor area	Age at basic renovation
1995/2	100	100	1	1
1996/1	99.21216	99.32787	0.998087	1.00075
1996/2	97.74959	98.92159	0.998859	0.989281
1997/1	97.49622	98.63903	0.998841	0.989561
1997/2	96.11946	97.77027	0.99769	0.985391
1998/1	96.39198	98.10098	0.997507	0.985035
1998/2	98.42627	100.4031	0.998867	0.981423
1999/1	99.86575	101.8856	0.998439	0.981708
1999/2	105.3214	107.4701	0.998874	0.981112
2000/1	106.7714	108.8327	0.99911	0.981934
2000/2	112.0688	114.7404	0.998852	0.977839
2001/1	114.5953	117.3275	0.999128	0.977566
2001/2	121.0327	123.8603	1.000274	0.976904
2002/1	122.4616	124.9895	1.000733	0.979057
2002/2	126.3071	129.3299	1.002821	0.97388
2003/1	126.7574	129.6671	1.002198	0.975416
2003/2	130.3482	133.7447	1.002474	0.972199
2004/1	130.5492	133.9151	1.00166	0.973249
2004/2	132.4315	136.3435	1.001474	0.969879
2005/1	132.3287	136.4151	1.000981	0.969093

Table 5.6b: Development of (weighted) mean rents of office premises per square metre, quality adjustment factors and change in mean prices standardised for quality 1995/2=100 in whole Finland (aggregation of observations (3.5a) p. 6))

Point in time	Laspeyres' index	Qality adjusted Laspeyres's index	Adjustments for quality	
			Floor area	Age at basic renovation
1995/2	100	100	1	1
1996/1	98.82051	99.128	0.997185	0.999713
1996/2	95.97759	97.11563	0.999528	0.988749
1997/1	95.15724	96.50556	0.997475	0.988525
1997/2	94.11951	96.08957	0.991907	0.987489
1998/1	94.21439	96.47679	0.989572	0.986841
1998/2	97.0128	99.74011	0.993825	0.978699
1999/1	97.67792	100.5674	0.9918	0.979298
1999/2	100.7931	103.9779	0.987694	0.981449
2000/1	102.457	105.4244	0.987417	0.984237
2000/2	108.5364	112.6151	0.978752	0.984705
2001/1	110.0883	114.2859	0.978138	0.9848
2001/2	117.7175	121.9085	0.979682	0.985648
2002/1	118.9247	123.0409	0.979808	0.986466
2002/2	125.9205	130.0981	0.98217	0.98546
2003/1	126.508	130.9965	0.977141	0.988328
2003/2	129.9037	133.3921	0.990495	0.983194
2004/1	129.9836	133.232	0.990403	0.985071
2004/2	132.4422	136.5734	0.989155	0.980383
2005/1	132.6564	136.3087	0.989496	0.983537

Tables 5.6.a and 5.6.b show the development of rents per square metre for office premises since 1995 by aggregation principles 3.5a and 3.5b. The aggregation method (i.e. eq. (3.5a) and (3.5b)) has no significance in the case of whole Finland - Laspeyres' classification indices are of almost the same magnitude irrespective of the method in the case of whole Finland (cf. columns 1 of Tables 5.6a and 5.6b). By contrast, the quality adjustments for age at basic renovation and floor area differ from each other according to the used aggregation method. In the method of aggregation of unit prices the need for quality adjustment for age at basic renovation amounts to approximately 3.1 per cent whereas in the case of weighted aggregation the respective need is half of this. Because the average age of the targets of rental is higher the index must be corrected upwards (See Figure 5.1). The unweighted aggregation method does not recommend any kind of quality adjustment for floor area. By contrast, in the weighted case the mean floor area weighted by observation falls systematically by approximately 20 to 25 per cent from 1995/2 to 2005/1. As Figure 5.2 shows, in the case of office premises growth in the floor area of the rental target raises the rent per square metre, so systematic fall in the mean floor area from the base time period to the present requires upwards correction of the index.

5.5 Index calculation with rents of shop premises

Figures 5.1 and 5.2 show that growth in age at basic renovation lowers the rent level in the case of shop premises. Systematic growth in the age at basic renovation requires upwards correction of the index.

Large changes in floor area are in practice born at three measurement points - in old agreements in Vantaa and Pori at 1998/2, in old agreements in Porvoo at 1999/2, in new agreements in central Tampere at 1999/2 and in old agreements in Helsinki's Ruoholahti at 2004/2. Large floor area changes reflect changes in the compilation of the data rather than actual changes of floor area. Large quality adjustments for floor area that are due to the compilation of the statistics are eliminated so that the index component concerned behaves more moderately and quality adjustment for floor area becomes insignificant in both aggregation methods over the 1995/2-2005/1 time period.

Table 5.7a: Development of (unweighted) mean rents of shop premises per square metre, quality adjustment factors and change in mean prices standardised for quality 1995/2=100 in whole Finland (aggregation of observations (3.5b) p. 6))

Point in time	**Laspeyres' index**	**Qality adjusted Laspeyres's index**	**Adjustments for quality**	
			Floor area	**Age at basic renovation**
1995/2	100	100	1	1
1996/1	99.50339	99.22401	1.002574	1.000241
1996/2	99.27926	99.75496	1.000795	0.99444
1997/1	98.88131	99.29038	1.001126	0.99476
1997/2	97.26445	98.12825	1.003551	0.98769
1998/1	97.02747	97.82475	1.004745	0.987166
1998/2	100.7283	102.4353	1.004141	0.97928
1999/1	100.2897	101.7986	1.005137	0.980142
1999/2	107.6702	110.2125	1.000926	0.976029
2000/1	108.5155	110.928	1.001425	0.976859
2000/2	108.1548	112.0193	0.99959	0.965898
2001/1	108.4946	111.8489	1.000362	0.969659
2001/2	114.2379	118.6534	0.998541	0.964193
2002/1	114.0913	118.3745	0.999792	0.964017
2002/2	118.9102	125.0264	0.996296	0.954616
2003/1	118.2212	124.213	0.996955	0.954669
2003/2	121.667	128.9282	0.99761	0.945941
2004/1	122.0147	129.0681	0.999469	0.945854
2004/2	124.2804	132.9663	0.99487	0.939496
2005/1	124.0024	132.5159	0.995554	0.939934

Quality adjustments for age at basic renovation are minor between successive points in time but cumulate from 1995/2 to 2005/1 to approximately 6 per cent in the unweighted method and to around 4.7 per cent in the weighted method. Because lower rents are paid for targets that are older at the time of basic renovation than for newer targets (See Figure 5.1), the index must be corrected upwards.

