Owner Occupied Housing

Final report on land price separation study

INE, Portugal for Eurostat

To be presented at the TAG Meeting

Global Office

2nd Technical Advisory Group Meeting
February 17-19, 2010
Washington DC
Luxembourg, March 2009

Working Group
"Harmonization of Consumer Price Indices"
to be held in Luxembourg

Thursday and Friday 26-27 March 2009
at 9.30 a.m., Room Quetelet

Item 5 of the Agenda

Final report on land price separation study

Paper by INE, Portugal for Eurostat

12/03/2009
Final report on land price separation study

Eurostat's introduction

From both Eurostat's and the main users' points of view, a harmonised treatment of the Owner Occupied Housing (OOH) and its prospective inclusion to the HICP would improve the comparability and reliability of the measurement of inflation. Therefore Eurostat has been running a pilot on OOH in which the net acquisition approach has been tested.

In the approach, one important element is dwellings prices, which are entered in the index. However, for some uses of house price indices it has been seen desirable to construct indices that would only track the prices of the structures and thus exclude the land price component. Thus Eurostat funded a study in which the objective was to find out if it is possible and how to separate the effect of land prices in house price index data using econometric techniques in the case when land prices cannot be directly observed. The result of this study called "Owner-Occupied Housing - Econometric Study and Model to Estimate Land Prices" is included in the form of a final report of the contractor.

In particular, the study was launched to provide answers to the following questions:

1) Is separation of the land price component and the structure component theoretically feasible and on what grounds?
2) What different approaches are available for the separation?
3) Which of the available method(s) suggested under 2) could be utilised for ongoing practical index compilation and why (statistical quality, costs, timeliness, reproducibility, etc.) and moreover what are the prerequisites of the implementation?
4) Does the specified econometric model produce reasonable (shadow) price data for the land component?
5) What are the minimum data requirements for the model?

One core outcome of this study was that, with the available data and the specific methods used for this study, the separation the land and structure from the compound price of dwelling does not at present seem feasible by using econometric modelling. In particular the treatment of location, so whether treated as a characteristic of the structure or of the land component, is critical for the model results. Thus the interactions between the prices or monetary values of structures, location and building plot requires further reflection.
Table of contents:

Introduction ..................................................................................................................... 2
I. Literature review ........................................................................................................ 2
II. Alternative approaches ............................................................................................ 4
III. Data requirements .................................................................................................... 5
References..................................................................................................................... 5

Econometric model for separating a housing price index into land and structure components .................................................................................................................... 6
1. Econometric model for separating a housing price index into land and structure components.................................................................................................................... 7
   1.1. Derivation of the regression model ....................................................................... 7
   1.2. Re-formulating the model to the case of apartments in blocks of flats ............... 12
   1.3. The structure-land decomposition and derivation of index weights.................... 13
   1.4. Pitfalls.................................................................................................................. 16
2. The single-family houses ....................................................................................... 17
   2.1 Data ....................................................................................................................... 17
   2.2 Computing a price index for structure using Statistics Finland’s price indices on single family houses and land plots for single-family houses ........................................ 17
   2.3 Hedonic regressions for the single-family houses ................................................ 20
      2.3.1. The simple model ......................................................................................... 20
      2.3.2. Hedonic model with characteristics affecting unit prices of land and structure ........................................................ ........................................ 24
      2.3.3 The adjacent period model ............................................................................ 31
      2.3.4 The Diewert specification .............................................................................. 34
   2.4 Summary of the findings and conclusions ............................................................ 35
3. Apartments in blocks of flats in Helsinki .............................................................. 38
   3.1 Data sources and Definitions ............................................................................. 38
   3.2. Dwellings on leased land plot ............................................................................ 39
   3.3 Data description .................................................................................................... 40
   3.4 Empirical results for the Helsinki apartments on blocks of flats ....................... 43
4. Summary and Conclusions .................................................................................... 50
References: ................................................................................................................... 52

Discussion of the Q&A session in the HICP meeting at Vienna on the 8th and 9th of September 2008 ................................................................................................... 53

Extensions to the Draft Final Report .......................................................................... 57
I. Discussion on the weight of the land component .................................................... 58
II. Alternative econometric approaches: simpler models ........................................... 58
III. Generalisation of the models ................................................................................ 61
   3.1 A simple generalisation ......................................................................................... 61
   3.2 A more complex generalisation .......................................................................... 68
IV. Conclusions ............................................................................................................ 73
References: ................................................................................................................... 74
Introduction

House prices may be seen as being comprised of two fundamental components, the structure price and the land price (e.g. where the structure sits on). For the purpose of compiling housing price indices, many times it is useful to have two separate indices referring to the components just mentioned. One such case is the incorporation of a housing index in HICP. Purchasing land may be looked at as an investment whereas purchasing the structure itself may be seen as “pure” consumption.

Nevertheless, land prices are not often directly available, especially in the case of apartments in blocks of flats. Without these data, it becomes difficult to disentangle the two effects in house prices and compute directly the required sub-indices.

The present study intends to assess the possibility of calculating such indices by means of indirect methods, based on hedonic regression techniques. To attain the objective, a theoretical framework is developed and an empirical study is conducted, using data on Finnish dwelling transactions.

The report is organised as follows. In the next section, brief summaries of related studies are presented. The second section contains an overview on alternative approaches on how to build indices for the present purpose. The third section presents the data requirements. The annex contains the stand-alone paper with the results of the econometric study that was carried out with data provided by Statistics Finland.

I. Literature review

Separating land prices from house prices has been a recent topic of discussion mainly in Statistical Offices. There are many reasons for doing this separation, the most important ones being connected to policy decision and making. Final purposes, though, are usually diverse in the literature reviewed and although this is an important topic (especially over the last years, with soaring interest rates turning even more attention to the housing market), there is not much research available; especially with the goals we have in mind. For this reason, only a brief review will be presented here.

1. The Price and Quantity of Residential Land in the United States

The authors (Davis and Heathcote, 2007) start by defining the nominal market value of residential land at date \( t \), \( p^l_t \), as the difference between the market value of the housing stock, \( p^h_t \), and the replacement cost of the stock of residential structures, \( p^s_t \):

\[
p^l_t = p^h_t - p^s_t
\]

In the equation, the prices are quality-adjusted per unit (of land, houses and structure) and the quantities are also quality-adjusted. Thus the label “land” is attached to anything that makes a house more valuable other than the replacement cost, adjusted for depreciation, of the physical structure.
It is considered that the growth rate of house prices is simply a weighted average of the growth rates of structure and land prices, assuming the stock of structure and land remains fixed. This being the case, the growth rate of land prices may be written as

\[
\frac{p_{lt+1}^l}{p_{lt}^l} = \frac{1}{w_l^t} \left[ p_{lt+1}^h \left( 1 - w_l^t \right) + p_{lt+1}^s \right]
\]  

(1)

The authors are then interested in applying equation (1) to the entire stock of housing in the United States with the intent of producing a residential land price index. In order to do that, they use index series for the house and structure prices and from those derive one for the weight of land prices on house prices.

