Combining Classification and Hedonic Quality Adjustment in Constructing a House Price Index

Eugen Koev*
Licentiate Thesis
Department of Economics
University of Helsinki
January, 2003

*Eugen Koev is currently economist at the Confederation of Unions for Academic Professionals in Finland (AKAVA), eugen.koev@akava.fi, tel. +358 9 150 2583, fax +358 9 150 2589

Abstract

Statistics Finland has relatively long experience in constructing indices of prices of old flats using both classification and time-dummy hedonic approaches. Each method has proved to have drawbacks. The feasible classification may be too rigid to capture relevant quality changes and the standard time-dummy hedonic approach is not easily interpretable in the context of traditional index number theory. To overcome the disadvantages of these methods the two approaches are combined. A hedonic-method quality adjustment is performed within each cell in a classification and then the index is computed by aggregating cell level quality adjusted prices using an index number formula. It is shown, that each step of the procedure has a very close analogue in the standard practice of statistical offices. A method for evaluating the aggregate contributions of the different characteristics on the overall quality adjustment is developed in order to make the hedonic method more transparent in the context of classical index number theory. Special attention in the discussion is paid to the interpretation of the age profile of house prices, since they are a mixture of two distinct, but empirically non-separable effects having different implication for the quality adjustment.

The method is applied for estimating quarterly indices for Finland during 1987-2000 using very large high-quality register data on all free-market transactions of dwellings in old blocks of flats and terraced houses. It turned out that the quality adjustment of the index was important in evaluation of short-term price movements on thin markets. In the long run the quality adjusted index series do not differ in any substantial way from the simple price averages trend, at least at aggregate levels. The reliability of the index is evaluated by simulation.

The method described in the paper is used in the Finnish official House Price Index 2000=100.

JEL Classification System: C43, E31, R31

Introduction

Most papers related to research in the field of price measurement are, unfortunately, disregarded by official statistics as too complicated and many academic researchers frown at official statistics for their rather unsophisticated nature. This paper is based on past experiences and current research whose results are used in the construction of the quarterly house price index published by Statistics Finland. Hopefully the topic will keep practitioners interested and at the same time the argumentation will satisfy academic readers as well. The discussion of methods used in compiling house price indices is currently particularly relevant, since issues related to prices of owner occupied houses have started to attract attention as it seems likely that housing expenses in owner occupied housing will be included in the European Union's Harmonised Consumer Price Index (HICP) via an index of prices of new dwellings.

As usually in the case of complex goods, the main concern in constructing a house price index is the quality adjustment method and regression methods for quality adjustment have a central place in our discussion. In the context of real estate and house price indices research in the field has been very much influenced by the fact that in the long run more than one sale of a particular dwelling will be observed. The first chapter of the paper is an overview of the mainstream methodology for constructing real estate indices. It covers the repeat-sales regression method suggested by Bailey, Muth and Nourse (1963) and further developed by Case and Shiller (1987) as well as the modern so-called hybrid-type models, for which an already classical reference is Quigley (1995). The chapter contains a discussions of the different assumptions behind these methodologies and points out, that all of them strongly rely on time-invariance of the regression parameters, an assumption which is dubious for long time spans.

The second chapter of the paper is a history of the House price index published by Statistics Finland. The standard classification method and the time-dummy hedonic method used in the past are presented and their shortcomings are discussed from the point of view of official statistics. The author's perception is that a very important concern of statistical agencies is how to link a regression based quality adjustment to the paradigm of classical index number construction methods in a way that ensures good transparency of the results. In chapter three an index construction method based on classification and within-class hedonic quality adjustment is proposed. To enhance the transparency of the procedure a method for evaluating the size of the overall quality adjustment and decomposing it into factors attributable to the different characteristics included in the hedonic regressions is developed.

Chapter four presents the data source, the empirical models and the regression estimation results. A section on the interpretation of the age profile of house prices is included. It turns out to be very important to distinguish between vintage and depreciation effects, which are both captured in the age profile. A method aiming to separate the two proposed by Englund, Quigley and Redfearn (1998) seems to be based on inappropriate identification of the regression coefficients. The author concludes that without outside sample information it is not possible to separate the two effects. It is argued that in the Finnish case interpreting the age profile as reflecting vintage rather than depreciation effect is the less erroneous assumption.

In chapter five the estimated quarterly indices for 1987-2000 are discussed. Data on all transactions of dwellings in housing share corporations provided by the Finnish tax authorities is used. The impact of the quality adjustment procedure is evaluated against the benchmark case in which sample characteristics quality differences are controlled for only by classifying the data. The reliability of the index is evaluated by a simulation method.

It is concluded, that the developed method is applicable generally in cases requiring complicated quality adjustment, such as wage indices.

1. Hedonic Indices for Real Estate Prices: an Overview

Hedonic indices are based on estimating the price relationship between the qualitative characteristics of a complex commodity and its market price. Theoretical foundations of the hedonic method are provided already by Rosen (1974) and Triplett (1983). In the context of housing prices the estimation of hedonic indices and the discussion around them differs in certain aspects from the discussion in the other typical application field of the technique, namely indices for consumer durable goods such as cars and computers. This is due to the differences in the nature of the markets and the data generating processes.

In the case of durable goods the use of hedonic technique is viewed as a means of accounting for the rapid technological quality improvement and the perpetual change of the durable goods on the market. In the case of housing price (or in general real estate) indices the need for regression modelling arise from the nature of the dwellings as goods and the specific features of the housing markets rather than rapid technological change.

House prices react quickly to changes in the economic environment and are considered an important economic indicator, so a real estate price index should be produced at least on a quarterly basis, many users would prefer a monthly one. At the same time the number of real estate sales over a short time period present only a very small fraction of the stock and not even all of them may be available for compiling the index. Each dwelling is endowed with an (almost) unique set of characteristics which determine price, giving rise to very large cross-section price differences. Under these circumstances it is very likely that observed price differences in different time periods as measured by simple statistics such as mean- or median change, or an index based on rough stratification of the dwellings will reflect changes in the quality mix of the dwellings. It is also possible that the quality mix of dwellings is systematically different in economic upturns and downturns, demographic changes may also shift sales towards dwellings endowed with some particular characteristic. One might expect an increasing transaction share in economically vivid regions with growing population. Such phenomena also render price measurement based on simple techniques potentially imprecise and even misleading.

Empirical research on regression methods for property price indices is grossly influenced by the possibility to observe the sales price of the same property (dwelling) at different time periods. In data sets covering a long time span some dwellings are sold more than ones and thus a subset of the data forms a panel. This particular feature of data on property prices has been recognised already in the very early research in the field and efficient utilisation of the panel information has become a major part of the mainstream research programme. In the following we try to summarise the theory.

1.1 The Repeat Sales Model

To introduce the notation, suppose, that the price of a dwelling is determined by the following rather general specification:

$$(1.1) \qquad p_{it} = \alpha_t + \boldsymbol{\beta}_t' \mathbf{x}_{it} + \zeta_{it} + \varepsilon_{it}$$

The subscript i refers to a specific dwelling and t to the time period of the sale. p_{it} is usually the log-price of the property, but it may be some other transformation of the price, $\mathbf{x}_{it}$ are the observed characteristics, α_t is an unknown constant and $\boldsymbol{\beta}_t$ is an unknown parameter vector. ζ_{it} is a dwelling-specific term reflecting idiosyncratic price effects and ε_{it} is a statistical error. Without loss of generality the normalisation $E(\zeta_{it})=0$ can be assumed. Empirical models in the literature are derived by assuming a specific structure of the error term and possibly restricting the parameter vector.

In their pioneering work Bailey, Muth and Nourse (1963) rely completely on a sample of repeat-sales. Suppose that a property is sold in periods τ and t, $\tau > t$. Assume that the property has remained in every respect the same between the two periods, i.e. both its observable and unobservable characteristics have not changed. Assume further that there is no change in the relative market valuation of the characteristics. These assumptions imply that $\mathbf{x}_{i\tau} = \mathbf{x}_{it}$, $\zeta_{i\tau} = \zeta_{it}$ and $\boldsymbol{\beta}_\tau = \boldsymbol{\beta}_t$. In this case the price difference of the dwelling i between τ and t is:

$$(1.2) \qquad p_{i\tau} - p_{it} = (\alpha_\tau - \alpha_t) + (\varepsilon_{i\tau} - \varepsilon_{it})$$

This extremely simple model can be estimated by ordinary least squares (OLS) under the assumption that $(\varepsilon_{it} - \varepsilon_{it})$ are independent with zero mean and constant variance. By appropriately normalising the coefficient for the initial period α_0, one gets the desired index directly from the coefficients α_t.

A more sophisticated variation of (1.2) proposed by Case and Shiller (1989) has become a benchmark case in the literature. Case and Shiller argue, that the property-specific value captured in the term ζ_{it} in equation (1.1) experiences random shocks over time:

$$(1.3) \qquad \zeta_{it} = \zeta_{i(t-1)} + v_{it} = \zeta_i + \sum_{j=1}^{t} v_{ij}$$

where ν_{it} is white noise and hence ζ_{it} is a random walk. Under these assumptions (1.2) becomes

$$(1.2') \qquad p_{i\tau} - p_{it} = (\alpha_\tau - \alpha_t) + (\nu_{i\tau} - \nu_{it}) + (\varepsilon_{i\tau} - \varepsilon_{it})$$

In the specification (1.2') the error term has two components and is heteroscedastic, since the variance of $(\nu_{i\tau} - \nu_{it})$ depends on the length of the interval (τ-t). Case and Shiller develop an obvious feasible generalised least squares (GLS) estimator, in which the squared OLS residuals of (1.2') are used in an auxiliary regression on a constant term and the interval (τ-t) to obtain estimates of the variances of ν_{it} and ε_{it}. These are then used to form the appropriate weight-matrix and re-estimate (1.2') by GLS.

The appeal of the repeat-sales model is that there is no need to know anything about the price-characteristics relationship, such as what are the relevant characteristics and the correct functional form. In this sense specifications (1.2) and (1.2') are robust. However, it assumes that the price-characteristic relation, whatever it is, is constant over time and that all characteristics of the properties are unchanged[1]. A very serious problem with the procedure is that it is extremely wasteful in terms of observations, since price information for properties whose selling price is observed only once is simply thrown away. An even more serious objection to the method is that properties sold more than once within a short time period are likely to be a non-random and non-representative sub-sample of all sales and hence the estimated index will be biased. These drawbacks of the method have been recognised in the 90's and have lead to much empirical work, whose purpose has been to develop procedures, which recognise that a subset of the data forms a panel (multiple sales of the same property), but makes use of the information of properties sold only once. Such methods are normally referred to in the literature as "hybrid".

1.2 The "Hybrid Models"

The insight of the "hybrid" models is that the panel of properties sold more than once may provide, under certain assumptions, the possibility for more efficient estimation than simply a pooled OLS procedure will do. Assume that $\boldsymbol{\beta}_t$ is time invariable, i.e. $\boldsymbol{\beta}_t=\boldsymbol{\beta}$. Then equation (1.1) becomes

$$(1.4) \qquad p_{it} = \alpha_t + \beta'\mathbf{x}_{it} + \zeta_{it} + \varepsilon_{it}$$

Under standard assumptions for the structure of the composite error term $\zeta_{it} + \varepsilon_{it}$, pooling all observations and using OLS will provide consistent estimates of the regression parameters. However by recognising the panel structure of the repeat-sales subset of the data, estimation can be improved in certain cases. The seminal paper of Quigley (1995) shows how the repeat-sales can be used first, to distinguish between effects of observed characteristics in the vector $\mathbf{x}_{it}$ and unobserved ones captured in the idiosyncratic term ζ_{it}, and second, to improve estimation efficiency by appropriate GLS procedure. Following Case and Shiller, Quigley (1995) assumes that the idiosyncratic term is described by (1.3), that is $\zeta_{it} = \zeta_i + \sum_{j=1}^{t} v_{ij}$ where ν_{ij} is essentially a white noise[2]. Then by re-arranging the error terms of (1.4) and defining $\eta_{it} = \sum_{j=1}^{t} v_{ij} + \varepsilon_{it}$ one obtains

$$(1.4') \qquad p_{it} = \alpha_t + \beta'\mathbf{x}_{it} + \zeta_i + \eta_{it}$$

The following structural assumptions are made: $E(\zeta_i)=0$, $E(\eta_{it})=0$, $E(\eta_{it}\,\eta_{j\tau})=0$, $E(\zeta_i\zeta_j)=0$, $E(\zeta_i\eta_{jt})=0$, $E(\zeta_i)^2=\sigma^2_\zeta$, $E(\eta_{it})^2=\sigma^2_\eta$ for all $i\neq j$ and $t\neq \tau$ and $E(\eta_{it}-\eta_{i\tau})^2$ is a quadratic polynomial of $(t-\tau)$. The proposed estimation procedure is the following. First

---

[1] In research based on "pure" repeat-sales methods this assumption is normally checked and only observations satisfying it are selected from the data

[2] I slightly re-interpret Quigley's (1995) presentation and don't go into full detail, he assumes for example a more complicated structure of the variance of the term ν_t, but this is not relevant for the discussion.

(1.4') is estimated using the sample of repeat-sales. The idiosyncratic effects are explicitly modelled by including property-specific dummies. The squared residuals of this model are then used to estimate σ^2_η and the time structure of $E(\eta_{it}-\eta_{i\tau})^2$. Then the model is estimated using again the same repeat-sales subset, but this time without the property-specific dummies. The squared residuals of this model together with the estimates of σ^2_η and the time-structure of $E(\varepsilon_{it}-\varepsilon_{i\tau})^2$ from the previous step are used to estimate σ^2_ζ. The structure of the variance-covariance matrix of (1.4') is now completely identified and in the final step all observations are used to estimate the equation by GLS.

1.3 Discussion

Different estimators are consistent and efficient under different conditions and choice of the most appropriate one is largely dependent on what particular problems the researchers think are of primary importance for the phenomena under study in general and for the particular data set used. Repeat-sales methods avoid a lot of modelling uncertainty related to omission of relevant explanatory variables and functional specification of the regression. Although bias due to non-randomness of the repeat-sales is recognised as a possibility by Quigley (1995), the hybrid method is primarily concerned with efficient use of all sample information and not with sample selection bias. This is because if the subset of repeat-sales in a data is a result of a non-random process, then the estimation results of the first step of Quigley's procedure, which relies on repeat-sales, are potentially biased and it is not self-evident whether the overall estimation is consistent. However, efficiency is achieved by imposing explicit structure of the regression function, a feature which pure repeat-sales methods avoid.

There are some other aspects of the hybrid model worth discussing. The procedure can be interpreted as an estimation of an unbalanced random-effect panel model. The central assumption of such models is that the two error terms ζ_i and η_{it} are both non-correlated with the observable variables The efficiency gain of the hybrid model estimator over either pooled regression or repeat-sales models comes from utilisation of both between-unit and within-unit price variability. Pooled regression utilises only between-group and repeat-sales only within-group variability (see Greene,1997, 618-620). Knowing this fact, it is not surprising that a normal empirical finding is that the hedonic index based on hybrid models has narrower confidence intervals as compared to estimates obtained from either pooled regression or repeat-sales.

Estimation of both the repeat-sales and the Quigley's hybrid model heavily relies on the assumption that the relative valuation of the different characteristics of the property are constant over time, implying time-constant parameter β. This assumption is very restrictive for analysing data collected over 10-years long time period as is the case e.g. in Case and Shiller (1979), Quigley (1995), Englund, Quigley and Redfearn (1998 and 1999). Meese and Wallace (1997) test different assumptions of the repeat-sales and the hybrid model. One conclusion is that estimates of β based on either repeat-sales or single sales are statistically different. The result is interpreted by the authors as supporting the hypothesis that repeat-sales prices are unrepresentative. The hypothesis of time-constant β is also rejected in this research based on data from two cities in 1970-1988. The authors suggest as an alternative a non-parametric regression specification allowing for time-variable parameters.

As obvious, there is not necessarily a best solution, since under different assumptions different procedure may be best. The author's opinion is that arguments against use of pure repeat-sales models are compelling. For time-periods of several years the assumption of parameter constancy in the case of houses is reasonable, so a hybrid-model of the type that Quigley (1995) suggests might be considered a good solution for short index series. However, the efficiency gain of utilising the repeat-sales information in estimation of the index is proportional to the fraction of repeat-sales in the overall data. The shorter the time span over which estimation is performed, the smaller the repeat-sales fraction and the smaller the benefit. Moreover, repeat-sales over short time periods are most likely subject to severe selection biases (in the very long run any house will be eventually sold more than once). For long time periods the hypothesis of parameter constancy has no a priori grounds, and as the work of Meese and Wallace (1997) shows, is likely to be violated. This considerations suggest that research in the field should perhaps be focused more on flexibility and robustness of the proposed methods rather than on their relative efficiency under perhaps unduly stringent assumptions.

From the point of view of official statistical agencies, there are several aspects to be considered when interpreting the research results and their suggestions. This requires some clarification of the language and the normal practice of statistical offices. The starting point of classical price index theory is that there exists a complete list of all goods and their unit prices. The items in that list are the same at all times. To make the concept operational, statistical offices would devise a classification of the goods into homogenous classes, so that goods within each class are (presumably) close substitutes. Today the classifications are increasingly co-ordinated at international level. Then from within each class one or several precisely specified items will be selected and their unit prices will be followed e.g. each month. The choice is normally made on some notion of representativeness of the items such as market share in the class. The observed prices of an item in the base and comparison periods form a “matched pair”. The precise specification of the items to be followed ensures, that the matched pair unit-price ratio reflects only real price changes and not quality differences. If more than one item from a class is selected, then the price change estimate for the class is obtained by simply averaging the matched pair price ratios, usually without weighting. The price index is then computed by aggregating class-level price- and quantity information using an index-number formula, which is almost always the Laspeyres.