Table 5.7b: Development of (weighted) mean rents of shop premises per square metre, quality adjustment factors and change in mean prices standardised for quality 1995/2=100 in whole Finland (aggregation of observations (3.5a) p. 6))

Point in time	Laspeyres' index	Qality adjusted Laspeyres's index	Adjustments for quality	
			Floor area	Age at basic renovation
1995/2	100	100	1	1
1996/1	99.0163	98.97105	1.000296	1.000161
1996/2	97.49307	97.98	0.999901	0.995129
1997/1	97.484	97.8992	1.00051	0.995251
1997/2	95.21988	96.58443	0.997685	0.98816
1998/1	94.74101	95.80639	0.998968	0.989901
1998/2	100.2305	102.3352	0.998994	0.98042
1999/1	99.95503	101.9382	0.999897	0.980646
1999/2	107.1107	108.9993	1.007505	0.975352
2000/1	106.6805	108.4594	1.00721	0.976558
2000/2	108.8736	111.3836	1.004563	0.973025
2001/1	109.1016	111.6815	1.00284	0.974133
2001/2	114.9847	118.2359	0.997126	0.975306
2002/1	114.9059	117.9852	0.99777	0.976078
2002/2	117.3236	121.1303	1.000757	0.967841
2003/1	117.019	120.7806	1.003149	0.965815
2003/2	121.6128	126.6202	1.002383	0.95817
2004/1	121.4494	126.1721	1.005144	0.957643
2004/2	123.6885	128.8335	1.008562	0.951915
2005/1	123.9989	129.3614	1.005777	0.95304

5.6 Office premises - indices for new agreements

According to the wishes of KTI, the index of new agreements compares new agreements in the base year to new agreements in the comparison point in time. New agreements are also analysed with both the unweighted and the weighted aggregation method. Because the volumes of data by micro area are small, especially the variable of floor area shows strong fluctuation. Small numbers of observations in the strata also create strong variation in rent levels, and compiling of separate statistics by stratum does not seem sensible. Their analysing should therefore be viewed as a kind of economic trend indicator. The only difference between the unweighted and the weighted aggregation method is quality adjustment for floor area which is almost insignificant in the case of the unweighted method but especially significant in the case of the weighted aggregation method. The index components are also almost identical in the two methods.

Table 5.8a: Development of (unweighted) mean rents of office premises per square metre, quality adjustment factors and change in mean prices standardised for quality from 1995/2=100 (aggregation of observations (3.5b) p. 6)). New agreements.

Point in time	Laspeyres' index	Qality adjusted Laspeyres's index	Adjustments for quality	
			Floor area	Age at basic renovation
1995/2	100	100	1	1
1996/1	100.5709	103.3729	1.001027	0.971896
1996/2	99.92888	101.7436	1.010822	0.971649
1997/1	101.3	102.9983	1.012366	0.971498
1997/2	102.323	104.5546	1.011894	0.967153
1998/1	105.0596	107.5731	1.014037	0.963115
1998/2	109.2535	111.0907	1.015245	0.968695
1999/1	119.1459	122.1053	1.013637	0.962636
1999/2	124.0084	125.714	1.021123	0.966027
2000/1	127.4844	129.7435	1.020941	0.962434
2000/2	133.7433	136.618	1.020651	0.959151
2001/1	138.7454	142.5455	1.02478	0.949805
2001/2	141.7797	144.7065	1.024612	0.956239
2002/1	141.749	144.2652	1.029918	0.954016
2002/2	139.3799	141.3422	1.024607	0.962434
2003/1	135.9739	138.0708	1.025272	0.960538
2003/2	135.464	137.6123	1.020186	0.964911
2004/1	139.4788	140.7046	1.020199	0.971661
2004/2	138.0409	138.8705	1.016787	0.977615
2005/1	137.9844	140.729	1.018105	0.96306

Table 5.8b: Development of (weighted) mean rents of office premises per square metre, quality adjustment factors and change in mean prices standardised for quality from 1995/2=100 (aggregation of observations (3.5a) p. 6)). New agreements.

Point in time	Laspeyres' index	Qality adjusted Laspeyres's index	Adjustments for quality	
			Floor area	Age at basic renovation
1995/2	100	100	1	1
1996/1	99.00129	100.4758	1.007138	0.978342
1996/2	104.1035	105.6744	1.005774	0.979479
1997/1	106.2454	106.4791	1.004879	0.992961
1997/2	102.328	104.9193	0.994285	0.980907
1998/1	104.7362	108.1368	0.993012	0.975369
1998/2	116.0304	118.8391	0.991554	0.984682
1999/1	120.8599	124.082	0.990734	0.983142
1999/2	120.8378	121.9465	1.011765	0.979385
2000/1	124.9377	127.3356	0.999208	0.981947
2000/2	133.6702	135.1137	0.99488	0.994408
2001/1	135.5084	137.9004	0.996615	0.985991
2001/2	144.2739	145.2572	0.989683	1.003585
2002/1	140.6419	144.0849	0.990335	0.985631
2002/2	136.9807	142.5933	0.969978	0.990372
2003/1	139.1733	146.3057	0.971976	0.978677
2003/2	133.1996	142.9218	0.949924	0.981105
2004/1	139.0316	149.4283	0.950285	0.9791
2004/2	140.37	154.5971	0.928466	0.977928
2005/1	141.447	155.1761	0.943096	0.966524

5.7 Shop premises - indices for new agreements

The volumes of data on new agreements concerning shop premises are small by stratum and especially in respect of floor area highly heterogeneous. In the small strata of shop premises rents per square metre and their range typically vary even more strongly than in the case of office premises. The floor areas of the targets of rental fluctuate strongly, and in consequence also the respective quality adjustment factors over time. At the level of the whole country new rental targets of shop premises have larger floor areas than in 1995, but their size fluctuates strongly by stratum. The age at basic renovation behaves more systematically - it grows from 1995 in the same way as in the analyses of other indices, so to retain the comparability of age at basic renovation the classification index must be corrected upwards. The central difference between the unweighted and weighted aggregations is in their Laspeyres' classification indices. The new agreements contain only few really large rental targets, which put even more emphasis on the significance of weighting. This explains the large differences between the Laspeyres' classification indices - aggregation of unit prices to stratum level produces random results compared to the weighted aggregation at the observation level.

At the level of the whole country, the mean floor areas of the strata remain relatively stable in the aggregation at the observation level and no major need to make adjustments for quality arises. By contrast, because the new agreements contain rental targets with very large floor areas in small strata, mean floor areas and ages at basic renovation vary strongly between the base and comparison periods in the weighted aggregation at the observation level. The following can be generalised about indices for new agreements: Non-stable price changes in new tenancy agreements for both office and shop premises principally arise from large structural changes in the small strata in age at basic renovation, floor area and rent per square metre. The situation is especially problematic in Helsinki's Itäkeskus.

Table 5.9a: Development of (unweighted) mean rents of shop premises per square metre, quality adjustment factors and change in mean prices standardised for quality from 1995/2=100 (aggregation of observations (3.5b) p. 6)). New agreements.