The data are quarterly. For structures, the authors use the price index for gross investment in new residential structures produced by the Bureau of Economic Analysis (BEA) within the National Income and Product Accounts (NIPA). For house prices, they use the repeat-sales-based index produced by the Office of Federal Housing Enterprise Oversight (OFHEO). A discussion on whether this index is appropriate is then presented. Afterwards, based on these two indices and more figures from BEA, the 2000 Decennial Census of Housing (DCH) and the 2001 Residential Finance Survey (RFS), an estimate for \( w_l \) is derived. With all the figures at hand the residential land price index is estimated.

2. Construction Price Indices and House and Property Price Indices 2006

Statistics Norway publishes three different indices in the housing market: output and seller’s selling price indices for new residential buildings and house price indices. The index most relevant to the present study is the price index for new detached houses. It is based on hedonic regression methods and land value is excluded.

A survey on investors is conducted, inquiring total purchase price, land cost and relevant characteristics of the dwelling. The sample for the survey is built based on register information of GAB (Ground Property, Address and Building Register). Therefore, the index excluding land value is compiled based on an internally calculated price.

The issue of land in urban area dwelling transactions (mainly flats) is not addressed.

3. Studies in Hedonic Resale Housing Price Indexes

The New Housing Price Index (NHPI) is a quarterly housing price index deflated of land value (Li et al, 2006). It is built using data acquired from a survey on contractors, reflecting the evolution of selling prices of new residential houses. The survey also collects contractors’ estimates of the current cost of land. These estimates allow the compilation of structure and land price indices, which are subsequently quality adjusted.

The goal of the paper is to assess the quality of the NHPI (including land). It is argued that NHPI is based on a small sample and that the quality adjustment is done by rather crude means. RHPI (Resale Housing Price Index) uses a bigger sample and quality
adjusts the index by hedonic methods\textsuperscript{1}. RHPI uses a sample of MLS (Multiple Listing Service) data on Ottawa transactions of new and old houses. The index built using this information is then compared with NHPI for Ottawa and the former is thought to be more accurate than the latter. The paper provides some interesting figures for the present research, for instance the correlation coefficient between lot area and selling price – which stands to be around 10% in Ottawa.

**II. Alternative approaches**

Apart from econometric modelling, there are two feasible alternatives. The first one is to use data on land prices: a direct approach not discussed in depth here as the goal of the present study is exactly to assess alternatives to direct approaches. The other possibility is to build an implicit land price index. In fact, this method is discussed under particular circumstances in the paper by Davis and Heathcote (2007), discussed previously.

_A direct approach_

This approach relies almost entirely on data of land prices, either based on real land transactions or on estimated land costs. Having access to such data, building up an index is then rather straightforward.

To our knowledge, there are no registries containing information on land prices for all dwelling transactions. Therefore, the simplest way to proceed is to constitute a sample of transactions and inquire for land share on the total house price. The sample may be of real estate agents or contractors, depending on national specificities. It may even be a sample of registered transactions (based on the fiscal registries, for example). A survey should then be conducted, inquiring for relevant variables. However, the land fraction on total cost is not, for obvious reasons, an adequate question to be made to home-acquirers, thus the preference for a sample based on real estate agents. Other sources might be appraisal institutions or firms, if there are any.

With estimates for land values on transactions, building two distinct indices is straightforward and care should only be exercised in order to control for compositional change in the sample and varying quality.

An alternative is to consider the possibility of using a SPAR index. This index relies on appraisals. It would be fairly easy to implement in the appraisal system a new variable stating an estimated proportion of the total appraisal that could be attributed to land. Appraisals would then be implicitly constituted by two components and, as a consequence, the two separate indices could be directly compiled. Of course, this approach depends beforehand on the evaluation of adequateness of a SPAR index to measure the evolution of OOH prices.

_Another indirect approach: implicit land price index_

\textsuperscript{1} We have not been able to know with certainty what kind of quality adjustment is implicit in the NHPI.
This is an obvious alternative. Considering there are house price indices and structure price indices, one can easily compute a land price index – it is basically an implicit index deriving from the two former ones. The main difficulty in this approach is to have an adequate structure price index. Such an index, in case it exists, is obtained using direct information, rendering this approach irrelevant (cf. a direct approach). On the other hand, one could use a proxy index. At first, a construction price index might seem reasonable. However, the index does not reflect the builder or contractor’s margin fee, among other amounts, turning the index into a rather crude option.

**III. Data requirements**

Data needs vary with the method chosen. The econometric approach is highly demanding – it is necessary to have information on the house characteristics, namely the dwelling size as well as the land plot size. The latter variable is usually not easily obtained and matching of databases is often required. The main difficulty rests on having this information for highly populated urban areas. The empirical section ahead will highlight all these issues.

On the other hand, even a direct approach requires information on many variables due mainly to the fact that structure and land indices must also be controlled for quality. If the quality adjustment is done by means of hedonic regression, then only the land plot size variable may be dropped from the list.

The particular case of the SPAR index is different though. Data needed are total transaction prices and the two component appraisals. For the indirect methods, data needed depends on the quality of already existent indices – good proxies make data requirements small, otherwise the method is also demanding.

**References**


Econometric model for separating a housing price index into land and structure components
1. Econometric model for separating a housing price index into land and structure components

Traditional house price indices measure the change of the price of a (fixed quality) dwelling. The overall price change measured can be, at least notionally, attributed to changes of the price of the structure of the dwelling (that is the price of the building) and changes of the price of land upon which the dwelling is built.

This is a problem for a consumer price index, such as HICP, which is supposed to measure “household final monetary consumption expenditure”, since price levels and price changes of a dwelling will, presumably, depend on both land and structure prices, and while a dwelling structure is essentially a durable good, land is an asset and therefore expenditure on land is not consumption expenditure but investment.

There are several different approaches which, in principle, can be used to separate dwelling and land prices. The most obvious is to construct a separate land price index and use it, together with a standard dwelling price index, to obtain an “implicit” index on dwelling structure. Another possible method, used for example by Statistics Canada, is to use expert valuations.

A third alternative suggested by Diewert (2007) is to construct an appropriate hedonic price index. This paper presents in detail a theoretical method which can be used for this purpose. The starting point is the idea of Diewert (2007) but, as discussed in the paper, the econometric model outlined there seems to have some serious drawbacks. This being the case, the specification suggested in this paper is different. In sections 1, 2 and 3 of the present chapter we develop the hedonic model and derive an overall Laspeyres-type aggregate index for dwellings together with its decomposition into price and structure components. Section 4 is a discussion of the pitfalls of the method.

1.1. Derivation of the regression model

The starting point for the model is the obvious idea that the price of a dwelling can be written as

\[ P_T = p^S S + p^L L \]

\( P_T \) is the total dwelling price, \( p^S \) and \( p^L \) denote unit prices (e.g. prices per square metre) for structure and land respectively, and \( S \) and \( L \) denote the number of structure and land units comprising the dwelling.