There are two distinct parts in this basic process for computing official price indices. First, there is the quality control part, achieved through the classification and the items selection from each class. This ensures that the index is comprehensive and that observed unit price changes are legitimate measures of price changes. At that level there is no index number problems involved, at least from the point of view of traditional index number theory. The second part is the aggregation. This is the focus of classical index theory, which defines and examines the properties of index formulae. A typical side-process is the calculation of a number of sub-aggregate level indices for different subsets of interest (i.e. food-price index in the consumer price index (CPI)) alongside with the "total" price index. A matter of interest and concern for the statistical agencies is the coherence between the total price index and the sub-indices and it requires, among other things, careful planing of the classification to ensure that a sub-index coverage is an union of distinct classes.

The need for quality adjustment in official indices arises when the above discussed basic quality-control procedure fails. In the CPI this is nowadays typically the case of household appliances and especially computers, since models on the market disappear very fast and are replaced by more sophisticated new ones. A quality adjustment procedure, as understood in statistical offices, should tackle that type of problems and should not affect other aspects of the index compilation such as aggregation issues.

An official index for house prices must be interpretable within the established paradigm, which, after all, has sound foundations in classical index number theory. From this point of view the real estate price indices described in the literature should be viewed as estimation techniques aimed at obtaining quality adjusted price ratios within some reasonably homogenous class of houses. Researchers typically provide computations for well defined geographical regions within which different houses can be viewed as reasonable substitutes. However the treatment of the topic is usually heavily concentrated around estimation techniques and econometric issues, the links to standard index construction practices and issues such as computation of an overall index are by-passed as self-evident or outside the scope of the research.

On the other hand, some recent research topics such as the so-called time-aggregation bias (Calhoun, Chinloy and Megbolugbe (1995), Englund, Quigley and Redfearn (1999)) , are not relevant from the standpoint of classical index number theory. The problem in a nut-shell is that prices of houses may change significantly from e.g. month to month and a quarterly index will fail to report such changes. Such smoothing of the series cannot be named a bias in the paradigm of classical index theory, where the length of the base and the comparison period can be freely determined. Of course, the loss of information on within-quarter price changes is highly relevant e.g. for purposes of investment risk analysis as pointed out in Calhoun, Chinloy and Megbolugbe (1995), but it nevertheless does not relate to the notion of bias as normally understood by index number experts.

Most important, the existing research is not explicit enough about which characteristics and how affect the quality adjustment. Although theoretical foundations of the hedonic method are provided already by Rosen (1974) and Triplett (1983), only very recently Triplett (2001) and Diewert (2001) have shown, for example, that the matched pairs method normally used in statistical offices can actually be expressed in a hedonic regression form. This seems to be a major reason why statistical agencies are still not comfortable with quality adjustment based on regression methods. Because the author's belief is that this line of research is crucial for the wider acceptance of hedonic quality adjustment in official statistics, an important part of the discussion in this paper is focused on how tangible estimates of the size and the magnitude of the quality adjustment of the characteristic included in the model can be obtained.

2. The History of the Finnish Official House Price Index

2.1 Main Features of the Finnish House Market

Around three fourths of the net wealth of Finnish households is in housing, meaning that home ownership is the only considerable wealth asset for most Finns. The proportion of owner-occupancy rate in Finland has declined in the 90's but is still as high as 60 percent. The decline in owner-occupancy rate is explained largely by the abolishment of rent controls and to a smaller extent by the appearance of new forms of occupancy, which stay between ownership and renting.

Single family houses and apartments in blocks of flats form 40 percent of the housing stock each, apartments in terraced houses account for 13 percent of the stock and the rest of the dwellings are apartments in other than residential buildings. The volume of new housing has been rather low since the end of the 80's, at present new dwellings form about one percent of the stock per year and at least 1/3 of them are intended for rental use. Building companies are involved mainly in the construction of blocks of flats and terraced houses, while most single family houses in Finland are still built using mainly family's own labour.

An institutional peculiarity on the Finnish housing markets is the rather technical distinction between real estate ownership and ownership of a dwelling in a housing share corporation. The land and the buildings in the case of blocks of flats and terraced houses are owned by so-called housing share corporations (asunto-osakeyhtiö). The owner of a dwelling formally owns those shares in the company, which explicitly entitle her rights to a specified dwelling. Single family homes typically form real estates, but for taxation reasons some single family houses are also organised as housing share corporations.

House markets transactions are concentrated in the bigger cities and apartments in blocks of flats account for over 50 % of the transactions of old (existing) dwellings. Single family house is transacted in only about 10-20 % of the cases. Most buyers buy the dwelling for own use.

Finnish house markets are believed to be very sensitive to interest rates, since a purchase of a dwelling is often financed by a loan, whose interest rate is linked to some short term market rate such as the six or twelve month Euribor. House loans are issued by deposit banks and in the past had short repayment period, five to ten years. The situation changed in the 90's and currently a loan can be normally taken with repayment period of 25 and even 30 years.

The Finnish house markets have been very volatile in the past 15 years. During the second half of the 80's liberalisation of the loan market together with booming economy lead to extremely strong price upsurge. In the subsequent recession prices fell to about half of their 1988-1989 levels. As the economy recovered, prices started to grow fast again in 1996. Rising housing prices were supported by interest rate stability secured by Finland's EU-membership and the government's commitment to the monetary union. A strong internal migration towards the capital region lead to scarcity and above-the-average housing price increases in Helsinki and the surrounding municipalities. Towards the end of 2000 the long upward trend in house prices at least temporarily bent down.

2.2 Data Sources and Definitions of Statistics Finland's House Price Index

The importance of the house market developments is widely recognised in Finland and house prices are an important topic of interest for politicians, social planners, economy analysts and ordinary people as well. Against this background it seems natural that Statistics Finland publishes a quarterly house price index since 1985. Afterwards computed historical series are available for some regions from 1970 onwards. Up to the end of 2001 the index was based on actual transaction prices of dwellings in housing share corporations, in practice apartments in blocks of flats and terraced houses. Information was gathered from the major real estate agents and covered about 1/3 of all transactions in that type of housing, but regionally and at different times the coverage has varied from 20 to 70 percent of all transactions. The data source and the methodology were renewed at the beginning of 2002. The new data source and the method are described in the following sections of the paper.

The old and the new index are based on the same definitions. The index covers transactions of dwellings in housing share corporations. Detached family houses forming a separate real estate are not included. The reason is twofold. First, real estate transaction information is registered separately. Second, since only a small number of real estate are carried out quarterly it will not be possible to reports the index at the same regional breakdown as the current one. Statistics Finland has started to produce a separate index for single family houses at rough regional level since the beginning of 2002. Over the last 15 years the price movements in single family houses have followed extremely closely the price movements of the dwellings covered by the house price index.

The house price index excludes new dwellings again because only few transactions are reported quarterly. Statistics Finland is working on the possibilities to collect more representative data on new house sales. Since there is no information on whether a flat is sold by the building company to a private person/household, all flats whose year of completion is the current or the previous calendar year are considered new and the ones with earlier year of completion are defined as old.

At present the house price index contains a series for the whole country (Åland not included), the provinces, major towns and other interesting geographical entities such as Greater Helsinki, comprising the municipalities of Espoo, Helsinki, Vantaa and Kauniainen. Within the regional breakdown separate series are published for type of building (block of flat or terraced house) as well as by number of rooms for the apartments in blocks of flats. The regions for which separate series are currently computed is presented in the Appendix. All results presented in the paper follow the definitions of the official index.

2.3 The Pure Classification Index

From 1985 till the second half of 1995 the house price index was compiled using a pure classification approach. Data was stratified first by region and within the region in the following way:

	Apartments in blocks of flat			Terraced house
Year of completion	1 room	2 rooms	at least 3 rooms	
up to 1960				
1961-1970				
after 1970				

Thus a classification by (region) X (type of building)X(number of rooms)X(age group) was created. Each cell in this classification was treated, in the terminology of the Harmonised Index of Consumer Prices (HICP), as an elementary aggregate. Simple arithmetic price ratios between the base and the comparison period were compiled and then aggregated into an index using the Laspeyres formula. The weights[3] were determined by the number houses in each elementary aggregate in 1980.

If in any particular period the number of transactions in an elementary aggregate cell was 0, then the index was computed under the assumption that the prices between the current and the previous quarter in the cell had remained unchanged. We shall argue later on, that this procedure does not bias the index unless some class with positive weight is constantly empty.

After several years of experience with this method it was considered, that it is not necessarily adequate for the smaller regions since the index series there exhibited too much quarter-to-quarter fluctuations possibly caused by uncontrolled quality variation of the flats sold. Since no further classification of the data was feasible, the solution was seen in using hedonic methods, which potentially allow to control for more quality characteristics than classification at the cost of imposing more structure on the way these enter the house price formation (i.e. the regression functional form). Thus the method used until the end of 2001 and described in the next section was adopted.

[3] Strictly speaking, the Laspeyres weights should be value shares

2.4 Time-Dummy Hedonic Regression

The variation observed in the price per square metre is grouped into three main factors: differences in the individual characteristics of the dwellings, the effect of location and residential area, and the sales period. The model applied is of the form:

$$(2.1)\quad Ln(p_{ati}) = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kti} + \sum_{a=2}^{N} \mu_a A_{ati} + \sum_{t=2}^{J} \lambda_t T_{ti} + \varepsilon_{ati}$$

The explanatory variable is the logarithmic price per square metre $ln(p_{ati})$, where sub-index i refers to the observation, a to the location and t to the time period. x_{kti} refers to the value of characteristic x_k in observation i at period t. A_{ati}'s are the location dummy variables which receive the value 1 if dwelling i is located in area a. T_{ti} ′s are the time dummy variables,.β_0 β_1 ...β_K, μ_2...μ_N, λ_1 ... λ_J are the regression parameters.

Estimation was based on data for eight quarters: the seven most recent and a fixed "base-period" quarter. The estimate of the price trend was obtained from the coefficients of the time dummy variables in the model. Since the explanatory variable was logarithmic, the coefficient estimate of the dummy variable indicated the change in price level in comparison with the "base" period in log- percents. The index point number was computed via the formula $I^t_0 = I_b e^{(\lambda_t)}$, where I^t_0 denotes the index point number for period t, I_b is the index point number in the base quarter and λ_t is the estimate of the coefficient of the time dummy variable for period t.

The locations in the model were identified at postal code area level. Other explanatory variables were floor area of the dwelling and its square root, number of rooms, age of the dwelling and its square root, type of building and physical condition indicators.

Equation (2.1) is a very standard specification in the hedonic index literature going back to Grilliches (1971) and still very popular. A particular feature of Statistics Finland's implementation was that instead of weighting results from separate regional regressions to obtain aggregate level indices, equation (2.1) was independently estimated for all subsets of the data for which an index was to be compiled. This feature lead to incoherent series. The problem is the following. If say, in Helsinki, prices for apartments in blocks of flats rose by 5 % from the previous quarter and prices of apartments in terraced houses rose by 3 %, then one would expect that the average price rise for all types of flats will be between 3 and 5 percent. This is not guaranteed by the described method, since the regression coefficients, in particular the coefficients of the time dummies for blocks of flats, flats in terraced houses and all flats are estimated by running separate, independent regressions.

It should be noted that this type of non-coherency problems are not uncommon and some famous classic index formulas such as the Fisher formula can certainly exhibit the same problem. A more subtle but similar problem is related to the fact that in a chained index the overall index change between two periods may be smaller or greater than the minimum or maximum change of its sub-indices. In practice, such situation may not exist or may not be noticed in official price statistics, especially in the CPI, first because CPI uses the Laspeyres formula, which is consistent in aggregation, second, the CPI is usually divided into at least 5-10 sub-indices and third it is seldom chained.

Obviously, appropriate weighting solves the coherency problem. However it will not remove the low transparency of the procedure. If one wishes to compare the results of the hedonic equation of the form (2.1) with say, an index based on classification etc., one would be able to state in what way the series differ and may suggest some plausible explanations, but no tangible numeric evidence of what regressor and how affected the quality adjustment and the index can be provided. The situation is to be compared with the normal practice of the CPI, where, say, the evaluation of the impact of gasoline price changes on the overall index is a standard procedure. So one has difficulties to demonstrate that the regression does its job appropriately especially in situations where the hedonic index persistently moves faster or slower than some unadjusted measure such as the simple price average. It makes it then rather understandable, that statistical offices feel uneasy about hedonic solutions of the above type, because the credibility of the statistics requires that there is some accounting of what was adjusted and why. The method proposed in the following section addresses this issue.

3. Methodology of the New Index

In view of the preceding discussion it is natural to search for combined methods that will eventually retain the good features of both the classification and the hedonic approach but will mitigate the problems associated with either of them. What is suggested here is first classifying the data with respect to the characteristics along which most price variation is observed. Then regression analysis is used to do cell-specific quality adjustment with respect to other important characteristics. Combining classification and hedonic regressions is not a novelty in itself, but in what follows the focus in on explicit evaluation and aggregation of the impact that different regressors have on the overall quality adjustment. The purpose is to obtain an index which fits as closely as possible into the traditional index number construction practices and at the same time makes full use of the hedonic quality adjustment procedure.

3.1 Classification

Location, type of building and number of rooms are the most fundamental characteristics of the dwelling, since they cannot be changed afterwards at all or only at large cost. It is also with respect of these characteristics that prices vary most. The regional stratification used was determined so as to form interpretable geographical entities with relatively similar price-level. The largest municipalities, for which separate index series are to be published were divided into two to four sub-regions by examining average prices of dwellings in 1995. On the other hand, smaller municipalities with few transactions were grouped together on regional basis. Within each region the dwellings were divided by type and number of rooms as follows

apartments in blocks of flats			apartments in terraced houses	
1 room	2 rooms	at least 3 rooms	1 or 2 rooms	at least 3 rooms

73,7 percent of the total price variation in the data for the year 2000 is between the cells suggesting that the adopted classification groups the observations into relatively homogenous groups.

After the data is classified, construction of a classical Laspeyres index straightforward. One proceeds by treating houses in each cell of this classification as perfect substitutes. Under this assumption, average price changes within each cell provide an unbiased estimate of price change and the index will be obtained by aggregating average prices across cells using Laspeyres' formula. We need only to agree on how cell price averages will be computed and how the cell quantities will be determined. In this case we use geometric price cell averages of square metre prices throughout. The rationale is that half-logarithmic regression modelling provides exact decompositions of within-cell geometric average price ratios into pure price and quality change components, and this feature is very attractive in what follows. Geometric averages are also recommended by HICP. The fixed Laspeyres quantities are the total floor area in the housing stock in each cell in 2000. Using weights based on stock rather than on the transacted dwellings in the base period is motivated by the fact that a house price index of old dwellings seeks to answer the question of how prices in the stock have developed on average. Purchasing a house is a very different concept than consuming housing services and thus weights based on transacted dwellings will have no clear meaning. With these conventions the house price index based on simply classifying the data is:

$$Ind_0^t = \frac{\sum_{i=1}^{N} (\overline{floor\_area}_0^i * n_0^i) * \overline{p}_t^i}{\sum_{i=1}^{N} (\overline{floor\_area}_0^i * n_0^i) * \overline{p}_0^i} \qquad (3.1)$$

where

N is the number of cells in the classification

$\overline{p}_0^i$ and $\overline{p}_t^i$ the average geometric prices at the base and the comparison period in cell i respectively

$\overline{floor\_area}_0^i$ and n_0^i the geometric average floor area and the number of dwellings in cell i in the base period

3.2 The Within-Cell Hedonic Quality Adjustment and its Decomposition

The above classification does not consider, among others, age of the house, floor area and micro-location, so price variation due to sample mix changes with respect to these characteristics will pass as price change in the index.

The quality adjustment strategy proposed is the following. A regression of standard type

$$Ln(p_{itj}) = \boldsymbol{\beta}_t' \mathbf{x}_{itj} + \boldsymbol{\varepsilon}_{itj} \qquad (3.2)$$

is specified either separately for each cell or for a larger section of the data including the necessary indicator variables to ensure that the sum of residuals for each cell are identically 0 by the properties of ordinary least squares (OLS) estimation. The sub-index i refers to the cell, t refers to the time period and j refers to the observation. p_{itj} is the price per square metre of floor area and $Ln(p_{itj})$ its natural logarithm. $\mathbf{x}_{itj}$ is the vector of characteristics, $\boldsymbol{\beta}_t$ is the vector of unknown parameters to be estimated and ε_{itj} is the statistical error term.