Point in time	Laspeyres' index	Qality adjusted Laspeyres's index	Adjustments for quality	
			Floor area	Age at basic renovation
1995/2	100	100	1	1
1996/1	106.9027	107.1388	1.002168	0.995637
1996/2	105.8499	108.6596	0.989752	0.98423
1997/1	107.3659	110.5942	0.983269	0.987329
1997/2	102.7713	105.9554	0.995987	0.973856
1998/1	107.8471	113.553	0.99278	0.956658
1998/2	114.8316	120.7564	0.99985	0.951079
1999/1	109.2124	115.6868	0.987877	0.95562
1999/2	124.8598	132.6495	0.990832	0.949985
2000/1	114.4661	121.5609	0.982164	0.958737
2000/2	111.4787	119.2062	0.9948	0.940064
2001/1	110.7087	118.7978	0.987606	0.943603
2001/2	112.6051	120.6158	0.987421	0.945478
2002/1	111.8421	120.8652	0.993392	0.931501
2002/2	105.3507	113.5894	0.998671	0.928704
2003/1	96.04156	104.5199	0.994036	0.924396
2003/2	101.6072	108.5527	1.006617	0.929864
2004/1	102.7707	111.4258	1.003584	0.91903
2004/2	94.75266	104.4357	0.99136	0.91519
2005/1	90.68876	100.4924	0.991592	0.910096

Table 5.9b: Development of (weighted) mean rents of shop premises per square metre, quality adjustment factors and change in mean prices standardised for quality from 1995/2=100 (aggregation of observations (3.5a) p. 6)). New agreements.

Point in time	Laspeyres' index	Qality adjusted Laspeyres's index	Adjustments for quality	
			Floor area	Age at basic renovation
1995/2	100	100	1	1
1996/1	102.3566	104.9494	0.988532	0.986609
1996/2	112.4552	115.1025	0.991633	0.985244
1997/1	122.8639	126.1627	0.988951	0.984733
1997/2	133.6802	137.7609	0.984311	0.985845
1998/1	124.8171	130.1068	0.979518	0.979404
1998/2	136.0767	138.5484	1.0051	0.977177
1999/1	124.7138	126.6485	1.0066	0.978267
1999/2	125.5761	126.0001	1.024178	0.973106
2000/1	116.7782	117.1027	1.017662	0.979922
2000/2	132.7105	133.241	1.030595	0.96645
2001/1	140.2231	140.7669	1.026282	0.970627
2001/2	145.9424	147.0226	1.018606	0.97452
2002/1	145.4299	148.7881	1.008515	0.969177
2002/2	155.421	160.5544	1.015349	0.953393
2003/1	154.9096	161.9211	1.024028	0.93425
2003/2	161.4271	167.5569	1.023784	0.941035
2004/1	166.1748	174.6923	1.034229	0.91976
2004/2	172.9625	178.4424	1.053962	0.919663
2005/1	172.4231	179.2244	1.045164	0.920479

7 CONCLUSIONS

The study examines semi-logarithmic model specifications in heterogeneously behaving cross-sectional strata. The examination combines classification and regression analysis. The models are specified as parallel in relation to the parameters and the specifications allow non-linearities of exogenous output variables. The standard price aggregation solution of the semi-logarithmic models for stratum level (i.e. elementary aggregate) is the geometric mean. Logarithmic unit prices are summed up by observation with weights of equal size in this method. The method has been applied widely in evaluations of pay differentials between women and men (Oaxaca (1973), Mincer (1974), Willis (1986), Card (1999), Vartiainen (2001), Bayard, Hellerstein and Troske (2003), Korkeamäki and Kyyrä (2002), Korkeamäki and Kyyrä (2003), Korkeamäki, Kyyrä and Luukkonen, (2004). Koev (2003) generalised the research method to index calculation by allowing non-linearities of price models and α and β heterogeneity over cross-sectional strata and time. Koev's log-Laspeyres' price index and its Oaxaca decomposition are shown in Annex 5 as parametric at the level of the whole country.

In Chapter 3, two new aggregation clauses based on the logarithmic mean (L. Törnqvist, 1935; Y. Vartia, 1976) are developed in which the aggregation of logarithmic means inside strata leads to either an unweighted or weighted logarithm of the arithmetic mean. Differences between the arithmetic and geometric means are derived accurately with the aggregation clauses in a case of an arbitrarily divided variable so that the commonly known second-order Taylor approximation can be precisely substituted with them. In Chapter 4 and Annex 5 these aggregation clauses are applied to heterogeneously behaving price, whose observations are aggregated into "average" price models at stratum level. Oaxaca decompositions are defined for the stratum level models in the typical manner, but instead of geometric mean prices they are developed with either weighted or unweighted arithmetic mean prices. The mathematical analysis is presented in Annex 5 for a linear model specification in heterogeneously behaving cross-sectional data. The indices are unusually presented in a parametric form at all aggregation levels.

Change in rents per square metre is calculated with Laspeyres' index formula at all aggregation levels. Index calculation from the micro class level to index categories at less detailed levels is performed with Laspeyres' logarithmic form (Y. Vartia, 1976, p. 128). Basing on Laspeyres' properties the method is consistent in aggregation.

Bibliography

Bailey M. J., Muth, R. F. and Nourse, H. O.: ”A Regression Model for Real Estate Price Index Construction”, JASA, vol. 58, 933-942, 1963.

Case, K. E. and Shiller, R. J.: ”Efficiency of the Market for Single Family Homes”, American Economic Review, vol. 79, 125-137, 1989.

van Dalen, J. and Bode, B.: ”Estimation Biases in Quality-Adjusted Hedonic Price Indices”, Roterdam School of Management, Rotterdam, 2004.

Davidson & MacKinnon : ”Estimation and Inference in Econometrics”, New York, Oxford University Press, 1993.

Hsiao, C.: Analysis of Panel Data”, Cambridge University Press, 1986.

Koev, E. : ”Palkkadiskriminaatio teollisuuden toimihenkilöillä”, ETLA discussion topics no. 568, Hel sinki, 1996. (in Finnish)

Koev, E. : ”Constructing a Hedonic Wage Index: Pilot Study for the Finnish Metal Industry”, EU Projects, 1997.

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)

Korkeamäki, O., Kyyrä, T. and Luukkonen, A.: ”Miesten ja naisten palkkaerot yksityisellä sektorilla”, VATT discussion intiatives no. 327, 2004. (in Finnish)

Kouvonen, S. and Suoperä, A.: ”VALPAS-tutkimus: Valtion ja Yleisten työmarkkinoiden palkkauskäytännöt vuosina 1995, 1996 ja 1997”, Ministry of Finance, 2000. (in Finnish)

Kouvonen, S. ja Suoperä, A.: ”KUNPAS-tutkimus: Kunnan ja Yleisten työmarkkinoiden palkka erot”, Association of Finnish Local and Regional Authorities, 2002 (in Finnish)

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

Quigley, R.: ”A Simply Hybrid Model for Estimating Real Estate Price Indexes”, Journal of Housing Economics, vol. 4, pp. 1 – 12, 1995.