Expression (1) simply states that the total price of a dwelling is the sum of the structure price and the land price. Suppose now that there is a data set containing information on sales prices, size of structure and size of the land plot upon which the dwelling is built. Let’s assume for the moment that the observations cover transactions of similar, single-family homes, situated in the same geographical area, e.g. Helsinki. Essentially the assumption is that dwellings differ with respect to the size of both the structure and the
land plot and are similar enough in other respects. Observations are available for a base period, referred to as period 0, and for $T$ subsequent periods. There are a total of $N(t)$ observations for each period $t = 0, \ldots, T$. Diewert (2007) turns equation (1) into an econometric model by assuming multiplicative errors with constant variances, that is:

\[
P_n^t = \left( p_0^n S_{n0}^t + p_0^n L_{n0}^t \right) \eta_{nt}, \quad n = 1 \ldots N(0), \quad \text{for } t = 0 \text{ (the base period)}
\]

\[
P_n^t = \left( \lambda_n p_0^n S_{nt}^t + \psi_n p_0^n L_{nt}^t \right) \eta_{nt}, \quad n = 1 \ldots N(t), \quad \text{for } t = 1, \ldots, T
\]

Here the subindex $n$ refers to the $n^{th}$ observation, the subindex $t$ to the period during which the transaction took place, $\eta_{nt}$ is the multiplicative error term, $\lambda_n$ and $\psi_n$ are the parameters of interest, describing the price ratios of structure and land unit prices between the initial (base) period 0 and period $t$ for $t = 1, \ldots, T$. Taking logarithms on both sides of (2) leads to:

\[
\ln(P_{n0}^t) = \ln \left( p_0^n S_{n0}^t + p_0^n L_{n0}^t \right) + \xi_{n0}, \quad \xi_{n0} = \ln(\eta_{n0}), \quad \text{for } t = 0 \text{ (the base period)}
\]

\[
\ln(P_{nt}^t) = \ln \left( \lambda_n p_0^n S_{nt}^t + \psi_n p_0^n L_{nt}^t \right) + \xi_{nt}, \quad \xi_{nt} = \ln(\eta_{nt}), \quad \text{for } t = 1, \ldots, T
\]

Diewert (2007) suggests (3) to be estimated as a system of nonlinear regression equations.

There are two particularities about the above specification that we would reconsider. First, transforming (2) into (3) turns a linear model into nonlinear one. The need for such transformation is dictated by the multiplicative error assumption in (2). While we accept the argument underlying the assumption – it is likely that the absolute size of the error term increases with the total price of the property – we think that re specifying (2) to include an additive (and heteroskedastic) error term is a better option to proceed with. In our view, (3) may unduly complicate the estimation procedure, and may even make the whole exercise unsustainable from an econometric point of view, as will be shown later on.

Second, there is no need to estimate a system of equations if the following reparametrisation is made:

\[
\delta_t = \lambda_t - 1 \quad \text{and} \quad \gamma_t = \psi_t - 1 \quad \text{for } t = 1, \ldots, T
\]

The new parameters $\delta_t$ and $\gamma_t$ are relative price-changes and not price-ratios. A value of $\lambda_t$ equal to 1.1 corresponds to a value of $\delta_t$ equal to 0.1. Both mean the same thing: the unit price for structure has increased by 10% between the base period and period $t$.

Now, starting from (1) but assuming an additive error term and using (4), the single-equation for dwellings sold in $T+1$ different periods becomes:

---

2 Equation (2) corresponds to (21) and (23) in Diewert (2007) and equation (3) to (22) and (24) in the same source. The notation here is slightly different.
\[ P_{nt}^T = \left( \sum_{i=0}^{T} \delta_i D_i + 1 \right) P_{0s}^S S_{nt} + \left( \sum_{i=0}^{T} \gamma_i D_i + 1 \right) P_{0l}^L L_{nt} + \varepsilon_{nt}, \quad t = 0, \ldots, T \]

where \( \varepsilon_{nt} \) is a heteroskedastic additive error term.

Equation (5) is still nonlinear in the set of parameters \((\delta_1, \ldots, \delta_T, \gamma_1, \ldots, \gamma_T, p_{0s}^S, p_{0l}^L)\) but linear in the alternative parameter set \((\delta_1 p_{0s}^S, \ldots, \delta_T p_{0s}^S, \gamma_1 p_{0l}^L, \ldots, \gamma_T p_{0l}^L, p_{0s}^S, p_{0l}^L)\) as seen from rewriting (5) as:

\[ (1.5') \quad P_{nt}^T = \left( \sum_{i=0}^{T} \frac{\delta_i p_{0s}^S}{\alpha_i} D_i + P_{0s}^S \right) S_{nt} + \left( \sum_{i=0}^{T} \frac{\gamma_i p_{0l}^L}{\beta_i} D_i + P_{0l}^L \right) L_{nt} + \varepsilon_{nt}, \quad t = 0, \ldots, T \]

Supposing that, except for the homoskedasticity of the error term, all other standard assumptions for that type of models can be plausibly made, one may conclude that ordinary least squares (OLS) will provide unbiased and consistent estimates of the unknown parameters of (5'), that is \(\delta_1, \ldots, \delta_T, \gamma_1, \ldots, \gamma_T, p_{0s}^S, p_{0l}^L\). By dividing throughout the OLS-estimates of \(\delta_i p_{0s}^S \ldots \delta_T p_{0s}^S\) by the OLS-estimate of \(p_{0s}^S\) and the OLS-estimates of \(\gamma_i p_{0l}^L \ldots \gamma_T p_{0l}^L\) by the OLS-estimate of \(p_{0l}^L\), one will obtain consistent estimates of the original set of parameters \(\delta_1, \ldots, \delta_T, \gamma_1, \ldots, \gamma_T\). These estimates should be numerically identical to the nonlinear regression estimates obtained directly from (5). However, standard linear regression output enables correct inference (p-values, confidence intervals) only for the parameter set \((\delta_1, \ldots, \delta_T, \gamma_1, \ldots, \gamma_T, p_{0s}^S, p_{0l}^L)\). For correct inference about the set \((\delta_1, \ldots, \delta_T, \gamma_1, \ldots, \gamma_T, p_{0s}^S, p_{0l}^L)\) one should apply nonlinear least squares directly to (5).

The heteroskedasticity of the error term implies that to obtain correct variances for the estimates one should use for example the method suggested by White (1980).

Equation (5) is a modification of the traditional time dummy variable hedonic regression. In this form the model does not include any other characteristics of the dwellings except for structure and plot size. This is probably not satisfactory, because both structure and land price are expected to depend on a number of characteristics other than size. Since the purpose is to obtain a “pure price index”, such characteristics should be included in the model, so that the time dummy coefficients do not erroneously reflect changes in the mix of characteristics of the dwellings sold in different time periods. The starting point for this extension is to write the base-period unit prices of structure and land for the \(n^{th}\) dwelling sold in period \(t, t = 0, \ldots, T\) as:

\[ p_{nt}^S(0) = \exp(\alpha^T x_{nt}^S) + \sigma_{nt} \quad \text{and} \quad p_{nt}^L(0) = \exp(\beta^T x_{nt}^L) + \nu_{nt} \]

where \(x_{nt}^S\) is a vector of characteristics related to the structure of the dwelling such as number of rooms, indicators of availability or absence of some facilities (type of heating, etc.) and \(\alpha\) is the parameter vector describing the relation of these characteristics to the unit price of structure. Similarly, \(x_{nt}^L\) is a vector of characteristics describing the quality of the plot, such as micro-location indicators, distance from city...
centre, availability of public transports, sea-view, etc and \( \beta \) is the corresponding parameter vector. \( \sigma_{nt} \) is the error term related to structure unit price and \( \nu_{nt} \) the error related to land unit price.

It is important to note that the vectors \( x^S_{nt} \) and \( x^L_{nt} \) are likely to have common elements because some characteristics are likely to affect both unit prices of land and structure. The logic is obvious: expensive structures are likely to be built on expensive locations, meaning that for example micro-location indicators can enter both the equation for \( p^S_{nt}(0) \) and for \( p^L_{nt}(0) \).