Denote the OLS-estimate of the parameter vector for cell i in the base period by $\hat{\mathbf{b}}_0^i$, and for the comparison period by $\hat{\mathbf{b}}_t^i$, the cell average vector of characteristics in the base period by $\overline{\mathbf{x}}_0^i$, the corresponding vector for the comparison period by $\overline{\mathbf{x}}_t^i$, and the geometric average prices in the two periods as $\overline{p}_0^i$ and $\overline{p}_t^i$ respectively. The following decomposition of the geometric average price ratio than is identically true:

$$(3.3) \qquad \frac{\overline{p}_t^i}{\overline{p}_0^i} = \exp\left[\hat{\mathbf{b}}_0^{i\,\prime}(\overline{\mathbf{x}}_t^i - \overline{\mathbf{x}}_0^i)\right]\exp\left[(\hat{\mathbf{b}}_t^i - \hat{\mathbf{b}}_0^i)'\overline{\mathbf{x}}_i\right]$$

The first term of the expression has clearly the interpretation of fraction of the price ratio due to quality difference of the sample mix at base period valuation of the characteristics. The second term can be interpreted is a price change due to changes in valuations, that is a “true price change”. This type of decomposition introduced by Oaxaca (1973) is well known in the literature on wage discrimination, but to the author’s knowledge is not commonly used in research related to hedonic indices. The quality adjusted price ratio is then

$$(3.4) \qquad \exp\left[(\hat{\mathbf{b}}_t^i - \hat{\mathbf{b}}_0^i)'\mathbf{x}_t^i\right] = \frac{\overline{p}_t^i}{\overline{p}_0^i \exp\left[\hat{\mathbf{b}}_0^{i\,\prime}(\overline{\mathbf{x}}_t^i - \overline{\mathbf{x}}_0^i)\right]} = \frac{\overline{p}_t^i}{\exp\left[\hat{\mathbf{b}}_0^{i\,\prime}\overline{\mathbf{x}}_t^i\right]},$$

since by the properties of OLS it is true that

$$(3.5) \qquad \overline{p}_0^i = \exp\left[\hat{\mathbf{b}}_0^{i\,\prime}\overline{\mathbf{x}}_0^i\right]$$

The quality adjusted cell average price in the comparison period, denoted as $\overline{p}_t^i(qa)$ is respectively

$$(3.6) \qquad \overline{p}_t^i(qa) = \frac{\overline{p}_t^i}{\exp\left[\hat{\mathbf{b}}_0^{i\,\prime}\overline{\mathbf{x}}_t^i\right]}\overline{p}_0^i$$

Denoting by N_t^i the number of observations in cell i in period t, the right hand side of (3.4) can be written also as

$$\text{(3.4')} \qquad \frac{\bar{p}_t^i}{\exp\left[\hat{\mathbf{b}}_0^{i\,\prime}\,\bar{\mathbf{x}}_t^i\right]} = \frac{\exp\left(\frac{1}{N_t^i}\sum_{j=1}^{N_t^i} Ln(p_{itj})\right)}{\exp\left(\frac{1}{N_t^i}\sum_{j=1}^{N_t^i}\hat{\mathbf{b}}_0^{i\,\prime}\,\mathbf{x}_{itj}\right)} = \left(\prod_{j=1}^{N_t^i}\frac{p_{itj}}{\exp(\hat{\mathbf{b}}_0^{i\,\prime}\,\mathbf{x}_{itj})}\right)^{\frac{1}{N_t^i}}$$

Since $\exp(\hat{\mathbf{b}}_0^{i\,\prime}\,\mathbf{x}_{itj})$ in the last term of (3.4') is a consistent price prediction of what a particular dwelling actually sold in period t would have been in the base period, it follows that $\frac{p_{itj}}{\exp(\hat{\mathbf{b}}_0^{i\,\prime}\,\mathbf{x}_{itj})}$ is a consistent prediction of the price ratio of this dwelling between the two periods and corresponds to a matched-pair price ratio with the difference that the denominator is not an actually observed price but a price prediction. Now it is clear, that the quality adjustment price ratio (3.4) has the interpretation of geometric average of matched-pairs price ratios and is completely analogous to what statisticians would normally compute at cell level in the classical index construction set-up.

It is worth noting that equation (3.2) is linear with respect of the coefficients $\boldsymbol{\beta}_t$ but not necessarily with respect to characteristic vector $\mathbf{x}_{itj}$, some of whose elements may be polynomials or other transformations of certain "basic" variables, e.g. age and squared root of age, which is our case as shall be seen later on in the text. In such situations the validity of results (3.3) – (3.6), especially the interpretation of (3.4) as an average of matched-pairs ratios, requires that the average $\bar{\mathbf{x}}_t^i$ is computed separately for each element, ignoring possible functional dependencies between the elements.

To clarify the point, let explicitly denote the characteristics vector $\mathbf{x}_{itj}$ as an appropriately defined function of some functionally unrelated characteristics vector $\mathbf{y}_{itj}$ that is $\mathbf{x}_{itj}$ = g($\mathbf{y}_{itj}$). In the above discussion it is assumed that the average characteristics vector $\bar{\mathbf{x}}_t^i$ is computed over the observations in cell i in period t as $\bar{\mathbf{x}}_t^i = \frac{1}{N_t^i}\sum_{j'=1}^{N_t^i}\mathbf{x}_{itj} = \frac{1}{N_t^i}\sum_{j'=1}^{N_t^i} g(\mathbf{y}_{itj})$.

A natural notation for the last term is $\overline{g(\mathbf{y}_{itj})}_t^i$. This is emphasised, because there is another possibility. Let $\bar{\mathbf{y}}_t^i$ denote the average of the "basic characteristics" $\mathbf{y}_{itj}$ in cell i

for period *t*. $\overline{\mathbf{y}}_t^i$ is interpreted as the characteristics of the average property sold. It would be natural to ask what is the quality-adjusted price ratio at the "representative" point in the space of characteristics defined as g($\overline{\mathbf{y}}_t^i$). If g(.) is non-linear, the answer to this question will differ from (3.4), because then $\overline{g(\mathbf{y}_{itj})}_t^i$ will not in general equal g($\overline{\mathbf{y}}_t^i$). However, according to Vartia and Koskimäki (2001), whose paper examines different quality adjustment alternatives, in many cases the difference will not have practical importance.

Another very important point is that decomposition (3.6) is not unique. The following equality is also identically true:

$$\text{(3.7)} \qquad \frac{\overline{p}_t^i}{\overline{p}_0^i} = \exp\left[\hat{\mathbf{b}}_t^{i\,\prime}(\overline{\mathbf{x}}_t^i - \overline{\mathbf{x}}_0^i)\right]\exp\left[(\hat{\mathbf{b}}_t^i - \hat{\mathbf{b}}_0^i)'\overline{\mathbf{x}}_0^i\right]$$

After some arithmetic manipulations it is seen that (3.7) implies the following quality adjusted cell average price for the comparison period:

$$\text{(3.8)} \qquad \overline{p}_t^i(qa)'' = \exp\left[\hat{\mathbf{b}}_t^{i\,\prime}(\overline{\mathbf{x}}_0^i)\right]$$

While the quality adjusted price representation (3.6) amounts to updating the base period prices by an evaluation of the price change of the observed comparison period characteristics mix, (3.8) is a direct evaluation of base period mix at comparison period prices. The term

$$\text{(3.9)} \qquad \frac{\overline{p}_t^i}{\exp\left[\mathbf{b}_0^{i\,\prime}\overline{\mathbf{x}}_t^i\right]}$$

certainly has a cell-level Paasche index interpretation and may seem inconsistent with the Laspeyres framework. However, another interpretation, expressed in the terminology of HCPI, is, that the variants at elementary aggregate level are constantly changing and then the quality adjustment procedure for the price of the comparison period variant mix is certainly not against the current principles of HCPI.

Using (3.6) rather than (3.8) is preferable in our case for the following reason. In the empirical part of the paper a quarterly index is computed, but a whole year is chosen as base period. Consequently there is much more data to estimate the base period coefficients than to estimate comparison period coefficients. The abundance of degrees of freedom allows using extensive set of location dummies covering separately areas where only few transactions per year occur. If one were to estimate regressions on quarterly basis, one would be restricted to much narrower set of location dummies to ensure that there are observations in each location for every period. With the chosen specification one is able to evaluate (3.9) and hence (3.6) for any subset of locations considered in the base period model that may occur in the comparison quarter's sample.

Another argument in favour of (3.6) over (3.8) is that the regressions for the comparison period do not have to be estimated. This is of great importance for official statistics where a production system should be as simple as possible. Still there is no restriction on the time structure of underlying parameters, using (3.6) simply does not require explicit estimates.

3.3 Evaluating the Impact of Characteristics on the Quality Adjustment

Evaluation of the effect on different characteristics on the quality adjustment and the index will greatly improve the transparency of the statistical procedure, facilitate its empirical evaluation and provide useful information for further analysis of the housing market.

An exact decomposition consistent with the discussion in section 3.2 is possible for index formulae having logarithmic representation, such as the Törnqvist formula. Here we discuss a more simple case for the log-Laspeyres formula defined in standard notation (see Vartia (1976)) as:

$$\text{(3.10)} \qquad \log\!-La_0^t \equiv \exp\Big(\sum_{i=1}^{N} w_0^i \ln\Big(\frac{\overline{p}_t^i}{p_0^i}\Big)\Big), \text{ where } w_0^i = \frac{p_0^i q_0^i}{\sum_{i=1}^{N} p_0^i q_0^i}$$

Now using (3.4) we have for the within cell quality adjusted index

$$\text{(3.11)} \qquad \log\!-La_0^t = \exp\Big(\sum_{i=1}^{N} w_0^i \ln\Big(\frac{\overline{p}_t^i}{\overline{p}_0^i \exp\Big[\hat{\mathbf{b}}_0^{i\,\prime}(\overline{\mathbf{x}}_t^i - \overline{\mathbf{x}}_0^i)\Big]}\Big)\Big)$$

$$= \exp\Big(\sum_{i=1}^{N} w_0^i \ln\Big(\frac{\overline{p}_t^i}{\overline{p}_0^i}\Big)\Big)\exp\Big(\sum_{i=1}^{N} w_0^i \hat{\mathbf{b}}_0^{i\,\prime}(\overline{\mathbf{x}}_0^i - \overline{\mathbf{x}}_t^i)\Big)$$

The first term of (3.11) is the pure classification log-Laspeyres index, and the second term is the explicit within-cell quality adjustment at aggregate level. One can group the estimated vector of characteristics valuation and the average characteristic vector into i.e. location (*L*) and size (*S*) components as follows:

$$\text{(3.12)} \qquad \hat{\mathbf{b}}_0^i \equiv \big[\hat{\mathbf{b}}_{0L}^i \hat{\mathbf{b}}_{0S}^i\big],\ \overline{\mathbf{x}}_0^i \equiv \big[\overline{\mathbf{x}}_{0L}^i \overline{\mathbf{x}}_{0L}^i\big],\ \overline{\mathbf{x}}_t^i \equiv \big[\overline{\mathbf{x}}_{tS}^i \overline{\mathbf{x}}_{tS}^i\big]$$

Then from (3.11) and using (3.12) one can decompose the quality-adjustment term into location and size components:

$$\text{(3.13)} \quad \exp\Big(\sum_{i=1}^{N} w_0^i \hat{\mathbf{b}}_0^{i\,\prime}(\overline{\mathbf{x}}_0^i - \overline{\mathbf{x}}_t^i)\Big) = \exp\Big(\sum_{i=1}^{N} w_0^i \hat{\mathbf{b}}_{0L}^{i\,\prime}(\overline{\mathbf{x}}_{0L}^i - \overline{\mathbf{x}}_{tL}^i)\Big)\exp\Big(\sum_{i=1}^{N} w_0^i \hat{\mathbf{b}}_{0S}^{i\,\prime}(\overline{\mathbf{x}}_{0S}^i - \overline{\mathbf{x}}_{tS}^i)\Big)$$

Naturally (3.13) can be extended to examine the aggregate effect on quality adjustment of each explanatory variable in the regression included.

Decomposition (3.13) does not hold exactly for the Laspeyres index, so either log-Laspeyres should be used[4] or (3.13) can be used as an approximation to evaluate the approximate effects of the quality adjustment factors on the Laspeyres index as follows. Define:

$$(3.14) \qquad \delta_0^t \equiv \frac{\ln(La_0^t)}{\ln((\log\!-La_0^t)}, \text{ so that } La_0^t \equiv \left\{\log\!-La_0^t\right\}^{\delta_0^t}$$

Then obviously, the quality adjustment of the Laspeyres index due to e.g. location differences can be evaluated by:

$$(3.15) \qquad \left\{\exp\sum_{i=1}^{N} \mathbf{b}_{0L}^{i}{}' (\overline{\mathbf{x}}_{0L}^{i} - \overline{\mathbf{x}}_{tL}^{i})\right\}^{\delta_0^t}$$

In our case the log-Laspeyres and the Laspeyres indices are so close, that (3.13) can be used directly without empirical problems.

[4] Log-Laspeyres index is always lower than or equal to the Laspeyres index with the same base year and weights (see Fisher (1922) and Vartia (1976)). In our case the difference between the two is negligible.

4. Empirical Results

4.1 Data Source

Research and planning work for renewing the House price index of Statistics Finland started in 2001 after Statistics Finland and the Finnish Tax Administration agreed, that information on transaction prices of apartments in blocks of flats and terraced houses will be provided on quarterly basis. The tax authority collects the prices in connection with the asset transfer tax, which is paid by the buyer and amounts to 1.6 percent of the price. The tax authority data has been available for statistical purposes since 1987, but only on an yearly basis and with delay of about 7 months, thus it could not be used for compiling a quarterly index. The index based on the method described in the following sections is already in use in Statistics Finland. The results reported here cover the period from 1987 till the end of 2000. This is the whole history of the tax authority data. The data of the taxation authorities covers ultimately all transactions of apartments, but on a quarterly basis it covers currently 2/3 share . This is because the buyers have two months time after the purchasing date to pay the tax and inform the authorities and there is some delay caused by information processing. In practice all transactions brokered by a real estate agent (those account for about two thirds of all transactions), will be reported with very little lag, since the real estate agent is responsible for the due payment of the tax and the standard practice is that the tax is paid and the tax form is filled at the moment the transaction takes place.

Table 4.1 below presents the estimated coverage of the data for the whole country and the major towns during the 3rd quarter of 2001. The estimates are based on the assumption that the total number of transaction is the same as during the 3rd quarter of 2000. The precision is good enough for the purpose of overall evaluation of the applicability of the data for constructing a quarterly index. The average coverage of the transactions in the new data seems sufficient with few exceptions. It should be kept in mind that the process is going on and improvement will be for sure achieved. The basic problem seems to be the clearly lower than the average coverage of the transactions in the last month of the quarter (September in the above case). This feature of the data will surely persist also in he future, although not so clearly as now. This suggests that weighing of the monthly observations might be needed in order to capture correctly the price development during the whole quarter. Because there was no "real-time" quarterly data from the tax authority, a pilot research was conducted to analyse the behaviour of the indices under different assumptions of the coverage of the actual quarterly data. This pilot work shall be shortly overviewed later in the text. Since there is no particular value of this research in terms of the central issues of the paper, the results presented further are based on the total data available since 1987 till 2000.

Table 4.1 The estimated coverage of the data during the 3rd quarter of 2001

Region	2001, 3:rd quarter		September 2001	
	reported transactions	%-share of total in 3:rd quarter of 2000	reported transactions	%-share of total in September of 2000
Whole Country	**11819**	**68,9**	**2093**	**35,9**
Helsinki	1807	70,1	233	24,6
Vantaa	568	75,3	118	44,9
Espoo	217	26,0	0	0,0
Porvoo	113	75,8	16	33,3
Tampere	959	87,3	174	49,9
Turku	697	68,4	135	39,2
Oulu	506	91,5	50	27,0
Lahti	380	82,1	72	44,7
Jyväskylä	323	77,8	35	30,4
Kuopio	232	58,3	10	7,0
Pori	136	43,9	16	17,0
Kotka	184	74,5	40	49,4
Kouvola	170	111,1	49	98,0
Hämeenlinna	223	95,7	70	88,6
Vaasa	200	81,3	32	35,6
Joensuu	246	79,6	58	49,6
Lappeenranta	200	92,6	46	73,0
Seinäjoki	29	16,1	1	1,9
Mikkeli	142	100,0	47	109,3
Rovaniemi	124	63,6	18	24,3
Rauma	131	76,2	32	60,4
Kajaani	111	76,6	15	44,1

Other important feature of the data is the extensive use of different registers. The information directly provided by the tax authority contains information only on the transaction price, the dwelling floor area and the municipality of transaction. Using the official apartment identification code prices are linked to other information such as type of building, number of rooms, year of construction and location (postal code and co-ordinates). The sources for this information are the taxation register of real estates, maintained by the tax authorities, and from the building and dwelling register maintained by the Population Register Centre. There are some problems with the use of all these registers concerning primarily new dwellings, for which information may be available only with delay, however long experience at Statistics Finland indicates, that what comes to old apartments the overall quality of the data is high.

In many cases the building in which the sold dwelling was situated is identified, but there is no certainty as to which one of several equally sized apartments was actually transacted. This feature of the data makes utilisation of repeat-sales information rather problematic. Although the results of Meese and Wallace (1997) discussed in section 1.3 quite clearly show that the assumption of time-constancy of the parameters in repeat-sales and hybrid models are very likely to be violated for data set gathered over long time period, it would have been an interesting exercise to examine what results these methods would have provided.

4.2 Regression Estimation

In line with the discussion in chapter 3 the following regression equations for each region were estimated for the whole year 2000 data

$$(4.1) \qquad Ln(p_{ij}) = \beta_0 + \sum_{l=1}^{L_i} \beta_l A_{lij} + \gamma_1(\text{floor\_area}_{ij}) + \gamma_2\sqrt{(\text{floor\_area}_{ij})} +$$

$$\delta_1 age_{ij} + \delta_2\sqrt{age_{ij}} + \sum_{k=1}^{3} \upsilon_k rooms_{kij} + \eta_1 TH_{ij} + \eta_2 (TH)_{ij} * (rooms3)_{ij} + \varepsilon_{ij}$$

Regressions are estimated separately for each region in the classification rather than for each cell (region X type of building X number- of- rooms class), because degrees of freedom for many cells are not enough to obtain reasonably stable estimates for the unknown coefficients. This means, that within each location the explanatory variables are restricted to have the same coefficients. For this reason the subscript *i* refers now to location rather than classification cell, *j* refers to the observation and the subscript for time period *t* is omitted, since in the estimation only the year 2000 data is used.

The general form of the regression model is of standard semi-log type. The dependent variable $Ln(p_{ij})$ is the price per square metre of floor area. The variables A_l , $l = 1.. L_i$ are postal code area indicators for the municipalities, which are separately examined, and municipality indicator variables for the rest of the regions. The variables $rooms_k$, k=1, 2, 3 are room-class indicators. Square roots of age and floor-area variables are included to capture non-linearity of the age and floor area profiles.

TH is an indicator for terraced houses and *TH*rooms*3 is an interaction, which takes value 1 if a terraced house apartment has 3 or more rooms.

The purpose of the model is to provide information for quality adjustment with respect to age, dwelling-floor area and micro-location of the dwelling. The room-number indicators as well as TH and *TH*rooms3* - indicators are included in the regression for technical reasons, to ensure by the properties of the OLS-estimator that the sum of residuals will be zero for all cells in the classification, since results (3.3)-(3.6) and the decomposition (15) hold exactly only if this is true. Obviously, they are strongly correlated with the floor-area, but given the large data, the estimated coefficients for the later are reasonable (see table 4.2).