Suoperä, A.: ”Kirkon ja muiden sektoreiden suhteelliset palkkaerot: Niiden analyysi ja tilastointi”, 2002. (in Finnish)

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. & Kurjenoja, J.: ”Naisten ja miesten palkkaero Suomen metalliteollisuudessa.” Studies of the Labour Institute for Economic Research, Helsinki, 1992. (in Finnish)

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.

Vartiainen, J.: ”Sukupuolten palkkaeron tilastointi ja analyysi”, Ministry of Social Affairs and Health, Gender equality series 2001:7, 2001. (in Finnish)

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 Jämför- barhet”, 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.

Annex 1: Stratification of tenancy agreements of offices premises into new and old agreements

Estimation areas	Municipality	Strata
Estimation area 1	Helsinki	Kluuvi
		Kaartinkaupunki
		Kamppi
		Sörnäinen
		Siltasaari
		Ruoholahti
		Vallila
		Lauttasaari
		Pitäjänmäki industrial area
		KruuPunaEiraUllaKatajaKaivo
		Etu- and Taka-Töölö
		Itä-, Länsi- and Pohjois-Pasila
		Ylä- and Ala-Malmi
		Herttoniemi industrial area and harbour
		Rest of Helsinki
Estimation area 2	Espoo	Otaniemi
		Tapiola
		Pohjois-Tapiola
		Leppävaara
		Kilo and Mankkaa
		Olari and Niittykumpu
		Rest of Espoo
	Kauniainen	Kauniainen
	Vantaa	Tikkurila and Viertola
		Tammisto, Pakkala, Veromies and Airport
		Myyrmäki and Martinlaakso
		Rest of Vantaa
Estimation area 3	Tampere	City centre
		Hervanta and Kauppi
		Tammela and Tulli
		Hatanpää and Lapinniemi
		Rest of Tampere
	Turku	City centre
		Kupittaa
		District 1
		Districts 8 and 9
		Rest of Turku
Estimation area 4	Jyväskylä	City centre
		Rest of Jyväskylä
	Kuopio	City centre
		Rest of Kuopio
	Lahti	City centre
		Rest of Lahti
	Oulu	City centre
		Rest of Oulu
Estimation area 5	Hämeenlinna	Hämeenlinna
	Joensuu	Joensuu
	Järvenpää	Järvenpää
	Kerava	Kerava
	Kouvola	Kouvola
	Lappeenranta	Lappeenranta
	Mikkeli	Mikkeli
	Pori	Pori
	Rovaniemi	Rovaniemi
	Vaasa	Vaasa
	Other municipalities	Other municipalities

Annex 2: Stratification of tenancy agreements of shop premises into new and old agreements

Estimation areas	Municipality	Strata
Estimation area 1	Helsinki	Kluuvi
		Kamppi and Kaartinkaupunki
		Kruununhaka, Punavuori, Eira, Ullanlinna, Katajanokka and Kaivopuisto
		Sörnäinen
		Itä-, Länsi- and Pohjois-Pasila
		Vallila
		Lauttasaari
		Linjat, Torpparinmäki, Alppila, Vallila and Hermanni
		Siltasaari and Ruoholahti
		Etu- and Taka-Töölö
		Meilahti and Ruskeasuo
		Vanhakaupunki and Herttoniemi industrial area
		Ylä- and Ala-Malmi
		Itäkeskus
		Rest of Helsinki
Estimation area 2	Espoo	Tapiola
		Olari,Mankkaa and Nöykkiö
		Espoonlahti
		Espoo centre
		Leppävaara and Matinkylä
		Rest of Espoo
	Kauniainen	Kauniainen
	Vantaa	Myyrmäki, Tammisto and Pakkala
		Tikkurila and Viertola
		Rest of Vantaa
Estimation area 3	Tampere	City centre
		Rest of Tampere
	Turku	City centre
		Itäkeskus (Varissuo)
		Teräsrautela
		Rest of Turku
Estimation area 4	Jyväskylä	City centre
		Rest of Jyväskylä
	Kuopio	City centre
		Rest of Kuopio
	Lahti	City centre
		Rest of Lahti
	Oulu	City centre
		Rest of Oulu
Estimation area 5	Hämeenlinna	City centre
		Rest of Hämeenlinna
	Joensuu	City centre
		Rest of Joensuu
	Järvenpää	Järvenpää
	Kajaani	Kajaani
	Kerava	Kerava
	Kokkola	Kokkola
	Kouvola	Kouvola
	Lappeenranta	Lappeenranta
	Mikkeli	Mikkeli
	Pori	City centre
		Rest of Pori
	Porvoo	Porvoo
	Raahe	Raahe
	Raisio	Raisio
	Rovaniemi	Rovaniemi
	Salo	Salo
	Savonlinna	Savonlinna
	Seinäjoki	Seinäjoki
	Vaasa	Vaasa
	Valkeakoski	Valkeakoski
	Other municipalities	Other municipalities

Annex 3: Estimation results for office premises in 1995/1-2001/2 (standard errors in brackets)
Brief description of estimation method in footnote of page seven.

Key figures/Point in time	1995/2	1996/1	1996/2	1997/1	1997/2	1998/1	1998/2	1999/1	1999/2	2000/1	2000/2	2001/1	2001/2
N	5354	5929	5999	6559	6942	7870	6838	7769	7770	8631	7562	8402	8065
Micro classes	109	98	97	99	109	108	108	98	111	99	110	98	110
Adj. R^2	0.426895	0.429819	0.421151	0.433939	0.470006	0.490966	0.522405	0.556201	0.571745	0.586475	0.607716	0.611371	0.61833
RMSE	0.309407	0.309175	0.308247	0.305671	0.29564	0.291719	0.283044	0.277785	0.267788	0.266979	0.275445	0.277968	0.274554
Constant	2.306092	2.328858	2.326821	2.389309	2.29619	2.342251	2.364772	2.392	2.566317	2.542243	2.428492	2.524138	2.54716
	(0.030655)	(0.029107)	(0.037119)	(0.033214)	(0.031621)	(0.027005)	(0.028509)	(0.028722)	(0.022136)	(0.021116)	(0.028239)	(0.026771)	(0.019046)
Floor area	-0.00002	-0.00001	-3.25E-6	5.532E-6	0.000024	0.000022	0.000025	0.000027	8.053E-6	4.777E-6	6.086E-8	1.019E-6	-0.00001
	(6.059E-6)	(5.922E-6)	(5.388E-6)	(5.181E-6)	(6.269E-6)	(5.949E-6)	(5.735E-6)	(5.308E-6)	(5.056E-6)	(4.589E-6)	(5.56E-6)	(5.306E-6)	(4.921E-6)
Floor area$^{1/2}$	0.004478	0.004259	0.003135	0.0023	0.000572	0.000737	0.000394	-0.00002	0.00043	0.000641	0.001394	0.001017	0.001986
	(0.000609)	(0.000586)	(0.000562)	(0.000536)	(0.000579)	(0.000545)	(0.000552)	(0.00051)	(0.00048)	(0.000446)	(0.000497)	(0.000475)	(0.000464)
Age at basic renovation	0.015118	0.014682	0.014681	0.017864	0.014171	0.01601	0.010804	0.009441	0.01931	0.016534	0.004245	0.008835	0.006277
	(0.002902)	(0.002768)	(0.003251)	(0.002869)	(0.002656)	(0.002322)	(0.002344)	(0.002358)	(0.001936)	(0.001865)	(0.002414)	(0.002303)	(0.001738)
Age at basic renovation$^{1/2}$	-0.14786	-0.15199	-0.14831	-0.173	-0.13342	-0.15173	-0.13194	-0.12372	-0.20119	-0.17824	-0.08548	-0.12185	-0.10761
	(0.018912)	(0.017982)	(0.021989)	(0.019388)	(0.018031)	(0.015422)	(0.015846)	(0.016055)	(0.012498)	(0.012025)	(0.016198)	(0.015359)	(0.010929)
$He(\alpha)$	1	1	1	1	1	1	1	1	1	1	1	1	1
	(0.018287)	(0.018391)	(0.015976)	(0.014664)	(0.01425)	(0.013805)	(0.014089)	(0.012708)	(0.012332)	(0.010632)	(0.009763)	(0.009)	(0.011081)
$He(x\beta)$	1	1	1	1	1	1	1	1	1	1	1	1	1
	(0.053312)	(0.037901)	(0.021919)	(0.023197)	(0.027037)	(0.022648)	(0.025923)	(0.034912)	(0.015389)	(0.016376)	(0.019267)	(0.028158)	(0.032453)
$\sum He(x\beta)/ N$ *(log-%)*	0.002246	0.003758	0.001047	0.001247	-0.0024	-0.00281	0.002184	0.002444	0.000816	0.002296	0.006326	0.006139	0.009251