The exponential form assumed in the unit price equations is important – it ensures that, for all possible combinations of the parameter vectors and the regressors, the estimated unit prices of land and structure are positive.

The next step is to substitute (6) into (5), which after re-arranging the right-hand side terms becomes:

\[
P^T_{nt} = \left( \sum_{i=1}^{r} \delta_i D_i + 1 \right) \frac{\exp(\alpha' x^S_{nt}) S_{nt}}{p^S_{nt}(0)} + \left( \sum_{i=1}^{r} \gamma_i D_i + 1 \right) \frac{\exp(\beta' x^L_{nt}) L_{nt}}{p^L_{nt}(0)} + \left( \sum_{i=1}^{r} \delta_i D_i + 1 \right) S_{nt} \sigma_{nt} + \left( \sum_{i=1}^{r} \gamma_i D_i + 1 \right) L_{nt} \nu_{nt} + \epsilon_{nt}
\]

(1.7)

Equation (7) is a nonlinear model with composite error term \( \theta_{nt} \), which is clearly heteroskedastic. Still, under standard assumptions, it can be estimated consistently by nonlinear least squares. To obtain correct confidence intervals and valid tests for the estimated parameters one should use heteroskedasticity-corrected covariance matrix estimates of the type suggested by White (1980). Both nonlinear least squares estimation and automatic computation of heteroskedasticity-corrected covariance matrices are nowadays available in specialised statistical software packages such as STATA® and TSP®. The SAS/STAT®, which is the main statistical software used by Statistics Finland and also seems to be commonly used by other statistical agencies, includes an otherwise sophisticated and user-friendly nonlinear least squares estimation procedure, but does not provide an option for computing the heteroskedasticity-corrected covariance matrix for the nonlinear regression.

As noted above the vectors of explanatory variables \( x^S_{nt} \) and \( x^L_{nt} \) most probably contain common elements. Erroneous exclusion of variables affecting both structure and land prices either from the structure unit price or the land unit price equation will most likely result in violation of the assumption of orthogonality between regressors and error terms, which will render the NLS-estimates of (7) inconsistent. Inconsistency of the parameter estimates will, in turn, lead to incorrect decomposition of the total dwelling price into structure and land components as well. To minimise this possible

---

3 The assumptions under which NLS is consistent, when the errors are independent and identically distributed (i.i.d.) and hence homoskedastic, are discussed, e.g., in Davidson & MacKinnon (1993), pp. 145-153. NLS consistency when the i.i.d. assumption does not hold is described on pp. 298-299.
problem, one may consider including all explanatory variables in both structure and land unit price equations.

From an econometric point of view, the possibility to develop model (7) starting from model (5) and using the unit price specification (6) is the biggest difference between the method developed in this paper and the method outlined in Diewert (2007). Using the reparametrisation of the time dummy variables presented in (4) to be able to write Diewert’s system of equations (3) as a single-equation model and inserting (6) into it one obtains:

\[
\ln(P_{nt}^T) = \ln\left(\sum_{i=1}^{T} \delta_i D_i + 1\right) \exp(\alpha'x_{nt}^S)S_{nt} + \left(\sum_{i=1}^{T} \gamma_i D_i + 1\right) \exp(\beta'x_{nt}^L)S_{nt} + \xi_{nt}
\]

(1.8)

In equation (8) the error term \(\phi_{nt}\) is inside the logarithm function and there is no way to obtain a composite additive error term corresponding to \(\theta_{nt}\) in equation (7). As a result, one cannot apply standard econometric theory to (8). In Diewert (2007) the problem is circumvented by simply assuming that there are no error terms in the unit price equations. This means that whatever the explanatory variables included in these equations, unit prices for land and structure will be completely determined by them. It is hard to consider such an assumption plausible.

In model (7), the parameter vectors \(\alpha\) and \(\beta\) are assumed to be time-invariant. The interpretation of the assumption is that there is no variation of the relative importance of the structure (land) characteristics for structure (land) unit price formation over time. Unit prices are assumed to change proportionally, but the rate of change can differ for structure and land as captured by the parameters \(\delta_i\) and \(\gamma_i\). In principle this parameter time-invariance assumption may be relaxed. This will lead to a more complicated decomposition of the dwelling price index into land and structure components. We will not discuss in detail the implications of such an extension here.

It is not plausible that price developments are the same for different dwelling types and especially for different geographical areas because both supply and demand factors can vary a lot within a country. Provided that the data used is representative enough, a single equation of the type (7) applied to the whole data set is likely to provide good estimates for the development of the land and structure prices “on average” if an appropriate set of dummy variables is included in the unit price equations to control for regional and dwelling-type price differences. However, it will not be possible in general to account for how exactly the potentially different regional and dwelling-type price developments are aggregated into an overall index (a measure of average price change).

\^4 In Diewert (2007), the specified relationship between unit prices and characteristics corresponding to (6) in the present paper is linear. This of course does not matter for the present argument.
So it still may be better to estimate separate equations like (7) for different geographical regions and perhaps for different dwelling types and aggregate them by an index formula. The geographical stratification should be solved based on knowledge about the differences in regional housing markets. Single-family homes should be anyway treated separately from apartments in blocks of flats. We discuss these issues in more detail in the following sections.

1.2. Re-formulating the model to the case of apartments in blocks of flats

The model developed in section 1 of the present chapter assumes that there is one dwelling built on each land plot, which is the case for single-family houses. Separating land and structure price in the case of blocks of flats creates extra data requirements and complicates estimation. This point is clarified in the discussion below.

An apartment in a block of flats will account for only a small fraction of the total bundle formed by the total residential structure and the land upon which it is built. In this case there is an extra question to be answered compared to the case of single family houses. Namely, how the total price of the land on which the block of flats is built is divided among the individual apartments. A line of reasoning is that, in a given building, bigger flats “use” larger fraction of land so that the total land price will be distributed among the apartments in proportion to their size. However, this is not the only possible, and perhaps not even a “fair”, mechanism. The value of a land plot very much depends on its location and a particular location will provide all owners the same environment quality irrespective of how big flat they own. So, it may be the case that the total land price is distributed equally among all apartments. Between the two limiting cases – land price distributed in proportion to apartment’s size and land price distributed equally – there is infinite number of possibilities and in reality we are most likely somewhere in the middle.

If, alongside with information on the size of the apartment transacted and the (total) size of the land plot, there is also information on both total number of dwellings in the building and total structure size (i.e., total dwelling space of all apartments) the model from equation (7) can be modified for the blocks of flats case. Let the subindex \( J \) denote a particular land plot. There are \( M_J \) dwellings built on that plot with total dwelling space \( S_J \). Let subindex \( nJt \) denote the \( n \)th apartment on the land plot \( J \), transacted in period \( t \). Then, (7) can be modified for example in the following way:

\[
P_{nJt}^T = \left( \sum_{i=1}^{T} \delta_i D_i + 1 \right) \exp(\alpha' x^S_{nJt}) S_{nJt} + p_{nJt}(0) \sum_{i=1}^{T} \gamma_i D_i + 1 \left( \exp(\beta' x^L_{nJt}) \left( \frac{1}{M_J} + (1 - \tau) \frac{S_{nJt}}{S_J} \right) L_{nJt} + \theta_{nJt} \right), \quad 0 \leq \tau \leq 1
\]

(1.9)

where \( \theta_{nJt} \) is a composite heteroskedastic error with a similar, but not exactly the same, structure as \( \theta_{nt} \) in equation (7). The parameter \( \tau \) measures to what extent total land
price is distributed among the apartments equally and/or according to their size. τ=0 means that total land price is distributed purely in proportion to apartment size and τ=1 that total land price is equally distributed among apartments in the building.