Some further comments on the choice of the explanatory variables are also in place. First of all, a very short list of apartments' physical characteristics is used. This is because our experience is that type of building and construction year are very strongly correlated with the availability or absence of other characteristics. The age of the dwellings is used as an explanatory variable and since all observations are from the year 2000, the age coefficients can equally well be interpreted as construction-year coefficients and therefore they capture the effects of such "omitted" characteristics. There is inherent ambiguity in the interpretation of the age coefficients in regressions like that and different interpretations affect the quality adjustment differently. The problem is discussed in detail in the following section.

The register data contains information about basic amenities, but in Finland virtually all dwellings have amenities such as some form of heating, hot water, WC and shower. Sauna is a Finnish peculiarity, which is standard equipment in apartments built since the beginning of the 90's and not available in apartments of blocks of flats built in the 70's or earlier. Garage is present in terraced house apartments and is almost always bought separately in the case of apartment in block of flats. An important variable in the data, provided earlier by major real estate agents was the agents' evaluation of the overall condition of the flat, but such information can't be obtained from the register.

The use of extended set of location indicators is justified on the basis that relative differences in price levels by location reflect differences in characteristics difficult to incorporate in the model directly, such as availability of different services, transport connections, recreational activities as well as intangible factors such as the image of the area, which is usually result of a long and complicated socio-economic process. But there is also a potential problem with the location dummies. A change in the relative price level in an area, captured by the coefficient of a location dummy may reflect changes in the quantities of the unobserved characteristics rather than changes in the relative scarcity of the characteristic or changes in buyers' tastes. In case this should happen, it should be in principle viewed as a quality change. A good example is the level of noise. Suppose an airport is just built near a particular location. The level of noise is probably a negative factor, which tends to reduce house prices. One can argue that a quality adjusted index should not view this as a price fall, because physically the same apartment now provides worse housing services than before because of increased noise level. In practice such issues are very difficult to treat in a completely correct way.

Most likely there are omitted variables in our model, but they do not necessarily bias the index, since what we need is not unbiased regression coefficients, but only an unbiased prediction of the price of the average characteristics vector at base period prices (see equation (3.6))[5]. Of course, parameters of wrong size and sign will render the decomposition (3.13) non-interpretable in economic terms.

Using the total data for the year 2000 (58,566 observations) altogether 64 regressions were estimated with a total of 557 location indicators. The Table 4.2 summarises the results in a concise form giving to expositional simplicity more importance than to mathematical purity. The main points remain valid also in the detailed report provided in the Appendix.

Table 4.2 Average estimated coefficients

Variable	**Average coefficient***	**Average t-value***
floor area	0.0067	3.49
square root of floor area	-0.1595	-4.70
Age	0.0111	4.29
square root of age	-0.1797	-6.44
1 room indicator	0.0328	1,42
at least 3 rooms indicator	0.0219	1,16

[5] This point is worth a short comment, because it seems to be sometimes misunderstood. Suppose the true equation is $y_i=\beta x_i+\gamma z_i+\varepsilon_i$ and there is a relation between z and x of the form $z_i=\delta x_i + \omega_i$ (i.e. the variables x and z are correlated) and z is omitted from the regression. The regression equation to be estimated is then $y_i=(\beta+\delta)x_i+(\varepsilon_i+\omega_i)$. Under standard assumptions the conditional expectation of $E(y_i|x_i)=0$ and OLS estimator will be unbiased for the "reduced form" parameter $(\beta+\delta)$. Consequently an unbiased prediction of y_i given x_i will be obtained. It is a different matter, that the estimated coefficient of x has no structural interpretation and is in this sense biased.

terraced house indicator	0,1106	4.18
Terraced house with at least three rooms interaction	0,0227	0,93
*Average values weighted by number of observations in each separate regression		

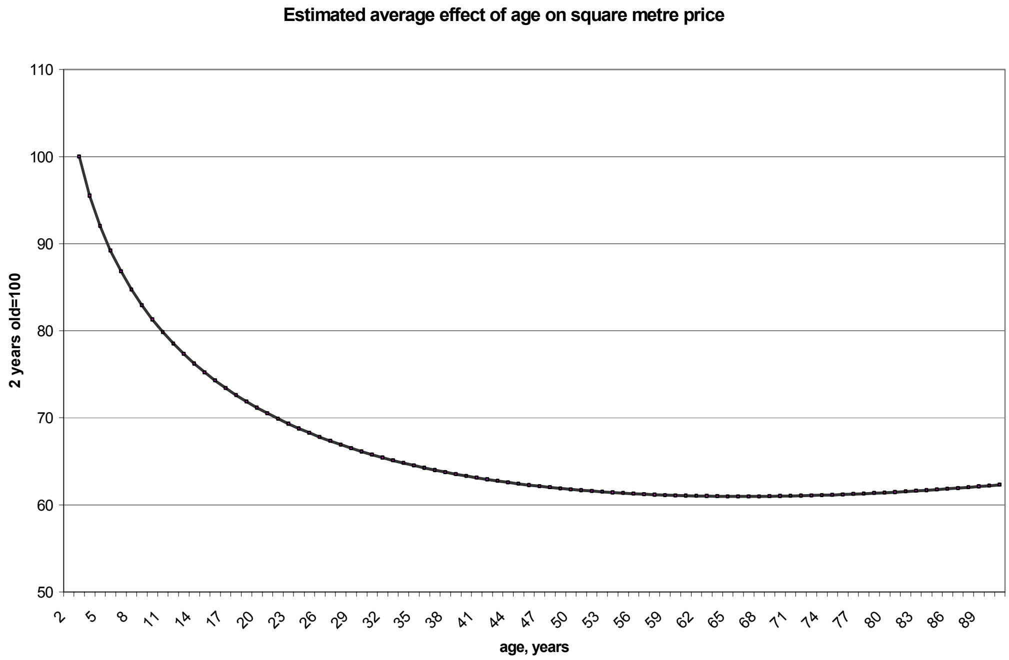
The average R^2 statistic is 38 %, meaning that the models explain on average 38 percent of the within-regional price variation. The statistic is not high, but two things must be noted. First, the regional classification alone already captures 70 percent of the total price variation in the data, so the classification by region and the regional regression together capture over 80 % of the total price variation. Second, the location indicators (not reported in the table because of their very large number) as well as the floor-area and age variables taken together are statistically highly significant, as expected.

Although the coefficients of the room-number indicators are insignificant, they are kept in the model to ensure that the sum of residuals for each room- and type-of-building class is 0, so that the above presented price/quantity decompositions hold as identities.

The above diagram presents the effect of size on prices. The curve is evaluated using all size-related variables in the model, both the room-number indicators and the estimated dwelling floor area polynomial. It appears that square metre prices fall rather quickly with dwelling floor area. The square metre price of a 60 m^2 flat is 75 % of the square metre price of a 20 m^2 flat. This feature can be understood both from the technological and demand side. It is clear, that construction of large flats has economies of scale over small ones. On the demand side, Finland's single-person households account for 37 % of all households and the share is still growing, so demand for small apartments is high. The trend becomes flat for large, over 100 m^2 dwellings.

4.2 Interpretations of the Age Profile: Depreciation versus Vintage Effect

The diagram below presents the estimated average effects of age on prices of dwellings. According to our estimates the price initially falls with age at a yearly rate of 4,5 percent but the rate decreases fast and a 10-year old dwelling will lose only 1,8 percent of its value within a year. The age effect on prices is practically zero for 40 year old dwellings and even turns slightly upwards for over 70 years old dwellings. The fast decline of the price for rather new dwellings may partly reflect extrapolation, since there are few observations of only few years old dwellings in the data. This will not cause serious problems, since the average age of sold dwellings is typically much higher.



The interpretation of the age profiles is of great importance. There are two distinct simultaneously acting reasons explaining the shape of the age profile. The first is that if repairs do not (on average) offset depreciation, then, ceteris paribus, an older home will be in worse condition than a newer one. Another factor affecting the shape of the profile is that newer homes presumably embody better planning and construction technology and probably are on average better equipped. This way of thinking gives also a reason why for old enough houses the age profile eventually turns up[6] - they represent certain historically valuable architectural style and thus have "museum value". In the text the value of a house associated with the architectural style and construction technology of a particular time period is referred to as vintage effect as opposed to depreciation effect associated with physical ageing.

For the purposes of quality adjustment of the index separation of the vintage and the depreciation effects would be desirable, since they have different implications for the quality adjustment. The depreciation effect implies, that if the average age of the dwellings in the sample increases, then the index should be adjusted upwards to reflect the decline of the average quality mix in the data. On the other hand, an increase in the average construction year of the sample would usually[7] call for downward adjustment of the index, because the sample mix embodies newer and better technology.

The problem is that separation of the depreciation and vintage effects is not possible in cross-section data, because the construction year, the selling period and the age are altogether by definition in perfect linear relation. An explicit recognition of the problem and an attempt to solve it by combining cross-section and panel data is provided by Englund, Quigley and Redfearn (1998). Their argument goes as follows. The starting point is specification in line with the hybrid model of Quigley (1995) discussed earlier. The log-price equation is specified in equation (6) in the quoted paper, presented here in notation consistent with the one used in earlier chapters as

$$(4.2) \qquad \mathrm{p}_{it} = \mathrm{P}_t + \boldsymbol{\beta}\mathbf{x}_{it} + \beta_y \mathrm{YR}_\mathrm{i} + \beta_\mathrm{d}\mathrm{AGE}_{it} + \zeta_i + \varepsilon_{it}$$

YR_i is the construction year of dwelling i in the data, AGE_{it} is the age of the dwelling at time period t, and the vector $\mathbf{x}_{it}$ contains the other explanatory variables, P_t is the time-dummy for the sales period, ζ_i is the dwelling-specific term and ε_it the error term. The authors suggest, that the perfect linear dependence between construction year, the selling period and age is solved by the following. They define a new "error" term as

$$(4.3) \qquad \gamma_{it} \equiv \beta_\mathrm{d}\mathrm{AGE}_{it} + \zeta_i + \varepsilon_{it}$$

[6] Another argument is that for old houses the value increase of the land exceeds the decline of value due to depreciation. This is not a valid point here, since the location indicators in the model control for land value differences
[7] If the average construction year is very old, then the "museum value" may reverse the argument.

and then estimate from the repeat-sales sub-sample

$$(4.4)\quad p_{it} = P_t + \boldsymbol{\beta}\mathbf{x}_{it} + \beta_y YR_i + \gamma_{it}$$

Then using residuals of the regression (4.4) as estimates of γ_{it} and the definition (4.3) they continue to estimate the AGE_{it} coefficient (that is the depreciation rate) from:

$$(4.5)\quad \gamma_{it} - \gamma_{i\tau} = (\beta_d AGE_{it} + \zeta_i + \varepsilon_{it}) - (\beta_d AGE_{i\tau} + \zeta_i + \varepsilon_{i\tau}) = \beta_d (AGE_{it} - AGE_{i\tau}) + (\varepsilon_{it} - \varepsilon_{i\tau}) = \beta_d (t - \tau) + (\varepsilon_{it} - \varepsilon_{i\tau})$$

Unfortunately, there is a flaw in the procedure. The assumption that omitting the age variable from the regression will leave the depreciation effect in the error term is incorrect, because age is not uncorrelated with the included variables but on the opposite, it can be expressed as a linear combination of selling time and construction years. Omitting the AGE_{it} from the regression (4.2) will not leave the depreciation effect in the modified error term γ_{it}. The depreciation effect will be completely augmented in the estimated coefficients of the time dummies and the construction year.

The following discussion clarifies the point, which should be quite obvious. In their paper Englund et al. use data from 1981 first quarter to 1993 and construct a quarterly index using time dummies to denote the selling quarter of a house. The dummies are formed in a more complicated way than normally to identify precisely the sale date within a quarter (footnote 2 in the quoted paper). For clarity of exposition, let's suppose here that the sale quarter dummies are formed in the usual way, that is a quarter dummy takes value 1 if the sale is during the quarter and 0 otherwise. Suppose also that the age is measured in years and quarters. It is identically true that AGE_{it}=yyyy/q-YR_i, where yyyy/q denotes the year and quarter of sale. Substituting this expression into (4.2) and rearranging it one obtains

$$(4.2')\quad p_{it} = P_t + \beta_d * \text{yyyy/q} + \boldsymbol{\beta}\mathbf{x}_{it} + (\beta_y - \beta_d) YR_i + \zeta_i + \varepsilon_{it}$$

For all dwellings sold in a particular quarter the term β_d*yyyy/q is constant and can be denoted as A_t and the parameter (β_y-β_d) as β_v. Now what equation (4.4) actually estimates is (4.2'), which can be written also as

$$(4.2'') \qquad \mathrm{p}_{it} = (\mathrm{P}_t + A_t) + \boldsymbol{\beta}\mathbf{x}_{it} + \beta_v \mathrm{YR}_i + \zeta_i + \varepsilon_{it}$$

The estimated coefficients of the time dummies are contaminated by the depreciation effect, which may bias the index if depreciation is significant. The estimated coefficient of construction year might appear to represent the "pure" vintage effect, but this is incorrect. Pure vintage effect could be estimated only if it were possible to evaluate the depreciation, which is not the case. Regressing the differences of the residuals of (4.2') for dwellings sold more than ones on the time between sales as in (4.5) does not estimate the depreciation, but can be rather viewed as a test for time-constancy of the parameters, which is assumed throughout in this setting. Englund et al. estimate the annual "depreciation" rate to be between 0,0023 and 0,00993, the estimates being statistically highly significant in all but one case. The result can be interpreted as evidence against the assumption of time-constant regression parameters.

The only plausible way to separate vintage and depreciation effect remains use of outside information such as expert judgement. Since such information is not easily obtained and its quality will be difficult to evaluate, one is left with the question of which factor is primary in determining the age profile. Most papers include age as explanatory variable in a model specification of the type (4.2), which means deciding implicitly in favour of depreciation. On a-priori grounds the choice is not obvious at all. Assuming that the downward sloping age-profile reflects depreciation means that one views the ageing of the stock as quality deterioration which may or may not be the case, depending on the renovation and repair activity. In Finland at least, most privately funded dwellings are regularly repaired and improved and are on average in very good technical condition. On the other hand newer vintages presumably take advantage of better planning and building techniques and for sure have more and better equipment than older ones. As mentioned earlier in Finland for example saunas are available in almost all apartments of newer vintages. Interpreting the age-profile as reflecting primarily vintage rather than age effect in the Finnish case is probably the choice closer to reality and this is how it is treated here in the quality adjustment procedure. There is no reason to interpret the growth of the average age in the Finnish stock of dwellings as indicator of housing quality deterioration.

The interpretation of the estimated age profile as reflecting vintage effects is achieved through the following simple variable transformation (re-interpretation). When computing the index for a particular period age is not calculated as physical age at the time of sale but as (2000-construction year)[8].

To understand more clearly the practical importance of this discussion, lets go through a simple example. Suppose that the quality adjustment of the index due to "age" is to be computed between 1990 and 1995. Suppose further that the dwellings sold in 1990 were on average built in 1975 and the ones sold in 1995 were on average constructed in 1980. Using the results from table 4.2 the quality adjustment of the index change between 1990 and 1995 due to difference in average construction year will be approximately

$$\exp\left\{\left(0.0111(2000-1975)-0.1797\sqrt{2000-1975}\right)-\left(0.0111(2000-1980)-0.1797\sqrt{2000-1980}\right)\right\}$$

that is about 3.9 percent downward correction reflecting that in 1995 the dwellings sold were of newer vintage and hence better and ceteris paribus more expensive. For comparison, suppose now that in both periods the average construction year were the same, 1975. Under the current interpretation of the age profile as reflecting vintage effects, the quality adjustment in the index will be zero. However, the physical age of the 1975-built dwellings in 1990 is 15 years and in 1995 20 years. If the age profile were interpreted as reflecting depreciation, then in this situation the quality adjustment would be computed as $\exp\left\{\left(0.0111(15)-0.1797\sqrt{15}\right)-\left(0.0111(20)-0.1797\sqrt{20}\right)\right\}$, which would result in about 5.4 percent ***upward*** correction of the index, reflecting that the 1975-built dwellings got older in 5 years and presumably in worse condition.

[8] This does not cause any problems for computing the index for periods up to 2001. From 2001 onwards there is the possibility that some sold flats are build after 2000 and then the square root of age cannot be calculated for them. In such cases the variable (2000-construction year) is set to 0. Since such cases will be very rare in the following 2-3 years, their effect on the index is negligible. Statistics Finland will regularly update the index and the estimated regressions in 3-4 year time periods, so errors will not accumulate. It might appear, that replacing age and its square root with construction year and its square root and re-estimating the regressions would be a clearly better solution, but it is not necessarily so. Of course in the latter case there will be no problems to perform the mathematical operations for dwellings built after 2000, but the danger of extrapolation is present.

5 The Indices

5.1 Some Clarifications

The indices for the period 1987-2000 were estimated using the total data and aplying both the classification and using a within-cell quality correction method. Aggregation of the index was done using both the Laspeyres formula (3.1) and the log-Laspeyres formula. Since there was practically no difference in the results, the official index of Statistics Finland uses the log-Laspeyres formula, for which the decompositions (3.11) and (3.13) hold exactly.

When deciding the classification care was taken not to have classes with systematically less than 5 observations, but on the other hand it was considered unnecessary restriction of the classification to require that each cell has at least five observations in every period. As a result in some (but few) cases there are less than 5 observations or even no observations at all in some cells. The effect of two alternative ways of treating cells with few observations was examined. First, the price change in a cell was imputed by using the estimated change in adjacent cells. For example, if in a particular region there happened to be less than 5 prices of 2-room apartments in blocks of flats, the quarterly change for the blocks of flats in the region is computed using only other than two-room apartments. The cell-level index for two-room apartments is updated by the estimated quarterly change. This amounts to assuming that the price development in the cell with few observations follows the same trend as the most likely “substitutes”. In some cases this assumption may not work very well since it is not clear how good substitute a single-room and two- or three-room apartments are. For this reason an alternative procedure was examined. The cell price indices were computed normally for all situations where there was at least one observation in the cell for the quarter. If there were no observations, it was assumed that the price in the cell remained unchanged from the previous quarter and the cell-level index was left unchanged. This may seem to bias the index, but in author's view it is not the case if the cell is not empty all the time. With a base index, as in this case, the whole price change between the base and the comparison period will be correctly estimated any time there are observations in the cell. What will go wrong are the estimated changes between periods with not enough observations. Since cells where few observations occur are also cells with small weight in the index, the difference in the results of the two procedures becomes negligible at rather low aggregate levels.