Annex 4: Estimation results for shop premises in 1995/1-2001/2 (standard errors in brackets)
Brief description of estimation method in footnote of page seven.

Key figures/Point in time	1995/2	1996/1	1996/2	1997/1	1997/2	1998/1	1998/2	1999/1	1999/2	2000/1	2000/2	2001/1	2001/2
N	6936	7279	7359	7757	7680	8563	7908	8721	8536	9201	9268	9911	9139
Micro classes	121	98	100	98	121	124	123	100	123	100	123	100	124
Adj. R^2	0.400145	0.395269	0.405003	0.408051	0.416228	0.422911	0.420592	0.417787	0.429261	0.433588	0.424156	0.42564	0.434972
RMSE	0.498521	0.501553	0.496856	0.498953	0.498573	0.495411	0.502085	0.503984	0.504268	0.509142	0.514653	0.514961	0.516865
Constant	3.154563	3.136424	3.254549	3.365586	3.308814	3.423605	3.403994	3.465816	3.634543	3.573166	2.978676	2.998781	3.225125
	(0.046436)	(0.045756)	(0.065202)	(0.063402)	(0.077013)	(0.073114)	(0.072206)	(0.064035)	(0.061529)	(0.06084)	(0.060141)	(0.057432)	(0.062073)
Floor area	0.000175	0.000176	0.000178	0.000183	0.000211	0.000216	0.000223	0.000229	0.000197	0.000191	0.000158	0.00016	0.000163
	(0.000015)	(0.000015)	(0.000015)	(0.000014)	(0.000016)	(0.000015)	(0.000018)	(0.000018)	(0.000014)	(0.000013)	(0.000014)	(0.000013)	(0.000014)
Floor area$^{1/2}$	-0.01726	-0.01746	-0.01836	-0.0188	-0.01991	-0.01998	-0.02108	-0.02108	-0.0185	-0.01881	-0.01542	-0.01574	-0.01649
	(0.001179)	(0.001162)	(0.001143)	(0.001112)	(0.001184)	(0.001111)	(0.001245)	(0.001197)	(0.00109)	(0.001058)	(0.001074)	(0.001044)	(0.00108)
Age at basic renovation	0.032452	0.029547	0.037033	0.043911	0.044126	0.049965	0.048058	0.044537	0.058764	0.050186	0.008373	0.007152	0.014824
	(0.004198)	(0.004134)	(0.005437)	(0.005253)	(0.006088)	(0.005942)	(0.005475)	(0.004952)	(0.004708)	(0.004703)	(0.004736)	(0.004525)	(0.004795)
Age at basic renovation$^{1/2}$	-0.28115	-0.26346	-0.32332	-0.37792	-0.37743	-0.4313	-0.40775	-0.40589	-0.48656	-0.43004	-0.12803	-0.12413	-0.19743
	(0.027649)	(0.027276)	(0.037813)	(0.036634)	(0.043516)	(0.041734)	(0.039694)	(0.035451)	(0.03367)	(0.033405)	(0.033134)	(0.031583)	(0.033954)
$He(\alpha)$	1	1	1	1	1	1	1	1	1	1	1	1	1
	(0.016308)	(0.016213)	(0.01496)	(0.014299)	(0.014244)	(0.013361)	(0.013831)	(0.013432)	(0.013264)	(0.013022)	(0.014852)	(0.013887)	(0.012652)
$He(x\beta)$	1	1	1	1	1	1	1	1	1	1	1	1	1
	(0.019099)	(0.019035)	(0.020535)	(0.019452)	(0.015467)	(0.014714)	(0.016045)	(0.014831)	(0.016764)	(0.019455)	(0.029383)	(0.033513)	(0.033195)
$\sum He(x\beta)/N$ (*log-%*)	0.008667	0.008697	0.008969	0.009733	0.016505	0.013105	0.006568	0.005877	0.001203	0.00121	0.003643	0.000957	0.001679

Annex 5: Price aggregation and index calculation of semi-logarithmic models for heterogeneously behaving cross-sectional data estimated with the OLS method.

Let us examine equation (2.1) estimated with the OLS method

$$(1) \qquad \log(p_{ikt}) = \hat{\alpha}_{kt} + x'_{ikt}\hat{\beta}_{kt} + e_{ikt},$$

where sub-index *i* refers to an obsevation, *k* to stratum A_k, and *t* to a point in time. Coefficients $\hat{\alpha}_{kt}, \hat{\beta}_{kt}$ are OLS estimates of the equation's unknown parameters and e_{ikt} is its error term, or residual. We will first generalise aggregation of observations of equation (1) without precise specification of aggregation weights w_{ikt}, i.e.