A typical observation in Finnish data is that the unit price in apartments in blocks of flats tends to diminish with apartment’s size. This may reflect both “economies of scale” in construction costs and relatively larger “land price share” of small dwellings.

Obviously, to estimate (9) total dwelling space or some other relevant measure of total structure size and total number of dwellings in the building should be known alongside with the size of the land plot and the dwelling space (or another relevant structure size measure) of the dwelling transacted and its transaction price. If there is information only on total structure size or on total number of dwellings, one does not have a choice but, depending on which variable is known, assume τ=0 or τ=1 and try if it will work. If there is no information on either total structure size or number of dwellings a meaningful model cannot be specified, because there is no plausible way to obtain an estimable measure of land price fraction allocated to the apartment sold.

1.3. The structure-land decomposition and derivation of index weights

It is taken as a starting point for the discussion in this section that dwellings are stratified by geographical region and perhaps by dwelling type. There are K mutually exclusive classes and a separate model of the type (7) has been estimated for each class. Let the symbol \( \hat{\cdot} \) denote estimated parameters and values in equations (5) to (7) and the subscript \( k \) a class in the classification.

For any particular property \( A \) in class \( k \) with structure size \( S_A \), plot size \( L_A \), structure-characteristics vector \( x_A^S \) and land-characteristics vector \( x_A^L \), the estimated total price for the base period 0 is

\[
\hat{P}_0^T = \hat{P}_{k0}^{S_A} S_A + \hat{P}_{k0}^{L_A} L_A
\]

(1.10) where :

\[
\hat{P}_{k0}^{S_A} = \exp(\hat{\alpha}_{k0} x_A^S) \text{ and } \hat{P}_{k0}^{L_A} = \exp(\hat{\beta}_{k0} x_A^L)
\]

Respectively for some subsequent period (the comparison period) \( t \in 1, \ldots, T \), the estimated price for the same property is

\[
\hat{P}_t^T = (\hat{\delta}_t + 1) \hat{P}_{k0}^{S_A} S_A + (\hat{\gamma}_t + 1) \hat{P}_{k0}^{L_A} L_A
\]

(1.11) The estimated price ratio between periods 0 and \( t \) is

\[
\left( \frac{\hat{P}_t^T}{\hat{P}_0^T} \right) = \left( \frac{\hat{\delta}_t + 1}{\hat{\gamma}_t + 1} \right)
\]
\[
\frac{\hat{P}_t^{TA}}{\hat{P}_0^{TA}} = (\hat{\delta}_k + 1) \frac{\hat{P}_0^{SA}S_A + \hat{P}_0^{LA}L_A}{\hat{P}_0^{SA}S_A + \hat{P}_0^{LA}L_A} = (\hat{\gamma}_k + 1) \frac{\hat{P}_0^{LA}L_A}{\hat{P}_0^{SA}S_A + \hat{P}_0^{LA}L_A}
\]

(1.12)

So we have the price ratio of a dwelling decomposed into structure and land components. The price of the structure component of the dwelling has changed by \((\hat{\delta}_k + 1)\) between base and comparison periods, and the price of the land component by \((\hat{\gamma}_k + 1)\). The overall price ratio is a weighted average of both price-change components, the weights being each component’s value share in the base period.

Equation (12) describes the decomposition of the total price change for a single dwelling in class \(k\) of the classification into structure and price components. For all dwellings in class \(k\), we now proceed to obtain an “elementary aggregate” index. For that purpose we need a sample of dwellings with fixed characteristics. Normally the sample will be chosen to represent the dwellings transacted in the base period, that is, period 0. Suppose that there is a total of \(N_k(0)\) dwellings transacted in the base period in class \(k\) and \(n_k(0)\) is the subset of transactions available in the sample. The sample-based estimate for the value shares of structure and land in the base period for class \(k\) are

\[
(1.13) \quad W_{0}^{Sk} = \frac{1}{n_k(0)} \sum_{i=1}^{n_k(0)} \frac{\hat{P}_{0}^{SI}S_i}{\hat{P}_{0}^{SI}S_i + \hat{P}_{0}^{LI}L_i} \quad \text{and} \quad W_{0}^{Lk} = \frac{1}{n_k(0)} \sum_{i=1}^{n_k(0)} \frac{\hat{P}_{0}^{LI}L_i}{\hat{P}_{0}^{SI}S_i + \hat{P}_{0}^{LI}L_i}
\]

and the elementary aggregate index in class \(k\), \(I(k)_0\) is:

\[
(1.14) \quad I(k)_0 = \frac{1}{n_k(0)} \sum_{i=1}^{n_k(0)} \frac{\hat{P}_{0}^{TI}}{\hat{P}_{0}^{TI}} = (\hat{\delta}_k + 1)W_{0}^{Sk} + (\hat{\gamma}_k + 1)W_{0}^{Lk}
\]

where \(W_{0}^{Sk}\) and \(W_{0}^{Lk}\) are the value shares of structure and land as defined in (13). This simple result is due to the fact that according to the econometric model specification in section 1, the change of relative unit prices of structure and land are the same for all dwellings in class \(k\).

The final element needed to complete the index is to derive class weights and to decompose the total expenditure on dwellings into consumption and investment share. Then one will be able to obtain appropriate weight by which the index should be included in Eurostat’s HICP (Harmonised Index of Consumer Prices). Obviously expenditure on structure will be viewed as consumption expenditure and expenditure on land as investment expenditure.

Here we describe the procedure for estimating weights and expenditure shares when there is information (based for example on registers or National Accounts data) on the
total expenditure of dwellings for each class. Let’s denote it by $TE_{k0}$. From the sample-based estimation results, the share of consumption expenditure to total expenditure and the share of investment to total expenditure are simply $W_0^{sk}$ and $W_0^{lk}$. Then the estimates of total consumption ($TC_{k0}$) and investment expenditure ($TI_{k0}$) in class $k$ are

$$TC_{k0} = TE_{k0}W_0^{sk} \text{ and } TI_{k0} = TE_{k0}W_0^{lk}$$

Now, the overall dwelling index between periods 0 and $t$, $I'$, is obtained by the following aggregation of the elementary aggregate indices (14)

$$I'_0 = \sum_{k=1}^{K} I(k)'_0 \frac{TE_{k0}}{\sum_{k=1}^{K} TE_{k0}} = \sum_{k=1}^{K} \left( (\delta_{tk} + 1)W_0^{sk} + (\gamma_{tk} + 1)W_0^{lk} \right) \frac{TE_{k0}}{\sum_{k=1}^{K} TE_{k0}}$$

$$= \sum_{k=1}^{K} \left( (\delta_{tk} + 1) \frac{TC_{k0}}{\sum_{k=1}^{K} TC_{k0}} + (\gamma_{tk} + 1) \frac{Tl_{k0}}{\sum_{k=1}^{K} TE_{k0}} \right)$$