Another aspect that was studied was the difference of the coverage of the data for different months in the quarter. The data received immediately after the end of a quarter under-represents transactions concluded during the last month of the quarter. The appropriate treatment is to give month weights to the observations. The procedure may be problematic, if the coverage in the last month is very low, since then a single observation may have considerable effect on the calculations. To study the effect of weighting random samples from each quarter drawn. The sample was 60 % for the first and second month and 20 % for the last month of the quarter. Then the indices were estimated first by giving a weight of 3 for the last month of the quarter and 1 for the other two months and then without weighting. The difference in the procedure did not bring about any significant change in the overall picture as compared to a benchmark case, where a flat 60 % sample per quarter was drawn.

The results presented in what follows are based on all observations for the years 1987-2000, so the above discussion is of little relevance for them, but is important for the "real time" computations. To conform the history with the actual practice now already established, the calculations are done using the currently "official" procedure at Statistics Finland. Price changes for cells with less than 5 observations are imputed and observations from different months are weighted by month weights in the calculation of cell level averages of prices and quality characteristics. The month weights are derived from the average number of transactions in each month in the years 1995-2000 Regional level results are presented in the Appendix.

5.2 Behaviour of the Indices and the Hedonic Quality Adjustment

Diagram 5.1 The overall index for Finland 1987-2000

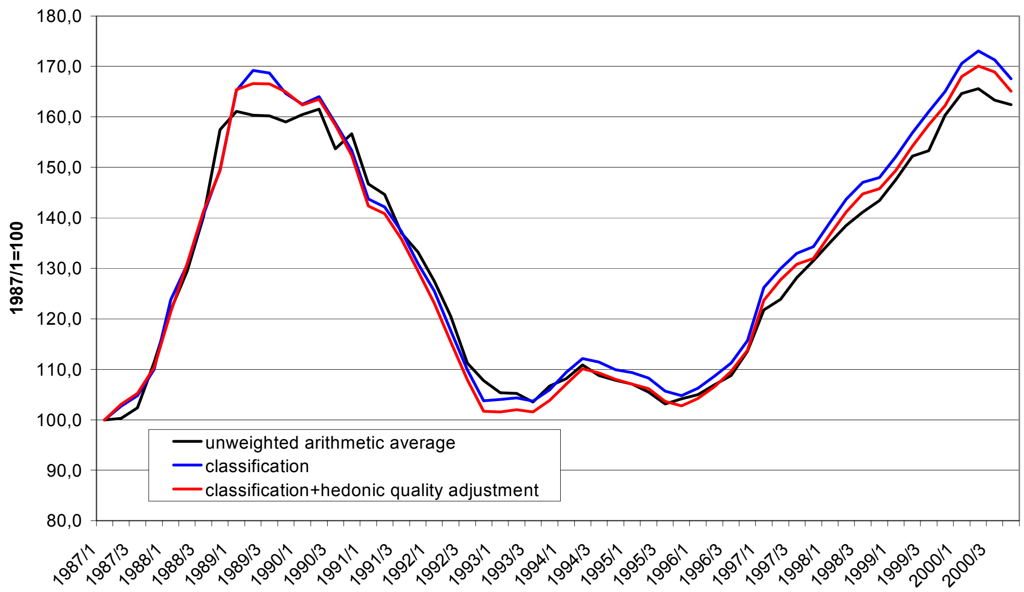
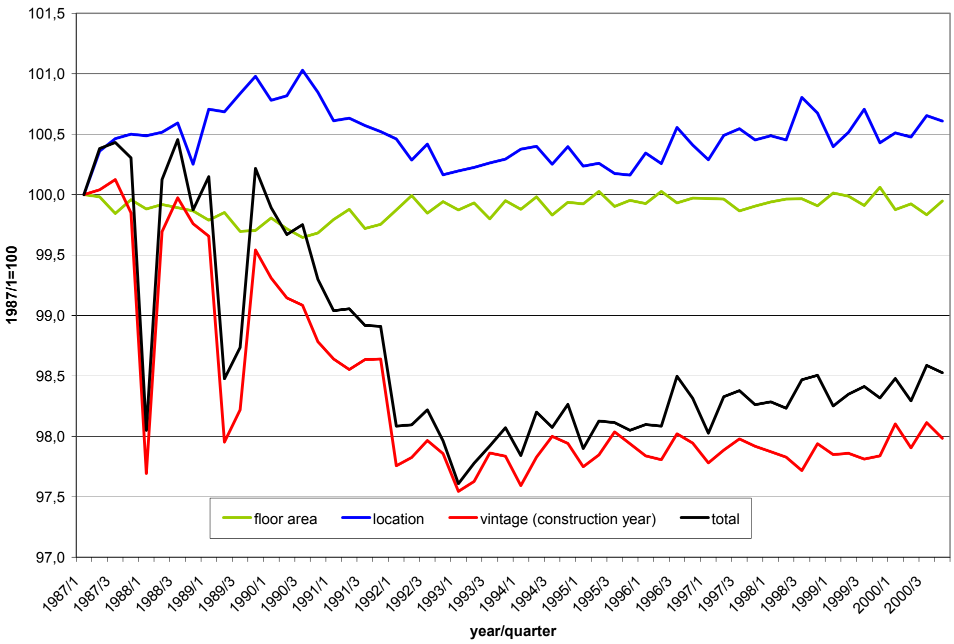


Diagram 5.1 presents results at the overall country level. The "classification" series is the index based on the classification described in section 3.1 and computed via log-Laspeyres formula (3.10) in the text. The unweighted average is computed by comparing the average square metre price in subsequent periods. The "classification + hedonic quality adjustment series" is the within-cell quality adjusted log-Laspeyres index, (3.11). At that level of aggregation even the simple average gives very similar picture of the long run price developments as the indices. Still, a closer look reveals a main problem of the simple average, which is not very serious in this particular case but is potentially dangerous. Both at the peak in 1989-1990 and during the price rise at the second half of the 90's the simple average series shows slower price movement than the indices. Obviously at that times the relative transaction volumes have shifted towards dwellings, where the square metre price is lower. Most likely this reflects shift in the demand towards relatively cheaper regions (locations).

The classification index takes account of the most important determinants of square meter price, region, type of building and number of rooms. This level of quality control seems to be quite enough at very aggregated levels. The within-class quality adjusted index does not change the overall picture. It is systematically below the classification index because by controlling for vintage effects (year of construction) it is taken into consideration that in the longer run newer and better equipped dwellings will enter the stock and improve the average quality of dwellings sold. A quality adjusted index should not interpret that feature as price rise. The point is clarified in diagram 5.2 presenting the overall quality adjustment and its components obtained via (3.13).

Diagram 5.2 The aggregate effect of quality adjustment 1987-2000, Finland, all dwellings



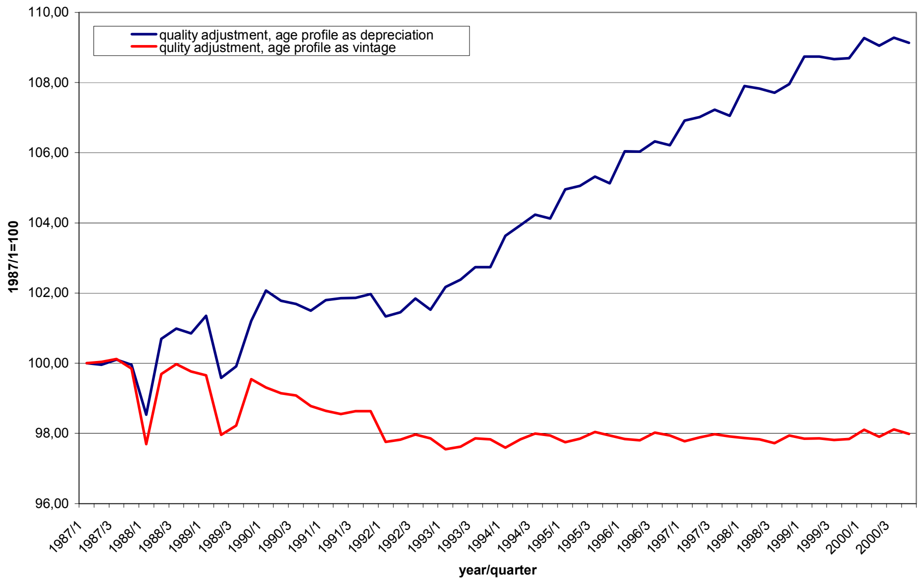
The most important factor affecting quality adjustment is the vintage effect. The high construction volumes at the end of the 80's are seen as a downward trend in the quality adjustment due to construction year: the average construction year became newer, which is interpreted in our method as indicating improvement in the average quality of the transacted dwellings. The trend practically ended after 1993. This observation is consistent with the very low construction volumes in the first half of the 90's and historically low construction volumes even after the economy recovered.

Although of very modest size, the quality adjustment due to micro-location of the dwellings supports economic intuition. As prices rose sharply at the end of the 80's the quality adjustment due to location goes upwards, reflecting shift of transaction volumes towards cheaper locations. The trend is reversed during the recession and turns up slightly again during the recent price upsurge.

The small negative quality correction due to floor area during the end of the 80's is also consistent with the logic of economic theory. The transaction volume has shifted towards smaller dwellings. The shift implies negative quality correction, since square metre prices are higher in smaller dwellings. The overall price of a smaller dwelling is, of course, lower than the price of a larger one, explaining why buyers would have to buy smaller flats when prices are very high.

The overall quality adjustment is negative most of the time and follows closely the vintage effect quality correction. Under the alternative interpretation of the age profile as reflecting depreciation rather than vintage effect discussed in section 4.2, the implications for the long term trend in quality adjustment would be the reverse, as illustrated in diagram 5.3.

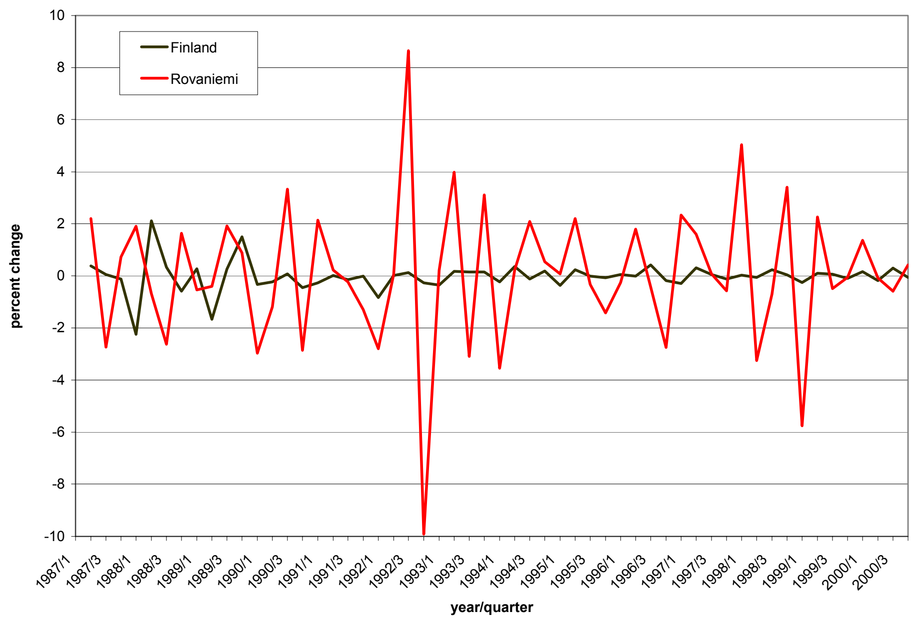
Diagram 5.3 Quality adjustment effects under alternative interpretations of the estimated age profiles



The red line in diagram 5.3 is the same as the quality adjustment due to vintage effects in diagram 5.2. The blue line is the quality adjustment generated under the interpretation of age profiles as reflecting depreciation. After the construction boom at the end of the 80's (seen with lag as a flat part of the line in the beginning of the 90's), new construction forms only a very small proportion of the dwelling stock. As a result each year the average age in the stock grows by almost 1 year. If the downward sloping age profiles reflect depreciation, the ageing of the stock means quality deterioration. In the Finnish case that interpretation leads to unduly strong upward quality correction of the index. It is a well known fact that dwellings in Finland are regularly repaired and even improved. The quality correction when age profile is interpreted as reflecting vintage effect is much smaller in absolute size, since at current construction volumes the average construction year (the average vintage) in the stock grows very slowly.

Although in the long run even the unadjusted average price changes provide a good picture of the price trend, the importance of the use of index formula and hedonic quality adjustment is better understood when viewing the house price index as a short term economic indicator. Then quarter to quarter changes rather than long term trends become important. In this respect unadjusted measures may be more misleading. The relative importance of quality adjustment is also much more important in small market areas than at overall country level.

Diagram 5.4 Quarterly changes in the overall hedonic quality adjustment in Finland and Rovaniemi



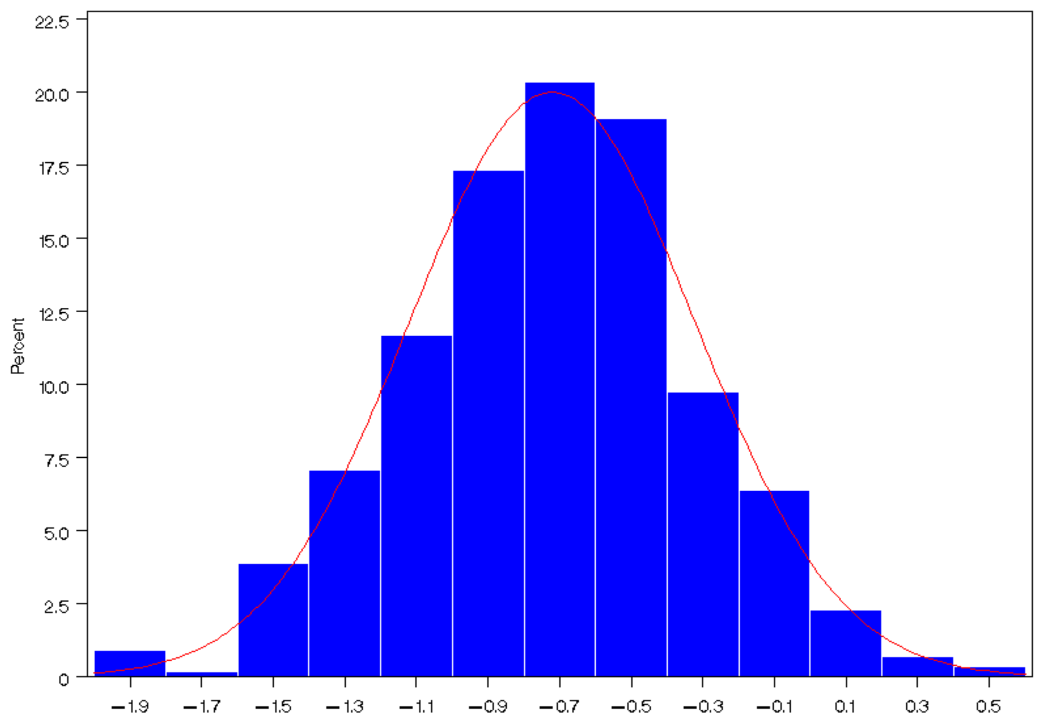
While at whole country level the quarter-to-quarter changes in overall hedonic quality adjustment is never larger in absolute size than 2 percent and is on average 0.3 percent, in Rovaniemi the average absolute quarter to quarter change is 1.96 percent and can be as large as 10 percent. Since 1 percent change in the quality correction between quarters implies approximately 1 percent change in the quarterly index, the strong influence of the hedonic quality adjustment for small markets such as Rovaniemi is obvious. As diagram 5.4 shows, a strong upward movement in the quality correction is usually followed by a strong downward movement and vice versa. This is due to the fact, that if in a particular quarter there is a random shift in the transaction volumes towards, say, high quality and high price dwellings, in the next quarter transactions of normal quality are likely to be observed. In the quarter-to-quarter quality adjustment change this is seen first as a peak downwards, to adjust for the better than the normal quality traded in the first quarter. This is followed by a peak upwards, since the return of transactions to dwellings of normal quality in the second quarter is, from the first quarter's prospective, a quality deterioration.

5.3 Precision of the Index

Precision in the index is evaluated by estimating the width of the 95 % confidence intervals for the quarterly percent change of the overall index and its sub-indices. The confidence interval widths are estimated by numerical simulation. The simulation set-up is the following. From the base year (2000) 565 samples with replacement are selected, each of the same size as the original data. Then from each sample the cell level average prices $\bar{p}_0^i$ and the regression coefficients for the cell level quality adjustment $\hat{\mathbf{b}}_0^i$ are computed. The vector of average prices and the matrix of regression coefficients from the k:th base year re-sampling are denoted as $\left[\overline{\mathbf{P}}_0, \hat{\mathbf{B}}\right]_k$. The quarterly change was arbitrarily chosen to be estimated between first and the second quarters of 1995. 565 samples with replacement of the same size as the original data set were selected from the data sets for the first two quarters of 1995 . Then cell level average prices $\bar{p}_{95/1}^i$ and $\bar{p}_{95/2}^i$ were computed. Denote the cell level average prices vectors for k:th re-sampling of the data in 1995 as $\left[\overline{\mathbf{P}}_{95/1}\right]_k$ and $\left[\overline{\mathbf{P}}_{95/2}\right]_k$ respectively. Now one has 565 independent "data sets" $\left[\overline{\mathbf{P}}_0, \overline{\mathbf{P}}_{95/1}, \overline{\mathbf{P}}_{95/2} \hat{\mathbf{B}}\right]_k$ From each data set index point number for the 1st and the 2nd quarter of 1995 and a quarterly percent change is computed at all levels of aggregation using exactly the same quality adjustment and aggregation procedures as explained in the text. In this way 565 independent values of the index point number are obtained for all sub-indices and the overall index. Since the average floor area and the number of dwellings in each class of the classification were evaluated independently using register information, these variables, which are used to evaluate the base period quantities, were treated as fixed (non-random) in the simulation exercise.