$$(2) \qquad \sum_i w_{ikt} \log(p_{ikt}) = \sum_i w_{ikt}\left(\hat{\alpha}_{kt} + x'_{ikt}\hat{\beta}_{kt}\right) + \sum_i w_{ikt} e_{ikt}$$

where $\hat{\alpha}_{kt} = \log\left(\overline{p}^G_{kt}\right) - \overline{x}'_{kt}\hat{\beta}_{kt}$ is the OLS estimator of the constant term of a semi-logarithmic model, in which $\log\left(\overline{p}^G_{kt}\right) = 1/n_{kt}\sum_i \log(p_{ikt})$ and $\overline{x}'_{kt} = 1/n_{kt}\sum_i x'_{ikt}$. Thus, the first right-hand parenthetical expression (i.e. fit) of equation (2) can be written as (direction: place the explicit clause of the constant term into equation (2) and multiply each observation level equation from the left by 'weights' w_{ikt})

$$(3) \qquad \sum_i w_{ikt}\left(\hat{\alpha}_{kt} + x'_{ikt}\hat{\beta}_{kt}\right) = \sum_i w_{ikt} \log\left(\overline{p}^G_{kt}\right) - \sum_i w_{ikt}\overline{x}'_{kt}\hat{\beta}_{kt} + \sum_i w_{ikt} x'_{ikt}\hat{\beta}_{kt}$$

and the weighted sum of the residual respectively as

$$(4) \qquad \sum_i w_{ikt} e_{ikt} = \sum_i w_{ikt}\log(p_{ikt}) - \sum_i w_{ikt}\log\left(\overline{p}^G_{kt}\right) + \sum_i w_{ikt}\overline{x}'_{kt}\hat{\beta}_{kt} - \sum_i w_{ikt} x'_{ikt}\hat{\beta}_{kt}.$$

Presentation of equation (2) with two factors (3) and (4) seems to unnecessarily complicate the aggregation problem. Its unconventional mathematics add to the confusion. However, the stratification has a clear objective but its nature is difficult to see at the moment. Once the analysis has been sufficiently generalised, it is time to examine its special cases by simplifying it. In other words, we will next study a few suitable options for weights w_{ikt}. 'Standard textbook selection' falls upon weights $w_{ikt} = 1/n_{kt}, \forall\, a_i \in A_k$. Then the price model for a stratum can be presented as (See e.g. Oaxaca, 1973; Willis, 1986; Card, 1999; Vartiainen, 2001; Bayard, Hellerstein and Troske, 2003; Korkeamäki and Kyyrä, 2002; Koev, 2003; Korkeamäki and Kyyrä, 2003; Korkeamäki, Kyyrä and Luukkonen, 2004)

$$(5) \qquad \log\left(\overline{p}^G_{kt}\right) = \hat{\alpha}_{kt} + \overline{x}'_{kt}\hat{\beta}_{kt},$$

where exogenous input variables are unweighted arithmetic means $\overline{x}'_{kt} = 1/n_{kt}\sum_i x'_{ikt}$, the output variable is a logarithmic unweighted geometric mean and the parameters are ordinary OLS estimates. The selection of weights $w_{ikt} = 1/n_{kt}$, $\forall\, a_i \in A_k$ is based on the standard OLS method in which one requirement is that residuals sum up to zero. Selection $w_{ikt} = 1/n_{kt}$, $\forall\, a_i \in A_k$ meets this requirement trivially. This requirement is usually not met with freely selectable weights as it does not even need to be met - meeting it in OLS estimation is sufficient. In other words, it does not prevent the use of other aggregation weights. Let us next examine weights $w_{ikt} = w^{Gw}_{ikt} = q_{ikt}\big/\sum_i q_{ikt}$ in which q_{ikt} connotes quantities at the observation level and its sum respectively their total for the strata level. Aggregation of equations (3) and (4) then leads to the price model of stratum A_k which can be expressed as

(6) $$\log\left(\overline{p}_{kt}^{Gw}\right) = \hat{\alpha}_{kt}^{Gw} + \overline{x}_{kt}^{\prime Gw}\hat{\beta}_{kt},$$

where $\log\left(\overline{p}_{kt}^{Gw}\right) = \sum_i q_{ikt}\log\left(p_{ikt}\right) / \sum_i q_{ikt}$ is the logarithmic weighted geometric mean, the constant term $\hat{\alpha}_{kt}^{Gw} = \log\left(\overline{p}_{kt}^{Gw}\right) - \overline{x}_{kt}^{\prime Gw}\hat{\beta}_{kt}$, in which $\overline{x}_{kt}^{\prime Gw} = \sum_i q_{ikt} x_{ikt}^{\prime Gw} / \sum_i q_{ikt}$.

Two other so-called elementary aggregate models for the strata are obtained analogously with weights (3.5a) and (3.5b) deduced in Chapter 3. For instance, with observation weights $w_{ikt} = w_{ikt}^{A} = L(p_{ikt}, 1) / L\left(\sum_{\mathrm{i}} p_{ikt}, n_{kt}\right)$ of equation (3.5b) the aggregation of equations (3) and (4) leads to

(7) $$\log\left(\overline{p}_{kt}^{A}\right) = \hat{\alpha}_{kt}^{A} + \overline{x}_{kt}^{\prime A}\hat{\beta}_{kt},$$

where $\log\left(\overline{p}_{kt}^{A}\right) = \sum_i w_{ikt}^{A}\log\left(p_{ikt}\right)$ is the logarithmic unweighted geometric mean, the constant term $\hat{\alpha}_{kt}^{A} = \log\left(\overline{p}_{kt}^{A}\right) - \overline{x}_{kt}^{\prime A}\hat{\beta}_{kt}$, in which $\overline{x}_{kt}^{\prime A} = \sum_i w_{ikt}^{A} x_{ikt}^{\prime}$. With weights $w_{ikt} = w_{ikt}^{Aw} = L(q_{ikt}p_{ikt}, q_{ikt}) / L\left(\sum_{\mathrm{i}} q_{ikt}p_{ikt}, \sum_{\mathrm{i}} q_{ikt}\right)$ (see Chapter 3) we obtain analogously

(8) $$\log\left(\overline{p}_{kt}^{Aw}\right) = \hat{\alpha}_{kt}^{Aw} + \overline{x}_{kt}^{\prime Aw}\hat{\beta}_{kt},$$

where $\log\left(\overline{p}_{kt}^{Aw}\right) = \sum_i w_{ikt}^{Aw}\log\left(p_{ikt}\right)$ is the logarithm of the unweighted geometric mean, the constant term $\hat{\alpha}_{kt}^{Aw} = \log\left(\overline{p}_{kt}^{Aw}\right) - \overline{x}_{kt}^{\prime Aw}\hat{\beta}_{kt}$, in which $\overline{x}_{kt}^{\prime Aw} = \sum_i w_{ikt}^{Aw} x_{ikt}^{\prime}$.

Elementary aggregate model (5) deduced from specification (1) is familiar to anyone having studied the basics of statistical science because is follows direct from the OLS method (or alternatively from the GLS method). Models (6), (7) and (8), by contrast, are not familiar because as far as we know they have never been presented before in a parametric manner in a specification (1) situation. The mathematics of these ‘average’ stratum level price models is necessary - index calculation is almost always based on either unweighted (aggregation of unit prices) or on weighted geometric or arithmetic means. Let us go on to examine differences between these methods and their index applications.