(1.16) $$= \sum_{k=1}^{K} \left( (\delta_{tk} + 1) \frac{TC_{k0}}{\sum_{k=1}^{K} TC_{k0}} + (\gamma_{tk} + 1) \frac{Tl_{k0}}{\sum_{k=1}^{K} Tl_{k0}} \right)$$

$$= I(S)'_0 + I(L)'_0$$

where $I(S)'_0$ and $I(L)'_0$ are the aggregate indices of structure and land prices defined as:

$$I(S)'_0 \equiv \sum_{k=1}^{K} \left( (\delta_{tk} + 1) \frac{TC_{k0}}{\sum_{k=1}^{K} TC_{k0}} \right) \text{ and } I(L)'_0 \equiv \sum_{k=1}^{K} \left( (\gamma_{tk} + 1) \frac{Tl_{k0}}{\sum_{k=1}^{K} Tl_{k0}} \right)$$

So we have obtained an overall price index for dwellings, which can be represented as a weighted average of a structure price index and a land price index, the weights being the shares of consumption and investment expenditure in total dwelling expenditure. Only $I(S)'_0$ will be included in HICP and its weight in HICP will be determined by the share of total consumption expenditure on dwellings, $\sum_{k=1}^{K} TC_{k0}$, in total final household consumption monetary expenditure.
1.4. Pitfalls

All usual problems stated in the already vast literature on hedonic regressions apply to the specifications developed in sections 1 and 2. These problems are related to possible biases due to omitted variables, functional form misspecification and the assumption of time-invariance of most parameters in the unit price equations. We will not repeat the arguments here but simply state that despite its shortcomings the hedonic method seems to be in many cases the best option available and is nowadays quite generally accepted also by official statistical agencies. However, the method developed here is intended to tackle a very particular problem and for this reason has some potential problems that are not common to hedonics in general and should be discussed here.

The key feature of the method described above is that it relies on variation in the structure size/land size ratio to identify separate structure and land price movements. To see that, suppose that all dwellings in the sample have the same ratio between the size of structure and size of plot denoted by \( r \). In this case there is a perfect linear relationship between structure and land sizes and the econometric model is not identified. Indeed, note that equation (5), for the dwellings sold in the base period \((T=0)\), can now be rewritten as:

\[
(1.18) \quad p_{n0}^T = p_0^S S_n + p_0^L L_n + \varepsilon_{n0} = p_0^S r L_n + p_0^L L_n + \varepsilon_{n0} = (p_0^S r + p_0^L) L_n + \varepsilon_{n0}
\]

This expression contains two unknown parameters, but only one independent explanatory variable, and there is no way to have an econometric model with unique solution. How serious this problem is can only be gauged once real data are available.
2. The single-family houses

2.1 Data

Statistics Finland uses data from the National Land Survey of Finland to compute the quarterly price indices on single-family houses and land plots for single family houses. We use these data to perform empirical tests for the econometric models developed in chapter 1.

The data cover all real-estate transactions - both on dwellings and other real-estates including construction land plots. It should be noted that in Finland apartments in blocks of flats and terraced houses are not considered real-estate property by law so these transactions are not recorded in this register. Legally the real estate is the land and the buildings on it. The real estate is owned by a “housing-share corporation”. The owner of an apartment in blocks of flats is a shareholder owning the shares, entitling him/her to use or rent a particular dwelling. Some single-family houses are also incorporated in housing-share corporations. However the majority of single-family houses are “normal” real-estates and their transactions are covered by the National Land Survey of Finland register. The data on single-family houses also exclude transactions on houses built on leased land.

The data contain information on the location of the property (municipality, postal-area code and co-ordinates), size of the land plot, size of the single-family house and year of construction.

2.2 Computing a price index for structure using Statistics Finland’s price indices on single family houses and land plots for single-family houses

The index series on both single-family houses and land-plots start on 1985. Both indices are calculated using a combination of regional stratification of the data and hedonic techniques.

Since the end of the 1990’s the prices for land plots have increased much faster than the total price of single-family houses. This implies that structure-component price of single-family houses has increased more slowly than the overall price index for single family houses.

A natural reference point for the empirical tests of the models developed in chapter 1 is to compute an implicit structure-price index using the available indices on total price for single-family houses and the land-plots. Using (1.16) from chapter 1 the (implicit) price index for structure is given by the following formula:

\[
I'_0(S) = \frac{I'_0(T)P^T_0 - I'_0(L)P^L_0}{P^T_0 - P^L_0} = \frac{1}{w_S} (I'_0(T) - w_L I'_0(L))
\]
where $I''_0(S)$ is the structure price-index, $I''_0(T)$ is the total price index for single-family houses and $I''_0(L)$ is the land-price index. $P''_0$ is the (average) total price of a single-family house in the base period and $P''_0$ is the (average) price of the land plot of a single-family house in the base period. $w_s$ is the share of structure of the total price in the base period and $w_l$ is the corresponding share of land. (2.1) actually corresponds to the double-deflation method used to compute price and quantity indices for value added in the National Accounts Statistics.

We have chosen as a base period years 2002 to 2007. Using the data on single-family houses and the statistics on land and single-family house prices, the necessary information on base period prices and weights to apply (2.1) was computed. It is presented in the following table:

<table>
<thead>
<tr>
<th>Year</th>
<th>Quarter</th>
<th>Number of transactions in the data</th>
<th>Average total price of the transacted house, euro</th>
<th>Average plot size of the transacted house, m²</th>
<th>Average price per m²</th>
<th>Average total price of the land plot, euro</th>
<th>Implicit structure price, euro</th>
<th>Land share of total price, %</th>
<th>Structure share of total price, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>1</td>
<td>1 758</td>
<td>105 952</td>
<td>2 254</td>
<td>9.7</td>
<td>21 886</td>
<td>84 085</td>
<td>20.6</td>
<td>79.4</td>
</tr>
<tr>
<td>2002</td>
<td>2</td>
<td>3 015</td>
<td>104 916</td>
<td>2 302</td>
<td>9.5</td>
<td>21 867</td>
<td>83 049</td>
<td>20.8</td>
<td>79.2</td>
</tr>
<tr>
<td>2002</td>
<td>3</td>
<td>2 679</td>
<td>102 614</td>
<td>2 440</td>
<td>8.0</td>
<td>19 521</td>
<td>83 094</td>
<td>19.0</td>
<td>81.0</td>
</tr>
<tr>
<td>2002</td>
<td>4</td>
<td>1 936</td>
<td>107 648</td>
<td>2 153</td>
<td>11.2</td>
<td>24 454</td>
<td>83 193</td>
<td>22.7</td>
<td>77.3</td>
</tr>
<tr>
<td>2003</td>
<td>1</td>
<td>1 785</td>
<td>111 959</td>
<td>2 316</td>
<td>10.1</td>
<td>23 388</td>
<td>88 572</td>
<td>20.9</td>
<td>79.1</td>
</tr>
<tr>
<td>2003</td>
<td>2</td>
<td>2 924</td>
<td>111 413</td>
<td>2 391</td>
<td>9.2</td>
<td>21 996</td>
<td>89 417</td>
<td>19.7</td>
<td>80.3</td>
</tr>
<tr>
<td>2003</td>
<td>3</td>
<td>2 852</td>
<td>111 431</td>
<td>2 419</td>
<td>8.5</td>
<td>20 495</td>
<td>90 936</td>
<td>18.4</td>
<td>81.6</td>
</tr>
<tr>
<td>2003</td>
<td>4</td>
<td>2 210</td>
<td>120 984</td>
<td>2 330</td>
<td>10.9</td>
<td>25 437</td>
<td>95 547</td>
<td>21.0</td>
<td>79.0</td>
</tr>
<tr>
<td>2004</td>
<td>1</td>
<td>1 950</td>
<td>124 793</td>
<td>2 381</td>
<td>9.9</td>
<td>23 602</td>
<td>101 191</td>
<td>18.9</td>
<td>81.1</td>
</tr>
<tr>
<td>2004</td>
<td>2</td>
<td>2 805</td>
<td>123 170</td>
<td>2 411</td>
<td>8.6</td>
<td>20 847</td>
<td>102 323</td>
<td>16.9</td>
<td>83.1</td>
</tr>
<tr>
<td>2004</td>
<td>3</td>
<td>2 831</td>
<td>121 421</td>
<td>2 377</td>
<td>9.0</td>
<td>21 300</td>
<td>100 120</td>
<td>17.5</td>
<td>82.5</td>
</tr>
<tr>
<td>2004</td>
<td>4</td>
<td>2 220</td>
<td>130 760</td>
<td>2 274</td>
<td>10.0</td>
<td>22 789</td>
<td>107 971</td>
<td>17.4</td>
<td>82.6</td>
</tr>
<tr>
<td>2005</td>
<td>1</td>
<td>1 962</td>
<td>137 596</td>
<td>2 254</td>
<td>10.1</td>
<td>22 758</td>
<td>114 838</td>
<td>16.5</td>
<td>83.5</td>
</tr>
<tr>
<td>2005</td>
<td>2</td>
<td>3 311</td>
<td>134 677</td>
<td>2 329</td>
<td>10.0</td>
<td>23 204</td>
<td>111 474</td>
<td>17.2</td>
<td>82.8</td>
</tr>
<tr>
<td>2005</td>
<td>3</td>
<td>3 010</td>
<td>132 260</td>
<td>2 384</td>
<td>9.0</td>
<td>21 429</td>
<td>110 831</td>
<td>16.2</td>
<td>83.8</td>
</tr>
<tr>
<td>2005</td>
<td>4</td>
<td>2 292</td>
<td>136 224</td>
<td>2 312</td>
<td>12.1</td>
<td>27 962</td>
<td>108 262</td>
<td>20.5</td>
<td>79.5</td>
</tr>
<tr>
<td>2006</td>
<td>1</td>
<td>2 094</td>
<td>149 876</td>
<td>2 272</td>
<td>11.9</td>
<td>27 023</td>
<td>122 853</td>
<td>18.0</td>
<td>82.0</td>
</tr>
<tr>
<td>2006</td>
<td>2</td>
<td>3 052</td>
<td>147 423</td>
<td>2 411</td>
<td>10.8</td>
<td>26 151</td>
<td>121 272</td>
<td>17.7</td>
<td>82.3</td>
</tr>
<tr>
<td>2006</td>
<td>3</td>
<td>3 202</td>
<td>143 009</td>
<td>2 445</td>
<td>11.3</td>
<td>27 660</td>
<td>115 349</td>
<td>19.3</td>
<td>80.7</td>
</tr>
<tr>
<td>2006</td>
<td>4</td>
<td>2 574</td>
<td>154 865</td>
<td>2 273</td>
<td>13.2</td>
<td>30 090</td>
<td>124 775</td>
<td>19.4</td>
<td>80.6</td>
</tr>
<tr>
<td>2007</td>
<td>1</td>
<td>2 270</td>
<td>159 574</td>
<td>2 300</td>
<td>10.7</td>
<td>24 554</td>
<td>135 020</td>
<td>15.4</td>
<td>84.6</td>
</tr>
<tr>
<td>2007</td>
<td>2</td>
<td>3 180</td>
<td>155 338</td>
<td>2 404</td>
<td>12.8</td>
<td>30 798</td>
<td>124 540</td>
<td>19.8</td>
<td>80.2</td>
</tr>
<tr>
<td>2007</td>
<td>3</td>
<td>3 152</td>
<td>147 564</td>
<td>2 431</td>
<td>11.4</td>
<td>27 717</td>
<td>119 848</td>
<td>18.8</td>
<td>81.2</td>
</tr>
<tr>
<td>2007</td>
<td>4</td>
<td>2 391</td>
<td>157 230</td>
<td>2 319</td>
<td>15.6</td>
<td>36 173</td>
<td>121 057</td>
<td>23.0</td>
<td>77.0</td>
</tr>
<tr>
<td>Average 2002-2007</td>
<td></td>
<td>131 207</td>
<td>24 736</td>
<td>106 471</td>
<td>18.9</td>
<td>81.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*from Statistics Finland's statistics on for single-family houses land plots for the corresponding period

Table 2.1: Land shares and weights derived from the published price statistics on single-family houses and land-plots for single family houses

So we use in the computations of the structure price index $w_s = 81.1\%$ and $w_l = 18.9\%$. The results for the whole country are presented in Diagram 1.
Diagram 2.1: The implicit structure price index for single family houses in Finland

Till the end of the 1990’s the structure price index follows very closely the overall price index. Structure prices show slower developments than total dwelling prices in the last decade. The result is simply consequence of the construction methodology and the unusually fast growth of land prices during the period. Since the land component is, according to the above calculations, only about one-fifth of the total price of a single-family house, the profile of the structure price index is very similar to the profile of the total price index, only a bit shifted “downwards”. The cumulative gap since the first quarter of 1985 till the fourth quarter of 2007 between the total single-family houses price index and the structure-price component is 8.9 percent, which is a non-negligible but still rather modest difference. The behaviour of the indices during the previous housing-market cycle 1985-1992 suggests that the land component is more volatile than the total index.

As already stated, according to the above calculations the land-price component in 2002-2007 is only about 20 % of the total price of a single-family house in Finland on average. The situation is very different in the Greater Helsinki region (the municipalities of Helsinki, Espoo, Vantaa and Kauniainen). There the prices per square metre for land plots available for construction of single-family houses are about ten times higher than the national average but the price of a single-family house is “only” about twice the national average. So the computed weight of land-component in the overall price index is 49 %. A further challenge is that the average number of transactions per quarter registered in Greater Helsinki is only about 100 for the single-family house and about 90 for the construction land for that type of houses. As a consequence, the land price index is very unstable. The results of the computations for Greater Helsinki are presented in Diagram 2.2

---

5 The index points for the indices scaled at 1985/1=100 for 2007/IV are 310,7 for the total dwelling price and 283,0 for the structure price.
6 Average land plot size is also much smaller in Greater Helsinki, so the total price of a land plot in Greater Helsinki is roughly 5-6 times more expensive then the country average
Diagram 2.2: The implicit structure price index for single family houses in Greater Helsinki

While the overall price index is reasonably smooth, the land-price index is very volatile, occasionally showing quarter-to-quarter changes of over 100 %. The computed structure price index mirrors the movements of the land price index and is as volatile as the land-price index.