Diagram 5.5 presents the histogram for the distribution of the quarterly change between 1995 1st and 2nd quarters of the overall within-cell quality adjusted index

Diagram 5.5 The distribution of the quarterly change of the overall index between 1st and 2nd quarters of 1995



According to the simulation results the true quarterly change is between -1.46 and 0.03 percent with 95 % probability. The length of the 95 percent confidence interval is 1.49 percent. The original point estimate is -0.79, well within the confidence interval bounds and very close to the simulations' average, -0.72. The distribution is symmetric and very close to normal, the normal approximation to the confidence bounds is very good.

Regional results are provided in the Appendix. The distributions are generally symmetric and $\bar{x} \pm \frac{1}{2} h$, where h is the width of the simulated 95 percent confidence interval, provides a very good approximation to the simulation. The price distribution in the underlying population is so dispersed, that for example in a relatively large market areas such as Lahti, the width of the 95 % confidence interval for the quarterly change is over 9 percent. From the municipalities for which separate statistics are currently published the confidence interval is widest in Kokkola, 16 percent.

The confidence intervals were computed also for the quarterly changes computed from the classification index. They are in most of the cases only slightly wider than for the within-cell quality adjusted index. The author does not think that this fact bears any particular meaning for the desirability of the within-cell quality adjustment procedure, since its primary purpose is to remove potential bias in quarter-to-quarter changes and not necessarily to increase the measurement precision. Of course, the situation would be more complicated if it had turned out that the within cell quality adjusted index quarterly changes had much larger spread than the classification index. Then one possibly has to consider trade-offs between bias and variance, which is a complicated exercise.

It will be quite reasonable to assume, that the variance of the quarterly change depends only on the size of the data. If this is the case, then the estimates of the confidence interval lengths provided here are somewhat conservative, since the number of transactions per quarter is currently 20-25 percent higher than in 1995.

Conclusions

The methodology proposed in the paper is rather straightforward and transparent. The magnitude of quality adjustment due to difference in sample mix can be evaluated for each quality characteristic at each aggregation level. This makes comparison between a standard classification index and the within-cell quality adjusted index presented in the paper very clear. The trends in the quality adjustment components provide some interesting insights. The regression coefficients are allowed to change freely over time while their explicit estimation for each period is not necessary. This largely facilitates the use of the method in production of fast statistics, which are intended to be used as an economic indicator.

The paper does not provide empirical results based on modelling techniques using information on repeat-sales of the same dwelling. The reason is that precise identification of the repeat-sales in the data set used was problematic.

The econometric methodology used on the part of the quality adjustment decomposition is related to the literature on measurement of wage differences, which is a large and important field in econometric analysis of micro data, but the relation of its results to quality adjustment in index number construction does not seem to be generally known.

In abandoning the idea of time-invariant coefficients the paper bears some resemblance to the work of Meese and Wallace (1997). Another similarity is the finding that the long run price trends estimated using simple statistics are very similar to the trend exhibited by a sophisticated index. It seems that in the case of housing prices the role of quality adjustment is important primarily in estimating short-term price movements in relatively thin markets.

The only persistent trend of the quality adjustment found in the research is related to the age profile of housing prices. The shape of the age profile is a mix of vintage and depreciation effects. In view of their importance for the long-run index trend, it would have been highly desirable to separate the effects. Unfortunately, this does not seem to be possible on the basis of sample information, since time of sale, construction year and age are by definition in exact linear relationship and empirical models including all three are unidentified. The decision to interpret age profile as reflecting only vintage effects is simply a practical solution motivated by conditions in Finland.

The width of the evaluated confidence intervals of the quarterly index change suggests that in many cases quarterly changes at municipal and sub-municipal level should be interpreted with care. The precision of the overall index is good. A feature not examined in the paper is the precision of the estimates, when not all information on prices and transactions is available. Currently information about two thirds of all transactions concluded in a quarter are received by Statistics Finland immediately. Information for the rest is received later. The published index points and quarterly changes for the current year are revised and the final estimates are ready in the beginning of the following year.

The applicability of the developed method is evidenced by its use at Statistics Finland. It is of general nature and applicable to other cases where classification and following prices of representative items from each class is difficult. A good example would be construction of wage indices, where both the job characteristics and the qualifications of the worker affect wages and hence neither following the wages in same jobs nor wages of the same individuals is under all circumstances satisfactory. Hopefully the paper presents a good example of incorporating academic research with the practical needs of official statistics.

References

Articles

Bailey M. J., Muth, R. F. and Nourse H. O. (1963) A Regression Model for Real Estate Price Index Construction, *Journal of the American Statistical Association, Volume 58*, *933-942*

Calhoun Charles A., Chinloy Peter, Megbolugbe Isaac F. (1995) Temporal Aggregation and House Price Index Construction, *Journal of Housing Research, 6(3), 419-438*

Case K. E. and Shiller R. J.(1989) The Efficiency of the Market for Single Family Homes, *American Economic Review, 79(1), 125-137*

Diewert E. (2001) Hedonic Regressions: A Consumer Theory Approach, *conference paper, Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, 2001*

Englund Peter, Quigley John M., Redfearn Christian L. (1998) Improved Price Indexes for Real Estate: Measuring the Course of Swedish Housing Prices, *Journal of Urban Economics, 44, 171-196*

Englund Peter, Quigley John M., Redfearn Christian L. (1999) The Choice of Methodology for Computing Housing Price Indexes: Comparisons of Temporal Aggregation and Sample Definition, *Journal of Real Estate Finance and Economics, 19:2, 91-112*

Fisher I.(1922) The Making of Index Numbers, Houghton Mifflin Company*, Boston*

Greene William H. (1997) Econometric Analysis (Third Edition), Prentice-Hall International, Inc., 1997

Griliches Z. (1971) Hedonic Price Indices for Automobiles: An Econometric Analysis of Quality Change*, Zvi Griliches (ed.) Price Indexes ad Quality Changes, 55-97*

Meese Richard A., Wallace Nancy E. (1997) The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches, *Journal of Real Estate Finance and Economics, 14, 51-73*

Quigley John M. (1995) A Simple Hybrid Model for Estimating Real Estate Price Indexes, *Journal of Housing Economics, 4, 1-12*

Oaxaca R. (1973) Male-Female Wage Differences in Urban Labor Markets, *International Economic Review, 14, 673-709*

Rosen S. (1974) Hedonic Prices and Implicit Markets: Product Differentiation and Pure Competition, *Journal of Political Economy,82, 34-55*

Triplett J. (1983) Concepts of Quality in Input and Output Price Measures: A resolution to the User Value-Resource Cost Debate, In M.F. Foss, Ed., *The U. S. National Income and Product Accounts: Selected Topics, NBER Studies in Income and Wealth*, University of Chicago Press, 1983, pp. 269-311

Triplett J. (2001), Handbook on Quality Adjustment of Price Indices for Information and Communication Technology Products, *forthcoming, OECD, Paris*

Vartia Y.O. (1976) Fisher's Five-tined Fork and Other Quantum Theories of Index Numbers, *Theory and Applications of Economic Indices, Proceedings of an International Symposium Held at the University of Karlsruhe April-June 1976, 1978, 271-295*

Vartia Y.O., Koskimäki Timo (2001) Beyond matched pairs and Griliches-type hedonic methods for controlling quality changes in CPI sub-indices: mathematical considerations and empirical examples on the use of linear and non-linear hedonic models with time-dependent quality parameters, *Discussion Paper, Department of Economics, University of Helsinki*

Other Sources

Asuntojen hinnat 1970 – II neljännes 1985, *HI 1985:6, Tilastokeskus*

Asuntojen hinnat 2001 3. neljännes, *Asuminen 2001:8, Tilastokeskus*

Journal of Real Estate Finance and Economics, 14, Kluwer Academic Publishers, 1997

Appendix

Regional classes

REGION	Explanation
Uusimaa	
Helsinki 1	postal code area 00100, 00120, 00130, 00140, 00150, 00160, 00170, 00180
Helsinki 2	postal code area 00200, 00210, 00240, 00250, 00260, 00270, 00280, 00290, 00300, 00310,00320, 00330, 00340, 00350, 00380, 00400, 00440, 00500, 00510, 00520, 00530, 00550, 00570, 00610, 00660, 00670, 00680, 00830, 00850
Helsinki 3	postal code area 00360, 00370, 00390, 00410, 00420, 00430, 00560, 00600, 00620, 00630, 00640, 00650, 00700, 00720, 00780, 00800, 00810, 00840, 00870
Helsinki 4	Rest of Helsinki
Espoo 1	postal code area 02100, 02110, 02120, 02130, 02160, 02170, 02750, 02700
Espoo 2	postal code area 00370, 02140, 02150, 02180, 02200, 02210, 02230, 02270, 02280, 02300, 02600, 02610, 02630, 02680, 02730, 02940
Espoo 3	Rest of Espoo
Vantaa 1	postal code area 01230, 01300, 01370, 01380, 01390, 01400, 01420, 01600, 01630, 01640, 01670, 01690, 01700, 01710
Vantaa 2	Rest of Vantaa
Greater Helsinki satelites	Municipalities of Hyvinkää, Järvenpää, Kerava. Riihimäki, Kirkkonummi Nurmijärvi, Sipoo, Tuusula, Vihti
Rest of Uusimaa	Municipalites in Uusimaa not in the above classes
Itä-Uusimaa	
Porvoo	
Rest of Itä-Uusimaa	
Varsinais-Suomi	
Turku 1	postal code areas 20100, 20110, 20120, 20140, 20500, 20520, 20700, 20900, 20960
Turku 2	postal code areas 20200, 20300, 20320, 20380, 20540, 20720, 20740, 20810, 20880
Turku 3	rest of Turku
Rest of Varsinais-Suomi	
Satakunta	
Pori 1	postal code areas 28100, 28130, 28360, 28500, 28660
Pori 2	rest of Pori
Rauma	
Rest of Satakunta	
Kanta-Häme	
Hämeenlinna 1	postal code areas 13100, 13130, 13200
Hämeenlinna 2	rest of Hämeenlinna
Rest of Kanta-Häme	
Pirkanmaa	
Tampere 1	postal code areas 33100, 33180, 33200, 33210, 33230, 33240, 33250, 33270, 33500, 33700
Tampere 2	postal code areas 33340, 33400, 33520, 33530, 33540, 33560, 33580, 33610, 33730, 34240
Tampere 3	Rest of Tampere
Rest of Pirkanmaa	

Regional classes

REGION	Explanation
Päijät-Häme	
Lahti 1	postal code areas 15100, 15110, 15140, 15150, 15800, 15850, 15900, 15950
Lahti 2	Rest of Lahti
Rest of Päijät-Häme	
Kymenlaakso	
Kotka 1	postal code areas 48100, 48210, 48310, 48600
Kotka 2	Rest of Kotka
Kouvola	
Rest of Kymenlaakso	
South Karelia	
Lapeenranta 1	postal code areas 53100, 53500, 53600, 53900
Lapeenranta 2	Rest of Lappeenranta
Rest of South Karelia	
Etelä-Savo	
Mikkeli 1	postal code areas 50100, 50170
Mikkeli 2	Rest of Mikkeli
Rest of Etelä-Savo	
Pohjois-Savo	
Kuopio 1	postal code areas 70100, 70110, 70300, 70600, 70800
Kuopio 2	Rest of Kuopio
Rest of Pohjois-Savo	
North Karelia	
Joensuu 1	postal code areas 80100, 80110, 80120, 80200
Joensuu 2	rest of Joensuu
Rest of North Karelia	
Central Finland	
Jyväskylä 1	postal code areas 40100, 40200, 40500, 40520, 40600, 40700, 40720
Jyväskylä 2	Rest of Jyväskylä
Rest of Central Finland	
South Ostrobothnia	
Seinäjoki	
Rest of South Ostrobothnia	

Regional classes

REGION	Explanation
Ostrobothnia	
Vaasa 1	postal code areas 65100, 65170, 65200, 65280
Vaasa 2	Rest of Vaasa
Rest of Ostrobothnia	
Central Ostrobothnia	
Kokkola	
Rest of Central Ostrobothnia	
North Ostrobothnia	
Oulu 1	postal code areas 90100, 90120, 90140, 90160, 90230, 90240, 90420, 90500
Oulu 2	Rest of Oulu
Rest of North Ostrobothnia	
Kainuu	
Kajaani	
Rest of Kainuu	
Lapland	
Rovaniemi	
Rest of Lapland	

Regional regression results

Location-dummy coefficients not reported, estimated coefficients' t-values in brackets

REGION	Intercept	floor area	square root of floor area	age	square root of age	terraced house	1 room	at least 3 rooms	terraced house with at least three rooms	obs. number	R^2
Uusimaa											
Helsinki 1	8,4526	0,0022	-0,0615	0,0042	-0,0672	0,1377	-0,0370	0,0707	..	1589	8,52
	(57,4807)	(2,2522)	(-3,2159)	(2,0184)	(-2,2051)	(0,4944)	(-1,8269)	(3,0757)	(..)		
Helsinki 2	8,4978	0,0056	-0,1196	0,0103	-0,1544	0,1975	0,0315	0,0246	-0,0619	4577	21,77
	(88,7757)	(8,7236)	(-10,3605)	(7,7070)	(-8,7179)	(4,7663)	(2,9154)	(2,1621)	(-1,3598)		
Helsinki 3	8,0525	0,0035	-0,1063	0,0065	-0,0854	0,0829	0,0544	-0,0047	0,0780	1469	20,69
	(47,7223)	(2,2582)	(-3,5352)	(1,9625)	(-2,4849)	(1,9939)	(2,2272)	(-0,2588)	(1,8009)		
Helsinki 4	8,9158	0,0093	-0,2288	0,0066	-0,1183	0,2018	0,0247	0,0761	0,0057	1731	37,28
	(51,7877)	(5,2549)	(-6,9548)	(1,7192)	(-3,0845)	(6,0716)	(1,0538)	(4,8819)	(0,1683)		
Espoo 1	7,7374	-0,0022	0,0171	-0,0045	0,0146	0,1218	0,0969	-0,0416	0,0265	621	12,43
	(32,5284)	(-1,3771)	(0,4356)	(-0,8397)	(0,2848)	(1,2122)	(2,0256)	(-1,1297)	(0,2661)		
Espoo 2	8,3672	0,0041	-0,1075	0,0021	-0,0629	0,1439	0,0475	-0,0231	0,0188	1294	17,96
	(38,9194)	(2,1413)	(-2,6288)	(0,5651)	(-1,9049)	(4,5338)	(1,3462)	(-0,9057)	(0,5562)		
Espoo 3	8,3602	0,0034	-0,1148	0,0119	-0,1865	0,1651	0,0492	-0,0243	0,0799	1328	35,66
	(59,2724)	(3,0884)	(-4,5545)	(4,0732)	(-6,7383)	(5,5412)	(1,7787)	(-1,1704)	(2,5233)		
Vantaa 1	8,8044	0,0064	-0,1777	0,0108	-0,1843	0,1084	0,0008	0,0500	0,0010	1150	43,52
	(51,7542)	(3,4080)	(-4,9791)	(3,0824)	(-5,7524)	(4,8438)	(0,0323)	(3,0439)	(0,0422)		
Vantaa 2	7,9092	0,0006	-0,0625	0,0105	-0,1628	0,1356	0,1193	-0,0031	0,1086	1619	51,93
	(52,9975)	(0,3676)	(-2,0557)	(2,7457)	(-4,8641)	(7,9416)	(6,3724)	(-0,2435)	(5,5959)		
Greater Helsinki satelites	8,2310	0,0058	-0,1611	0,0028	-0,0950	0,1311	0,0336	0,0394	0,0354	3730	46,40
	(90,7931)	(6,0297)	(-8,8368)	(1,9386)	(-6,9244)	(11,6125)	(2,3986)	(3,4509)	(2,5039)		
Rest of Uusimaa	8,2102	0,0057	-0,1259	0,0190	-0,2575	-0,0186	0,0437	-0,0130	0,0593	1022	28,49
	(37,7199)	(2,5735)	(-3,0240)	(8,4549)	(-9,7189)	(-0,7135)	(1,2701)	(-0,4227)	(1,6869)		

Location-dummy coefficients not reported, estimated coefficients' t-values in brackets