In Chapter 3 we deduced the difference between a weighted arithmetic and an unweighted geometric mean price for logarithmic prices (See Chapter 3 equation (3.6)). The difference between these parameters is obtained trivially in a case of specification (1) when we subtract equation (5) from equation (8).

(9) $$\sum_i \left(w_{ikt}^{Aw} - n_{\mathrm{kt}}^{-1}\right)\log(p_{ikt}) \equiv \hat{\alpha}_{kt}^{Aw} - \hat{\alpha}_{kt} + \left(\overline{x}_{kt}^{\prime Aw} - \overline{x}_{kt}^{\prime}\right)\hat{\beta}_{kt}$$

A respective simple parametric presentation can also be used to calculate differences between other statistics. The differences between the models can naturally be used to calculate differences of relative changes (index) in the respective statistics.

We will finally examine index calculation with stratum price models as a parametric presentation. The perspective is entirely new because as far as we know indices aggregated from the micro index level have never before been presented with paramerisized price models. We will free the analysis from choice of index formula and present index weights generally as $w_{kt}^{*}, k = 1, \ldots, K$. We will observe the analysis in the Annex and aggregate price models (5) over all strata (instruction: place the constant term estimator into equation (7) and use the basic aggregation clause in the footnote of page 6 (Vartia, 1979)).

$$\sum_k w_{kt}^* \log\left(\overline{p}_{kt}^G\right) = \sum_k w_{kt}^* \hat{\alpha}_{kt} + \sum_k w_{kt}^* \overline{x}_{kt}' \hat{\beta}_{kt} \Leftrightarrow$$

$$\log\left(\overline{p}_t^G\right) = \log\left(\overline{p}_t^G\right) - \overline{x}_t' \hat{\hat{\beta}}_t - \sum_k w_{kt}^* \overline{x}_{kt}' \left(\hat{\beta}_{kt} - \hat{\hat{\beta}}_t\right) + \overline{x}_t' \hat{\hat{\beta}}_t + \sum_k w_{kt}^* \overline{x}_{kt}' \left(\hat{\beta}_{kt} - \hat{\hat{\beta}}_t\right) \Leftrightarrow$$

$$\log\left(\overline{p}_t^G\right) = \hat{\hat{\alpha}}_t + \overline{x}_t' \hat{\hat{\beta}}_t, \tag{10}$$

where the constant term $\hat{\hat{\alpha}}_t = \log\left(\overline{p}_t^G\right) - \overline{x}_t' \hat{\hat{\beta}}_t$, $\log\left(\overline{p}_t^G\right)$ is weighted geometric mean (in log-scale) and $\hat{\hat{\alpha}}_t, \overline{x}_t', \hat{\hat{\beta}}_t$, respectively, are means calculated with index weights - they are 'population parameters' and are calculated as weighted means of stratum level parameters. When index weights sum up to one, $\sum_k w_{kt}^* \overline{x}_{kt}' \left(\hat{\beta}_{kt} - \hat{\hat{\beta}}_t\right)$ equals precisely weighted covariance $\text{cov}\left(\overline{x}_t', \hat{\hat{\beta}}_t\right)$ between input variables and their parameter equivalents.

Let us examine a few index weight options. When we select base period value share weights $w_{k\tau}^{Koev} = v_{k\tau} \big/ \sum_k v_{k\tau}$ (Koev, 2003, p. 26) for strata A_k and aggregate prices models (5) over the strata, Koev's Log-Laspeyres' received the following expression (the last phase is based on the co-called Oaxaca decomposition (Oaxaca, 1973))

$$\sum_k w_{k\tau}^{Koev} \log\left(\frac{\overline{p}_{kt}^G}{\overline{p}_{k\tau}^G}\right) = \log\left(\frac{\overline{p}_t^G}{\overline{p}_\tau^G}\right) = \hat{\hat{\alpha}}_t + \overline{x}_t' \hat{\hat{\beta}}_t - \hat{\hat{\alpha}}_\tau - \overline{x}_\tau' \hat{\hat{\beta}}_\tau = \left(\overline{x}_t' - \overline{x}_\tau'\right) \hat{\hat{\beta}}_\tau + \overline{x}_t' \left(\hat{\hat{\beta}}_t - \hat{\hat{\beta}}_\tau\right) + \left(\hat{\hat{\alpha}}_t - \hat{\hat{\alpha}}_\tau\right) \tag{11}$$

whose exp transformation produces Koev's index of dwelling prices

$$\frac{\overline{p}_t^G}{\overline{p}_\tau^G} = \exp\left\{\left(\overline{x}_t' - \overline{x}_\tau'\right) \hat{\hat{\beta}}_\tau\right\} \exp\left\{\left(\hat{\hat{\alpha}}_t + \overline{x}_t' \hat{\hat{\beta}}_t\right) - \left(\hat{\hat{\alpha}}_\tau + \overline{x}_t' \hat{\hat{\beta}}_\tau\right)\right\} \tag{12}$$

The left-hand side of the index expresses change in geometric mean prices weighted with fixed value shares for the base period (classic Log-Laspeyres' classification index), the first right-hand term describes the price ratio due to quality difference at base period valuation of the charasteristics and the last term expresses change in prices standardised for quality. As product of the difference between the quality adjustment and the index standardised for quality we obtain the precise change in geometric mean prices. Thus index (12) expresses the following: Some of the change in geometric mean prices is explained by quality change in characteristics and the remainder by price change standardised for quality (i.e. qualitative characteristics at base period). J. van Dalen and B. Bode (2004) shows in the paper 'Estimation Biases in Quality-Adjusted Hedonic Price Indices', that for the log-transformed multiplicative hedonic regressions (12) is biased index formula. The next example is not.

In the second example we examine the logarithm of Laspeyres' index (Vartia, 1976, p. 126)

$$\log\left(\frac{\sum_k q_{k\tau} \overline{p}_{kt}^{Aw}}{\sum_k q_{k\tau} \overline{p}_{k\tau}^{Aw}}\right) = \sum_k w_{k\tau}^L \log\left(\frac{\overline{p}_{kt}^{Aw}}{\overline{p}_{k\tau}^{Aw}}\right), \text{ where } w_{k\tau}^L = \frac{L\left(q_{k\tau}\overline{p}_{kt}^{Aw}, q_{k\tau}\overline{p}_{k\tau}^{Aw}\right)}{L\left(\sum_k q_{k\tau}\overline{p}_{kt}^{Aw}, \sum_k q_{k\tau}\overline{p}_{k\tau}^{Aw}\right)} \tag{13}$$