REGION	Intercept	floor area	square root of floor area	age	square root of age	terraced house	1 room	at least 3 rooms	terraced house with at least three rooms	obs. number	R^2
Itä-Uusimaa											
Porvoo	8,7661	0,0152	-0,2948	0,0122	-0,1613	0,1738	0,0148	0,0566	-0,0307	494	25,62
	(21,9291)	(2,8794)	(-3,2643)	(5,4236)	(-5,5140)	(3,4080)	(0,2744)	(1,3939)	(-0,5238)		
Rest of Itä-Uusimaa	8,7369	0,0187	-0,3488	0,0081	-0,1435	-0,0083	-0,0931	-0,0538	0,0475	182	49,96
	(9,9966)	(1,5991)	(-1,7573)	(1,1297)	(-1,8837)	(-0,1338)	(-0,8861)	(-0,6780)	(0,5843)		
Varsinais-Suomi											
Turku 1	8,6170	0,0121	-0,2312	0,0119	-0,1797	0,0122	-0,0233	-0,0197	0,1245	1283	11,83
	(48,4669)	(7,1950)	(-7,9579)	(4,5182)	(-5,0623)	(0,3640)	(-0,8960)	(-0,7741)	(2,2144)		
Turku 2	7,7106	0,0035	-0,1088	0,0016	-0,0653	0,1920	0,0631	0,0614	0,0422	1556	29,18
	(54,0711)	(2,1937)	(-3,8792)	(0,8197)	(-2,9476)	(7,3385)	(2,7740)	(3,5788)	(1,2685)		
Turku 3	8,2115	0,0098	-0,1909	0,0341	-0,3150	0,4558	0,1484	-0,0273	-0,0572	748	52,55
	(27,6072)	(2,7817)	(-3,1467)	(5,4021)	(-4,9962)	(7,1530)	(3,8929)	(-1,2463)	(-0,9003)		
Rest of Varsinais-Suomi	7,3407	0,0029	-0,0624	0,0139	-0,2025	0,1016	0,0965	-0,0069	0,0233	2660	42,57
	(58,7301)	(2,0110)	(-2,4161)	(8,3391)	(-12,1398)	(7,1318)	(4,9549)	(-0,4348)	(1,2099)		
Satakunta											
Pori 1	8,6091	0,0123	-0,2676	0,0078	-0,1490	-0,1447	-0,0829	0,0655	0,0255	504	34,01
	(35,0146)	(3,9495)	(-5,0191)	(3,7935)	(-6,0867)	(-3,2149)	(-2,0772)	(2,1496)	(0,4717)		
Pori 2	7,3672	0,0029	-0,1017	0,0054	-0,1136	0,2229	0,0634	0,0175	0,0046	463	62,20
	(27,7431)	(0,9636)	(-1,8315)	(1,5629)	(-3,3754)	(6,0616)	(1,4445)	(0,5164)	(0,1132)		
Rauma	8,8942	0,0109	-0,2988	0,0091	-0,1805	0,1792	-0,0799	0,0986	0,0135	562	46,40
	(36,0511)	(4,0962)	(-6,1610)	(3,1086)	(-5,3734)	(5,2398)	(-2,3861)	(3,5560)	(0,3336)		
Rest of Satakunta	7,5030	-0,0026	-0,0208	0,0243	-0,3013	0,0793	-0,0006	-0,0046	0,0243	644	36,69
	(29,3257)	(-0,7758)	(-0,3725)	(6,2230)	(-8,0955)	(2,8408)	(-0,0133)	(-0,1127)	(0,5589)		

Location-dummy coefficients not reported, estimated coefficients' t-values in brackets

REGION	**Intercept**	**floor area**	**square root of floor area**	**age**	**square root of age**	**terraced house**	**1 room**	**at least 3 rooms**	**terraced house with at least three rooms**	**obs. number**	R^2
Kanta-Häme											
Hämeenlinna 1	8,6169	0,0135	-0,2464	0,0125	-0,1817	-0,0457	0,0163	-0,0169	0,0791	402	25,48
	(24,8227)	(2,7358)	(-3,2070)	(2,5834)	(-3,3533)	(-0,6579)	(0,4494)	(-0,5280)	(0,9654)		
Hämeenlinna 2	8,7571	0,0147	-0,3191	0,0103	-0,1601	0,1702	0,0253	0,0055	0,0164	376	38,49
	(21,9160)	(3,0579)	(-3,6824)	(1,8170)	(-3,1358)	(4,5906)	(0,4407)	(0,1296)	(0,3540)		
Rest of Kanta-Häme	9,1726	0,0172	-0,3340	0,0320	-0,4089	0,0650	-0,0588	0,0236	-0,0126	753	40,36
	(36,4045)	(5,2210)	(-6,2077)	(10,4060)	(-12,1066)	(2,4961)	(-1,7843)	(0,7943)	(-0,3700)		
Pirkanmaa											
Tampere 1	8,5128	0,0071	-0,1366	0,0096	-0,1617	0,0141	0,0221	0,0300	0,0158	1388	20,72
	(64,5512)	(4,9605)	(-5,3924)	(5,7167)	(-7,7099)	(0,3432)	(1,1193)	(1,5378)	(0,3261)		
Tampere 2	8,2038	0,0055	-0,1529	-0,0048	-0,0299	0,0942	0,0233	0,0197	0,0756	873	30,22
	(39,4086)	(2,0635)	(-3,3134)	(-1,9346)	(-1,2267)	(4,0602)	(0,8537)	(0,8749)	(2,7361)		
Tampere 3	8,2637	0,0093	-0,2201	-0,0039	-0,0571	0,1439	0,0479	-0,0161	0,1051	1450	49,62
	(48,3486)	(4,3004)	(-5,8530)	(-1,5679)	(-2,4634)	(6,1249)	(2,1527)	(-1,0105)	(3,9392)		
Rst of Pirkanmaa	7,5275	0,0052	-0,1389	0,0025	-0,1038	0,0879	0,0050	-0,0032	0,0549	2183	50,25
	(57,4496)	(3,1860)	(-4,8723)	(15,9682)	(-17,5797)	(5,8668)	(0,2339)	(-0,1872)	(2,6916)		

Location-dummy coefficients not reported, estimated coefficients' t-values in brackets

REGION	Intercept	floor area	square root of floor area	age	square root of age	terraced house	1 room	at least 3 rooms	terraced house with at least three rooms	obs. number	R^2
Päijät-Häme											
Lahti 1	9,1421	0,0118	-0,2398	0,0190	-0,2757	-0,0315	0,0341	0,0297	0,1326	824	38,98
	(41,6582)	(4,7540)	(-5,3748)	(6,9217)	(-9,3927)	(-0,4255)	(1,0977)	(1,1155)	(1,6312)		
Lahti 2	8,5159	0,0100	-0,2648	0,0004	-0,1217	0,3184	0,0040	0,1005	-0,0493	939	46,20
	(32,8558)	(3,7193)	(-5,2678)	(0,0817)	(-2,2693)	(7,9989)	(0,1218)	(4,6192)	(-1,1866)		
Rest of Päijät-Häme	9,3542	0,0091	-0,2280	0,0647	-0,6952	0,1526	-0,0251	0,0882	0,0003	932	39,67
	(30,4840)	(2,4525)	(-3,5046)	(13,2692)	(-14,5708)	(5,8864)	(-0,6152)	(3,0496)	(0,0083)		
Kymenlaakso											
Kotka 1	8,4259	0,0110	-0,2103	0,0126	-0,2369	0,1136	-0,0266	0,0708	-0,1273	459	48,97
	(40,2471)	(4,1949)	(-4,8893)	(4,8404)	(-7,6890)	(2,2476)	(-1,0304)	(2,6567)	(-2,1997)		
Kotka 2	7,5557	0,0077	-0,1328	0,0063	-0,1572	0,3170	0,1093	0,0005	-0,0561	379	55,26
	(19,9532)	(1,5918)	(-1,6820)	(1,2971)	(-2,8005)	(8,1728)	(2,3872)	(0,0157)	(-1,2019)		
Kouvola	8,7534	0,0159	-0,3340	0,0091	-0,1547	0,1977	-0,0658	0,0740	0,0105	618	41,55
	(35,0641)	(5,0245)	(-6,0677)	(2,2979)	(-3,9982)	(5,7641)	(-1,9680)	(2,6445)	(0,2482)		
Rest of Kymenlaakso	7,4279	-0,0004	-0,0233	0,0171	-0,2658	0,0807	0,0552	0,0103	0,0040	669	34,68
	(28,4597)	(-0,1197)	(-0,4175)	(4,8462)	(-7,6821)	(2,9780)	(1,4448)	(0,3309)	(0,1110)		

Location-dummy coefficients not reported, estimated coefficients' t-values in brackets

REGION	Intercept	floor area	square root of floor area	age	square root of age	terraced house	1 room	at least 3 rooms	terraced house with at least three rooms	obs. number	R^2
South Karelia											
Lapeenranta 1	7,7583	-0,0046	-0,0145	0,0032	-0,0923	0,1735	0,0204	0,0617	-0,1115	342	44,05
	(29,3881)	(-1,0592)	(-0,2162)	(0,8323)	(-2,4478)	(4,0525)	(0,6554)	(2,1004)	(-1,9414)		
Lapeenranta 2	8,2109	0,0041	-0,1365	0,0173	-0,2337	0,2063	0,1058	0,0626	0,0192	409	53,08
	(43,7080)	(2,1244)	(-3,7710)	(4,2710)	(-6,1983)	(6,6697)	(3,8539)	(2,8198)	(0,5193)		
Rest of South Karelia	8,0555	0,0047	-0,1498	0,0055	-0,1598	0,0512	-0,0179	0,0524	-0,0256	532	33,50
	(35,1647)	(1,9769)	(-3,5713)	(1,0277)	(-3,0289)	(1,7595)	(-0,5394)	(1,5876)	(-0,5955)		
Etelä-Savo											
Mikkeli 1	8,3496	0,0023	-0,0990	0,0142	-0,2482	-0,0653	0,0387	0,0744	-0,0627	324	37,55
	(20,0116)	(0,3704)	(-0,9886)	(2,9395)	(-5,0128)	(-1,5949)	(0,9247)	(2,0383)	(-1,1291)		
Mikkeli 2	8,7160	0,0142	-0,3167	0,0256	-0,2818	0,2405	0,0410	0,0140	0,0808	235	71,18
	(24,2725)	(2,9440)	(-3,8420)	(6,0995)	(-7,4901)	(9,1316)	(0,7934)	(0,4959)	(2,4234)		
Rest of Etelä-Savo	8,3568	0,0066	-0,1379	0,0334	-0,4344	-0,0532	0,0412	0,0170	0,0253	855	44,80
	(30,1191)	(1,8673)	(-2,3387)	(7,3737)	(-9,8228)	(-1,9636)	(1,1513)	(0,5202)	(0,6805)		
Pohjois-Savo											
Kuopio 1	8,7004	0,0102	-0,1891	0,0132	-0,2062	0,1115	0,0662	0,0164	-0,0167	522	34,73
	(33,5732)	(3,1575)	(-3,4043)	(4,3878)	(-6,5158)	(1,4401)	(1,9074)	(0,5804)	(-0,1688)		
Kuopio 2	8,7187	0,0103	-0,2528	0,0118	-0,1974	0,0815	0,0116	0,0598	0,0689	875	37,87
	(43,2519)	(4,5838)	(-6,1711)	(3,2413)	(-5,2635)	(3,7150)	(0,4196)	(3,1626)	(2,6560)		
Rest of Pohjois-Savo	7,9399	0,0041	-0,1136	0,0225	-0,3285	0,0617	0,0454	0,0179	0,0219	1190	44,77
	(36,1873)	(1,3817)	(-2,3455)	(6,7413)	(-10,0042)	(2,8906)	(1,6084)	(0,6725)	(0,6947)		

Location-dummy coefficients not reported, estimated coefficients' t-values in brackets

REGION	Intercept	floor area	square root of floor area	age	square root of age	terraced house	1 room	at least 3 rooms	terraced house with at least three rooms	obs. number	R^2
North Karelia											
Joensuu 1	9,0558	0,0068	-0,1735	0,0439	-0,4839	0,0884	0,0391	0,0385	-0,1335	497	38,49
	(33,0866)	(2,1695)	(-3,1364)	(10,7781)	(-11,6006)	(1,9170)	(1,0812)	(1,2040)	(-2,2305)		
Joensuu 2	8,3111	0,0030	-0,1023	0,0238	-0,3266	0,0857	0,1054	0,0104	0,0317	473	36,59
	(29,1436)	(1,1461)	(-2,1080)	(2,4210)	(-3,6940)	(3,2680)	(2,0298)	(0,3054)	(0,8395)		
Rest of North Karelia	8,6790	0,0106	-0,2100	0,0308	-0,4473	-0,0794	0,0720	-0,0145	0,0727	587	60,56
	(29,5470)	(2,4166)	(-3,0486)	(6,1366)	(-9,6787)	(-3,1634)	(2,0154)	(-0,4034)	(1,9110)		
Central Finland											
Jyväskylä 1	8,5322	0,0041	-0,1386	0,0155	-0,1894	0,1020	-0,0169	0,0186	0,0283	749	33,30
	(48,6114)	(2,0917)	(-4,0410)	(5,3693)	(-6,4640)	(2,5464)	(-0,7148)	(0,8488)	(0,6402)		
Jyväskylä 2	7,7619	0,0066	-0,1608	-0,0049	-0,0344	0,0902	0,1339	-0,0170	0,0611	617	51,90
	(31,4065)	(2,2174)	(-3,0551)	(-0,7490)	(-0,5862)	(3,3190)	(3,9205)	(-0,7696)	(1,9359)		
Rest of Central Finland	7,6565	0,0033	-0,0879	0,0140	-0,2345	0,0268	0,0758	-0,0469	0,1215	1263	52,62
	(37,6393)	(1,1464)	(-1,8754)	(5,4466)	(-9,3423)	(1,4664)	(2,8471)	(-1,9835)	(4,5049)		
South Ostrobothnia											
Seinäjoki	7,3753	0,0024	-0,0617	-0,0066	-0,0614	0,0352	0,1951	-0,0676	0,0585	530	51,24
	(39,3275)	(1,1610)	(-1,5943)	(-1,6046)	(-1,7304)	(1,2045)	(5,0220)	(-2,2964)	(1,7017)		
Rest of South Ostrobothnia	7,5005	0,0003	-0,0549	0,0068	-0,2012	0,0408	0,0504	-0,0214	0,0918	546	50,74
	(21,4272)	(0,0659)	(-0,6717)	(0,8960)	(-3,2988)	(1,0524)	(0,8955)	(-0,4003)	(1,6837)		

Location-dummy coefficients not reported, estimated coefficients' t-values in brackets

REGION	**Intercept**	**floor area**	**square root of floor area**	**age**	**square root of age**	**terraced house**	**1 room**	**at least 3 rooms**	**terraced house with at least three rooms**	**obs. number**	R^2
Ostrobothnia											
Vaasa 1	8,7148	0,0164	-0,2957	-0,0025	-0,0448	-0,0866	0,0473	0,0072	0,0252	604	25,22
	(28,8658)	(3,9973)	(-4,3476)	(-0,9345)	(-1,3960)	(-1,0994)	(1,2253)	(0,2170)	(0,3757)		
Vaasa 2	7,2113	-0,0029	0,0297	0,0056	-0,1283	0,1479	0,1688	-0,0479	0,0861	271	58,84
	(24,7737)	(-1,0006)	(0,5105)	(1,1101)	(-2,5975)	(4,0761)	(3,8250)	(-1,5599)	(2,0426)		
Rest of Ostrobothnia	8,4055	0,0052	-0,1398	0,0319	-0,4171	-0,0573	-0,1055	-0,0029	0,0732	504	41,90
	(27,8476)	(1,2315)	(-2,0437)	(6,8906)	(-8,7460)	(-1,6841)	(-2,4176)	(-0,0926)	(1,7767)		
Central Ostrobothnia											
Kokkola	8,6136	0,0123	-0,2591	0,0005	-0,1348	0,0450	0,0294	-0,0295	0,0197	315	52,66
	(21,3655)	(2,3161)	(-2,9427)	(0,0708)	(-2,0829)	(1,0571)	(0,5598)	(-0,8132)	(0,3922)		
Rest of Central Ostrobothnia	5,2200	-0,0487	0,7228	0,0432	-0,5132	-0,1518	0,1689	0,0214	0,1283	85	52,69
	(4,4142)	(-2,4417)	(2,4477)	(1,6039)	(-2,3563)	(-1,4229)	(1,0151)	(0,1381)	(0,8935)		
North Ostrobothnia											
Oulu 1	8,6655	0,0109	-0,2380	0,0097	-0,1435	0,1618	-0,0088	0,0433	-0,0266	1036	39,17
	(53,9337)	(5,1466)	(-6,8460)	(4,7422)	(-6,8104)	(5,8588)	(-0,3983)	(2,3134)	(-0,8654)		
Oulu 2	9,0177	0,0152	-0,3386	0,0011	-0,0901	0,1590	0,0025	0,0393	0,0273	775	53,65
	(49,1405)	(6,5665)	(-8,3341)	(0,4427)	(-4,1431)	(8,3531)	(0,0972)	(1,9584)	(1,1680)		
Rest of North Ostrobothnia	7,7337	0,0065	-0,1636	-0,0038	-0,0872	0,0323	0,0581	0,0670	-0,0463	1092	44,62
	(39,8353)	(2,4867)	(-3,7195)	(-0,8969)	(-2,5974)	(1,3968)	(1,6232)	(2,0513)	(-1,3433)		

Location-dummy coefficients not reported, estimated coefficients' t-values in brackets

REGION	Intercept	floor area	square root of floor area	age	square root of age	terraced house	1 room	at least 3 rooms	terraced house with at least three rooms	obs. number	R^2
Kainuu											
Kajaani	8,2091	0,0058	-0,1778	0,0093	-0,1656	0,1369	0,0355	0,0748	-0,0199	481	50,80
	(24,1465)	(1,1581)	(-2,2275)	(1,9089)	(-3,7949)	(4,1515)	(0,9028)	(2,6442)	(-0,5185)		
Rest of Kainuu	7,6235	-0,0110	0,1258	0,0479	-0,5194	-0,2089	0,0673	0,0775	-0,0300	132	60,07
	(8,0249)	(-0,6971)	(0,5314)	(3,5941)	(-4,6881)	(-3,8469)	(0,7318)	(0,8084)	(-0,3214)		
Lapland											
Rovaniemi	9,9691	0,0236	-0,4525	0,0265	-0,3386	0,1114	-0,0413	0,0672	-0,0265	534	62,05
	(32,4412)	(5,7603)	(-6,5675)	(4,9239)	(-6,4847)	(4,0094)	(-1,1472)	(2,5083)	(-0,7812)		
Rest of Lapland	9,3431	0,0225	-0,4455	0,0264	-0,3289	0,0571	-0,0766	0,0559	-0,0430	720	40,70
	(30,3036)	(5,5991)	(-6,7542)	(4,0675)	(-5,3695)	(1,7001)	(-1,6219)	(1,3998)	(-0,9712)		