By locating the estimator of the constant term in price model (8) and aggregating the differentiated price models by stratum thus obtained over a stratum we receive (with the help of the basic aggregation clause (Vartia, 1979) and Oaxaca decomposition (Oaxaca, 1973))

$$\sum_k w_{k\tau}^L \log\left(\frac{\overline{p}_{kt}^{Aw}}{\overline{p}_{k\tau}^{Aw}}\right) = \log\left(\frac{\overline{p}_t^{Aw}}{\overline{p}_\tau^{Aw}}\right) = \left\{\left(\overline{x}_t'^{Aw} - \overline{x}_\tau'^{Aw}\right) \hat{\hat{\beta}}_\tau^{Aw} + \overline{x}_t'^{Aw} \left(\hat{\hat{\beta}}_t^{Aw} - \hat{\hat{\beta}}_\tau^{Aw}\right) + \left(\hat{\hat{\alpha}}_t^{Aw} - \hat{\hat{\alpha}}_\tau^{Aw}\right)\right\} \tag{14}$$

With exp transformation of the equation we obtain a parametric presentation of Laspeyres price index

$$(15) \qquad \frac{\sum_k q_{k\tau}\overline{p}_{kt}^{Aw}}{\sum_k q_{k\tau}\overline{p}_{k\tau}^{Aw}} = \exp\left\{\left(\overline{x}_t'^{Aw} - \overline{x}_\tau'^{Aw}\right)\hat{\hat{\beta}}_\tau^{Aw}\right\}\exp\left\{\left(\hat{\hat{\alpha}}_t^{Aw} + \overline{x}_t'^{Aw}\hat{\hat{\beta}}_t^{Aw}\right) - \left(\hat{\hat{\alpha}}_\tau^{Aw} + \overline{x}_t'^{Aw}\hat{\hat{\beta}}_\tau^{Aw}\right)\right\}$$

in which all parameters $\hat{\hat{\beta}}_\tau^{Aw}$ and $\hat{\hat{\beta}}_t^{Aw}$, input and output variables are weighted means calculated with weights $w_{k\tau}^L, k = 1,...,K$ defined in equation (13). The constant term is defined as $\hat{\hat{\alpha}}_t^{Aw} = \log\left(\overline{p}_t^{Aw}\right) - \overline{x}_t'^{Aw}\hat{\hat{\beta}}_t^{Aw}$. Equation (15) represents unbiased Laspeyres index formula for the log-transformed multiplicative hedonic regressions developed for are weighted aritmetic averages and expressed in parmetric form. Similar result naturally can be deduced for unweighted aritmetic averages (see, eq. (7) and Tables 5.10a and Tables 5.10b).

Table 5.10a: Price aggregation of a semi-logarithmic model from observation level to stratum level for OLS estimated behaviour (so-called elementary aggregate)

	Price aggregation weights	'Elementary aggregate'
Unweighted geometric	$w_{ikt} = 1/n_{kt}$	$\log\left(\overline{p}_{kt}^G\right) = \hat{\alpha}_{kt} + \overline{x}_{kt}'\hat{\beta}_{kt}$, defined in equations (2) and (5)
Weighted geometric	$w_{ikt} = w_{ikt}^{Gw} = q_{ikt}/\sum_i q_{ikt}$	$\log\left(\overline{p}_{kt}^{Gw}\right) = \hat{\alpha}_{kt}^{Gw} + \overline{x}_{kt}'^{Gw}\hat{\beta}_{kt}$ defined in equation (6)
Unweighted arithmetic	$w_{ikt} = w_{ikt}^A = L(p_{ikt},1)/L\left(\sum_i p_{ikt}, n_{kt}\right)$	$\log\left(\overline{p}_{kt}^A\right) = \hat{\alpha}_{kt}^A + \overline{x}_{kt}'^A\hat{\beta}_{kt}$ defined in equation (7)
Weighted arithmetic	$w_{ikt} = w_{ikt}^{Aw} = L(q_{ikt}p_{ikt}, q_{ikt})/L\left(\sum_i q_{ikt}p_{ikt}, \sum_i q_{ikt}\right)$	$\log\left(\overline{p}_{kt}^{Aw}\right) = \hat{\alpha}_{kt}^{Aw} + \overline{x}_{kt}'^{Aw}\hat{\beta}_{kt}$ defined in equation (8)

Table 5.10b: Summary of hedonic indices of a semi-logarithmic price model in OLS optimum between time periods τ, t

	Price index	Index formula as parametric expression: aggregated from differences between elementary aggregate models at stratum level: See Oaxaca decomposition and basic aggregation clause equation (10).
Unweighted geometric	Log-Laspeyres, see eq. (12)	$\frac{\overline{p}_t^G}{\overline{p}_\tau^G} = \exp\left\{(\overline{x}_t' - \overline{x}_\tau')\hat{\hat{\beta}}_\tau\right\}\exp\left\{\left(\hat{\hat{\alpha}}_t + \overline{x}_t'\hat{\hat{\beta}}_t\right) - \left(\hat{\hat{\alpha}}_\tau + \overline{x}_t'\hat{\hat{\beta}}_\tau\right)\right\}$
Weighted geometric	Log-Laspeyres, analogy eq. (12) (12)	$\frac{\overline{p}_t^{Gw}}{\overline{p}_\tau^{Gw}} = \exp\left\{\left(\overline{x}_t'^{Gw} - \overline{x}_\tau'^{Gw}\right)\hat{\hat{\beta}}_\tau^{Gw}\right\}\exp\left\{\left(\hat{\hat{\alpha}}_t^{Gw} + \overline{x}_t'^{Gw}\hat{\hat{\beta}}_t^{Gw}\right) - \left(\hat{\hat{\alpha}}_\tau^{Gw} + \overline{x}_\tau'^{Gw}\hat{\hat{\beta}}_\tau^{Gw}\right)\right\}$
Unweighted arithmetic	Laspeyres, analogy eq. (12) (12)	$\frac{\sum_k q_{k\tau}\overline{p}_{kt}^A}{\sum_k q_{k\tau}\overline{p}_{k\tau}^A} = \exp\left\{\left(\overline{x}_t'^A - \overline{x}_\tau'^A\right)\hat{\hat{\beta}}_\tau^A\right\}\exp\left\{\left(\hat{\hat{\alpha}}_t^A + \overline{x}_t'^A\hat{\hat{\beta}}_t^A\right) - \left(\hat{\hat{\alpha}}_\tau^A + \overline{x}_t'^A\hat{\hat{\beta}}_\tau^A\right)\right\}$
Weighted arithmetic	Laspeyres, see eq. (15)	$\frac{\sum_k q_{k\tau}\overline{p}_{kt}^{Aw}}{\sum_k q_{k\tau}\overline{p}_{k\tau}^{Aw}} = \exp\left\{\left(\overline{x}_t'^{Aw} - \overline{x}_\tau'^{Aw}\right)\hat{\hat{\beta}}_\tau^{Aw}\right\}\exp\left\{\left(\hat{\hat{\alpha}}_t^{Aw} + \overline{x}_t'^{Aw}\hat{\hat{\beta}}_t^{Aw}\right) - \left(\hat{\hat{\alpha}}_\tau^{Aw} + \overline{x}_t'^{Aw}\hat{\hat{\beta}}_\tau^{Aw}\right)\right\}$