Average yearly index point numbers and quality adjustments, overall index and major Finnish municipalities, 1987-2000, 2000=100

total hedonic adjustment=(adjustment due to floorarea)x(adjustment due to location)x(adjustment due to construction year)

Index, classification+hedonic quality adjustment =(Classification index)x(total hedonic Adjustment)

	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Finland														
Classification index	61,14	79,61	97,83	93,56	81,15	66,79	61,25	64,90	62,71	64,71	76,66	84,63	92,99	100,00
Adjustment due to floor area	1,0005	0,9999	0,9987	0,9982	0,9989	1,0002	0,9999	1,0001	1,0006	1,0007	1,0003	1,0005	1,0010	1,0000
Adjustment due to location	0,9977	0,9990	1,0024	1,0031	1,0002	0,9977	0,9968	0,9980	0,9965	0,9983	0,9988	1,0004	0,9995	1,0000
Adjustment due to construction year (vintage)	1,0202	1,0127	1,0083	1,0107	1,0060	0,9982	0,9968	0,9981	0,9986	0,9987	0,9986	0,9981	0,9981	1,0000
total hedonic Adjustment	1,0184	1,0117	1,0093	1,0120	1,0052	0,9961	0,9936	0,9962	0,9957	0,9977	0,9977	0,9990	0,9986	1,0000
Index, classification+hedonic quality adjustment	62,26	80,54	98,75	94,69	81,57	66,53	60,86	64,65	62,44	64,56	76,49	84,54	92,86	100,00
Greater Helsinki (Helsinki, Espoo, Vantaa)														
Classification index	58,98	79,52	97,05	90,48	74,69	58,23	54,89	60,32	57,36	59,35	73,05	81,18	90,56	100,00
Adjustment due to floor area	0,9991	0,9992	0,9974	0,9967	0,9975	0,9997	0,9999	0,9995	1,0002	0,9998	0,9993	0,9996	1,0005	1,0000
Adjustment due to location	0,9964	0,9962	1,0009	1,0024	1,0005	0,9972	0,9962	0,9966	0,9960	0,9981	0,9948	1,0012	0,9984	1,0000
Adjustment due to construction year (vintage)	1,0095	1,0094	1,0067	1,0040	1,0014	0,9968	0,9976	0,9981	0,9979	0,9983	0,9990	0,9968	0,9973	1,0000
total hedonic Adjustment	1,0050	1,0048	1,0050	1,0031	0,9994	0,9938	0,9937	0,9942	0,9941	0,9962	0,9931	0,9977	0,9962	1,0000
Index, classification+hedonic quality adjustment	59,28	79,90	97,53	90,77	74,64	57,87	54,54	59,97	57,02	59,13	72,54	80,99	90,22	100,00
Finland without Greater Helsinki														
Classification index	63,04	79,68	98,51	96,28	87,11	75,08	67,25	69,08	67,68	69,67	79,89	87,69	95,12	100,00
Adjustment due to floor area	1,0018	1,0006	0,9997	0,9995	1,0002	1,0006	1,0000	1,0007	1,0009	1,0015	1,0012	1,0012	1,0014	1,0000
Adjustment due to location	0,9988	1,0014	1,0036	1,0036	1,0000	0,9982	0,9974	0,9991	0,9969	0,9985	1,0023	0,9997	1,0005	1,0000
Adjustment due to construction year (vintage)	1,0294	1,0156	1,0097	1,0165	1,0100	0,9994	0,9962	0,9981	0,9992	0,9991	0,9982	0,9992	0,9988	1,0000
total hedonic Adjustment	1,0299	1,0176	1,0131	1,0196	1,0101	0,9981	0,9936	0,9979	0,9970	0,9991	1,0017	1,0001	1,0006	1,0000
Index, classification+hedonic quality adjustment	64,93	81,09	99,79	98,17	87,99	74,94	66,82	68,94	67,48	69,60	80,03	87,71	95,18	100,00
Helsinki														
Classification index	57,87	77,55	94,39	87,84	71,67	56,18	54,13	59,93	56,81	58,98	72,35	80,11	89,61	100,00
Adjustment due to floor area	1,0005	1,0012	1,0000	0,9987	0,9994	1,0010	1,0005	1,0006	1,0008	1,0009	1,0007	1,0004	1,0013	1,0000
Adjustment due to location	1,0017	1,0017	1,0070	1,0067	1,0059	1,0000	0,9994	0,9996	0,9989	1,0019	0,9962	1,0035	0,9991	1,0000
Adjustment due to construction year (vintage)	1,0011	1,0011	0,9989	0,9977	0,9972	0,9946	0,9948	0,9958	0,9950	0,9963	0,9967	0,9944	0,9955	1,0000
total hedonic Adjustment	1,0032	1,0041	1,0059	1,0030	1,0025	0,9955	0,9947	0,9961	0,9947	0,9991	0,9936	0,9983	0,9958	1,0000
Index, classification+hedonic quality adjustment	58,05	77,87	94,95	88,10	71,85	55,93	53,84	59,70	56,51	58,93	71,88	79,98	89,23	100,00
Espoo														
Classification index	60,39	81,08	100,25	94,69	79,79	62,33	56,24	61,86	59,29	60,41	74,38	83,25	92,45	100,00
Adjustment due to floor area	0,9944	0,9925	0,9912	0,9909	0,9915	0,9956	0,9976	0,9960	0,9981	0,9965	0,9954	0,9982	0,9991	1,0000
Adjustment due to location	0,9857	0,9851	0,9882	0,9934	0,9879	0,9904	0,9878	0,9884	0,9895	0,9885	0,9905	0,9972	0,9968	1,0000
Adjustment due to construction year (vintage)	1,0215	1,0211	1,0188	1,0136	1,0101	1,0011	1,0034	1,0018	1,0029	1,0021	1,0046	1,0015	0,9997	1,0000
total hedonic Adjustment	1,0013	0,9983	0,9979	0,9978	0,9894	0,9871	0,9888	0,9862	0,9905	0,9870	0,9904	0,9969	0,9956	1,0000
Index, classification+hedonic quality adjustment	60,46	80,94	100,04	94,48	78,94	61,53	55,61	61,01	58,72	59,62	73,67	82,99	92,05	100,00
Vantaa														
Classification index	63,02	88,69	107,26	98,58	83,66	63,01	56,83	59,64	56,88	59,51	74,62	83,52	92,62	100,00
Adjustment due to floor area	1,0001	0,9999	0,9940	0,9965	0,9976	1,0000	1,0007	0,9992	1,0006	0,9995	0,9983	0,9977	0,9988	1,0000
Adjustment due to location	0,9868	0,9861	0,9900	0,9951	0,9940	0,9944	0,9936	0,9949	0,9918	0,9950	0,9951	0,9958	0,9975	1,0000
Adjustment due to construction year (vintage)	1,0362	1,0360	1,0295	1,0230	1,0096	1,0020	1,0033	1,0043	1,0054	1,0026	1,0021	1,0021	1,0030	1,0000
total hedonic Adjustment	1,0226	1,0214	1,0130	1,0144	1,0011	0,9964	0,9975	0,9984	0,9977	0,9971	0,9956	0,9956	0,9993	1,0000
Index, classification+hedonic quality adjustment	64,45	90,59	108,66	99,99	83,76	62,78	56,69	59,54	56,75	59,34	74,29	83,15	92,55	100,00

Average yearly index point numbers and quality adjustments, overall index and major Finnish municipalities, 1987-2000, 2000=100

total hedonic adjustment=(adjustment due to floorarea)x(adjustment due to location)x(adjustment due to construction year)

Index, classification+hedonic quality adjustment =(Classification index)x(total hedonic Adjustment)

Tampere														
Classification index	53,90	73,40	89,95	83,46	73,42	61,64	56,15	60,41	59,24	61,69	75,43	83,74	93,39	100,00
Adjustment due to floor area	1,0033	1,0013	1,0007	0,9971	1,0016	1,0015	1,0021	1,0017	1,0013	1,0018	1,0015	1,0013	1,0010	1,0000
Adjustment due to location	0,9992	0,9989	1,0010	1,0101	1,0027	1,0008	1,0020	1,0026	1,0001	1,0061	1,0035	1,0063	1,0034	1,0000
Adjustment due to construction year (vintage)	1,0255	1,0268	1,0200	1,0150	1,0082	0,9987	0,9946	1,0026	1,0040	1,0035	1,0019	0,9985	0,9993	1,0000
total hedonic Adjustment	1,0281	1,0270	1,0217	1,0223	1,0126	1,0010	0,9987	1,0069	1,0053	1,0114	1,0068	1,0061	1,0037	1,0000
Index, classification+hedonic quality adjustment	55,42	75,37	91,90	85,32	74,34	61,71	56,07	60,82	59,55	62,39	75,95	84,25	93,73	100,00
Turku														
Classification index	65,88	85,35	107,64	102,95	90,33	74,68	68,90	72,30	68,46	69,96	81,94	89,06	96,39	100,00
Adjustment due to floor area	0,9979	0,9967	0,9966	0,9975	1,0032	1,0016	1,0005	1,0015	1,0016	1,0027	1,0027	1,0009	1,0025	1,0000
Adjustment due to location	0,9971	0,9993	1,0013	1,0089	1,0022	1,0027	1,0004	1,0011	0,9991	1,0024	1,0003	0,9983	0,9977	1,0000
Adjustment due to construction year (vintage)	1,0103	1,0086	1,0072	1,0024	0,9974	0,9955	0,9934	0,9954	0,9963	0,9940	0,9945	0,9948	0,9951	1,0000
total hedonic Adjustment	1,0053	1,0045	1,0051	1,0088	1,0029	0,9997	0,9943	0,9980	0,9970	0,9991	0,9975	0,9940	0,9953	1,0000
Index, classification+hedonic quality adjustment	66,23	85,73	108,18	103,86	90,59	74,66	68,50	72,16	68,25	69,90	81,73	88,52	95,93	100,00
Oulu														
Classification index	52,44	68,81	80,09	79,64	76,27	70,24	62,45	63,53	62,09	66,92	78,44	88,17	94,51	100,00
Adjustment due to floor area	1,0081	1,0058	1,0042	1,0055	1,0031	1,0040	1,0052	1,0080	1,0058	1,0066	1,0060	1,0054	1,0062	1,0000
Adjustment due to location	0,9980	1,0016	1,0015	1,0029	1,0028	1,0049	1,0036	1,0027	1,0058	0,9978	1,0086	1,0039	1,0087	1,0000
Adjustment due to construction year (vintage)	1,0341	1,0330	1,0285	1,0256	1,0199	1,0119	1,0116	1,0075	1,0085	1,0095	1,0051	1,0027	1,0006	1,0000
total hedonic Adjustment	1,0404	1,0406	1,0344	1,0342	1,0260	1,0210	1,0205	1,0183	1,0203	1,0140	1,0197	1,0120	1,0156	1,0000
Index, classification+hedonic quality adjustment	54,56	71,61	82,84	82,37	78,25	71,71	63,73	64,69	63,35	67,85	79,99	89,23	95,98	100,00

95 percent confidence interval of the quarterly change, 1^{st}-2^{nd} quarter 1995

Region	Index based on classification and hedonic quality adjustment					Index based on classification				
	2.5 % point, normal approximation	2.5 % point, simulation	97.5 % point, simulation	97.5 % point, normal approximation	length of the 95 % confidence bound, simulation	2.5 % point, normal approximation	2.5 % point, simulation	97.5 % point, simulation	97.5 % point, normal approximation	length of the 95 % confidence bound, simulation
Finland	-1,50	-1,46	0,03	0,06	1,49	-1,79	-1,79	-0,11	-0,13	1,67
Finland without Greate	-0,85	-0,83	0,81	0,80	1,64	-1,02	-0,94	0,70	0,68	1,64
Greater Helsinki	-2,95	-2,86	-0,03	-0,10	2,84	-3,38	-3,39	-0,35	-0,39	3,04
Uusimaa	-2,70	-2,57	-0,05	-0,16	2,53	-3,16	-3,12	-0,37	-0,45	2,75
Helsinki	-3,69	-3,67	0,03	-0,06	3,69	-4,28	-4,22	-0,39	-0,43	3,82
Espoo	-4,16	-3,96	3,01	2,90	6,96	-4,34	-4,14	2,83	2,81	6,97
Vantaa	-3,49	-3,33	1,15	1,10	4,48	-3,85	-3,70	1,43	1,35	5,13
Greater Helsinki satelit	-4,45	-4,34	0,09	0,20	4,44	-4,76	-4,65	0,74	0,79	5,39
Itä-Uusimaa	-9,36	-7,61	8,99	10,24	16,60	-8,44	-7,29	14,60	14,26	21,89
Porvoo	-9,07	-8,92	7,02	6,93	15,95	-7,47	-6,86	11,31	10,39	18,18
Varsinais-Suomi	-2,07	-1,98	2,15	2,09	4,13	-1,44	-1,37	2,82	2,80	4,20
Turku	-2,56	-2,64	2,96	2,87	5,61	-1,95	-1,81	3,84	3,65	5,65
Satakunta	-2,38	-2,19	4,38	4,11	6,57	-2,45	-2,28	4,45	4,56	6,73
Pori	-1,15	-0,77	8,10	8,07	8,87	-2,16	-2,13	8,47	8,43	10,60
Rauma	-5,24	-5,33	7,18	7,37	12,52	-7,38	-7,38	5,04	5,40	12,42
Kanta-Häme	-4,34	-4,24	3,05	2,89	7,29	-6,34	-6,03	2,40	1,87	8,43
Hämeenlinna	-5,92	-5,79	2,48	2,52	8,27	-7,70	-7,43	2,59	2,54	10,03
Pirkanmaa	-2,03	-2,00	2,08	1,89	4,08	-2,65	-2,61	1,52	1,42	4,14
Tampere	-1,47	-1,38	3,09	2,87	4,47	-1,93	-1,74	3,01	2,88	4,75
Päijät-Häme	-2,29	-2,20	4,26	4,48	6,46	-0,87	-0,85	6,21	6,20	7,06
Lahti	-1,18	-1,55	7,80	7,59	9,35	0,11	-0,04	9,66	9,48	9,70
Kymenlaakso	-5,98	-5,88	0,78	0,86	6,66	-6,33	-6,19	1,72	1,55	7,91
Kotka	-11,17	-11,20	-1,20	-1,12	9,99	-14,03	-14,05	-2,13	-1,87	11,92
Kouvola	-10,87	-10,94	1,29	0,88	12,23	-10,17	-9,75	3,83	3,63	13,58

95 percent confidence interval of the quarterly change, 1st-2nd quarter 1995

Region	Index based on classification and hedonic quality adjustment				
	2.5 % point, normal approximation	2.5 % point, simulation	97.5 % point, simulation	97.5 % point, normal approximation	length of the 95 % confidence bound, simulation
South Karelia	-9,19	-8,97	-1,18	-1,31	7,79
Lappeenranta	-11,20	-11,12	-0,01	-0,25	11,11
Etelä-Savo	-7,40	-7,33	0,79	0,79	8,12
Mikkeli	-13,13	-12,90	0,33	0,08	13,23
Pojhois-Savo	-1,73	-1,55	3,93	3,67	5,49
Kuopio	-4,92	-4,65	1,90	1,65	6,55
North Karelia	-3,71	-3,37	4,76	4,55	8,13
Joensuu	-6,12	-6,20	3,77	3,65	9,97
Central Finland	-2,42	-2,36	1,98	2,16	4,34
Jyväskylä	-3,55	-3,73	1,50	1,51	5,23
South Ostrobothnia	-2,80	-2,60	6,76	6,51	9,36
Seinäjoki	-8,16	-7,82	5,32	4,48	13,14
Ostrobothnia	0,73	0,77	9,99	9,78	9,22
Vaasa	1,09	0,95	13,22	12,90	12,27
Central Osthrobothni	-3,38	-1,10	14,35	16,27	15,45
Kokkola	-1,78	-1,54	14,51	15,43	16,05
North Osthrobothnia	-1,70	-1,61	3,91	3,89	5,52
Oulu	-1,45	-1,43	5,67	5,33	7,10
Kainuu	-11,71	-10,73	3,07	1,99	13,79
Kajaani	-9,67	-9,41	1,36	0,87	10,77
Lapland	-4,57	-4,11	7,83	6,94	11,94
Rovaniemi	-5,00	-4,82	7,43	7,62	12,25

Region	Index based on classification				
	2.5 % point, normal approximation	2.5 % point, simulation	97.5 % point, simulation	97.5 % point, normal approximation	length of the 95 % confidence bound, simulation
South Karelia	-8,85	-8,46	-0,66	-0,84	7,81
Lappeenranta	-8,96	-8,97	1,79	1,85	10,76
Etelä-Savo	-7,31	-7,05	1,77	1,64	8,82
Mikkeli	-13,74	-13,25	1,75	1,76	15,00
Pojhois-Savo	-2,12	-1,87	4,29	4,02	6,16
Kuopio	-4,96	-4,49	2,97	2,47	7,47
North Karelia	-4,44	-4,10	4,76	4,68	8,86
Joensuu	-7,37	-7,45	3,26	3,13	10,71
Central Finland	-2,65	-2,66	2,44	2,53	5,10
Jyväskylä	-5,50	-5,55	0,70	0,60	6,25
South Ostrobothnia	-3,95	-3,67	6,46	6,15	10,13
Seinäjoki	-10,04	-9,70	4,90	4,30	14,59
Ostrobothnia	-0,07	0,12	9,91	9,58	9,79
Vaasa	0,60	0,32	14,22	13,46	13,90
Central Osthrobothni	-3,72	-0,53	16,78	19,84	17,31
Kokkola	0,68	1,06	20,85	20,78	19,78
North Osthrobothnia	-2,57	-2,76	3,74	3,79	6,50
Oulu	-1,59	-1,58	6,60	6,24	8,18
Kainuu	-14,06	-13,72	0,64	-0,04	14,37
Kajaani	-14,16	-13,77	-1,99	-2,53	11,78
Lapland	-8,66	-8,45	2,86	2,38	11,31
Rovaniemi	-8,43	-8,21	6,05	6,33	14,26