2.3 Hedonic regressions for the single-family houses

2.3.1. The simple model

In this section we present and discuss the results from the model

\[
P_{nt}^{T} = \left( \sum_{i=1}^{T} \delta_i D_i + 1 \right) p_0^{S} S_{nt} + \left( \sum_{i=1}^{T} \gamma_i D_i + 1 \right) p_0^{L} L_{nt} + \epsilon_{nt}, \quad t = 0, \ldots, T
\]

This is (1.5) repeated. The model contains no other characteristics than the structure and the land size of a single-family house as well as separate time trends to capture the potentially different price developments of land and structure. As discussed earlier such a model is likely to be a reasonably good specification only for analysis of price developments in a particular regional “sub-market”. Anyway we start by estimating this model for the whole of Finland, because it presents a good reference point for the more complex analysis, where regional indicators and other characteristics of the house and its land plot are included in the model.
The above table summarises the estimation results for 24 quarters - from the first quarter of 2002 to the last quarter of 2007. According to the results the point estimate for the unit price of structure for the reference period, the first quarter of 2002, is 804.27 euros per square metre of dwelling space, implying that the structure of a single-family house with 100 m² would have cost 80 427 euros at that time. The standard error of the estimate is relatively small (21.69 euros) and therefore the coefficient is statistically extremely significant and the 95 % confidence interval for the estimate is reasonably narrow: 761.8 - 846.8 euros per square metre.

The estimated unit price for land is 1.8327 euros per m² of land plot, implying that the land share of the price of a house built on land plot of 2000 m² would have been 3665 euros in the beginning of 2002. This is a very low estimate compared to the observed actual prices of construction plots for single family houses – 9.7 euros per square metre according to the published statistics for the first quarter of 2002. The 95 % confidence interval is between 0.55 and 3.12 euros.

A very important assumption of the estimated model is that there is no overall intercept term, because the presence of an intercept is inconsistent with the basic idea – that the total price of a house can be expressed as the sum of the structure and the land prices.

For practical purposes the assumption implies that the model fits reasonably well if the size of a possible intercept is close to zero, even if not exactly zero. In this model a
positive intercept of say 1000 means that the estimated sum of land and structure prices is, on average, 1000 euros too low compared to the observed total price. Given that average transaction price for the single-family houses is well above 100,000 Euros in the period 2002-2007, an average relative error of the order of 1 percent is not too serious.

Unfortunately, in the simple model presented above the error is about 10%, which is a clear indication that the simplest model of the type (5) applied to the whole country is not satisfactory. To see that, we have simply calculated the average residual between predicted and observed total prices for each quarter and for the sample as a whole. The relative size of the residual (over all quarters) is 14,363 euro, 10.9% of the average single-family price (131,207 euro) in the sample. The quarterly results are summarised in Diagram 2.3.

![Diagram 2.3: The average residual as fraction of the average total price by quarter](image)

The average residual for the different quarters varies between 9 and 12 percent, suggesting that the model systematically undervalues the overall price of the single-family houses by about 10 percent. Although it seems quite obvious that the model in this form is not satisfactory, we computed the land, structure and overall price indices using the model estimates. Since the relative size of the average quarterly residual is rather stable, despite its shortcomings the model may still provide reasonable estimates of the time trends of the different price components and a reference point for the more sophisticated model.
Diagram 2.4: The total price index and the structure price index computed from the simple regression model

Diagram 2.4 presents the results for the overall price index and the structure component alongside with the official price index for single-family houses and the implicit structure price index computed in section 2.2. All indices exhibit very similar overall trends, but some differences in the quarter to quarter behaviour. The official price index shows somewhat larger overall price change between 2002/I - 2007/IV, 51.6 percent, while the other indices show virtually the same cumulative change of 40-43 percent. The correlation of the quarter to quarter changes of the official price index and the overall price index computed here is 75.5%, which is not low but far from perfect either. The correlation of the quarter to quarter changes between the implicit structure price index from section 2.2 and the structure price index computed here is about 70%.

All in all the overall price index and the index for structure estimated in this section do not appear unreasonable. Unfortunately the behaviour of the estimated land-price index is completely unsatisfactory, as seen from Diagram 2.5.

Diagram 2.5: The land price index estimated from the simple model in this section
The land price index estimated on the basis of the model from this section is extremely volatile and the quarter-to-quarter changes computed from it are virtually uncorrelated with the quarter-to-quarter changes of the official index (correlation 13%).

The reason why the overall price index estimated from the simple model in this section is still reasonable is because the land weight is very low – only 6% when the whole sample period, 2002-2007, is taken as a base period. This result is of course largely a consequence of the fact that the estimated unit price for land for the reference period was very low, as discussed earlier.

2.3.2. Hedonic model with characteristics affecting unit prices of land and structure

In this section we try to improve the simple model from the previous section by making unit prices of land and structure dependent on different characteristics. The model corresponds to equation (1.7) in chapter one. More precisely, the model here is of the form

\[
P_{n}^{T} = \left( \sum_{i=1}^{11} \gamma_{i}x_{n}^{i} + 1 \right) \exp(\alpha x_{n}^{S}) p_{n}^{S} s_{n} + \left( \sum_{i=1}^{11} \gamma_{i}x_{n}^{i} + 1 \right) \exp(\beta x_{n}^{L}) p_{n}^{L} l_{n} + \theta_{n}
\]

The difference from (1.7) is that the constant terms from the unit price expressions are taken “outside” the exponential terms \( \exp(\alpha x_{n}^{S}) \) and \( \exp(\beta x_{n}^{L}) \) and are represented by the parameters \( p_{n}^{S} \) and \( p_{n}^{L} \), which we call reference unit prices for structure and land. These “reference unit prices” correspond in their interpretation to the intercept term in an ordinary regression. For example, if all explanatory variables in the structure unit price equation for land are set to zero, the estimate of the unit price for land is the “reference unit price”.

Before estimating the model some checks of the data were performed. In principle we use validated data used in the production of the official statistics, so we were very careful with the additional restriction that we imposed. The official statistics rule out all observations where the total price is lower than about 17 000 euros (100 000 FIM) and where the dwelling space is less than 40 m^2. The largest house in the data was 440 m^2 and the most expensive 1.42 million euros. We dropped from the set 25 observations in which the land plot size was very low (less than 400 m^2) or very high (over 20 000 m^2). We also ruled out 14 observations, because the distance from the municipality centre was over 50 km, which is not plausible for municipalities outside Eastern and Northern Finland. For Eastern and Northern Finland even over 100 km distances were accepted (only several such cases in the data). Thirty nine observations had missing values for the distance from the municipality centre and were subsequently dropped from the estimation\(^7\).

\(^7\)We attempted to include an indicator variable for the variables with missing value for the distance variable, but it lead to numerical problems. The number of the observations dropped due to missing
The model contains fifteen regional indicators and two indicators for the municipality size (the largest and the smallest municipalities). These are included in both the structure and the land unit price equations to capture regional price level differences. Also the structure size (square metres of dwelling space) is included as an explanatory variable in both the structure and the land unit price equations. The construction year of the house (variable re-scaled as construction year - 1900) as well as an indicator for unknown construction year are included only in the structure unit price equation. Respectively the size of the land plot and its distance from the centre of the municipality are included as explanatory variables only in the land unit price equation. The squared distance is included, because trial estimations indicated a clear presence of nonlinear effects of this variable. The model includes also some of the interactions (altogether nine) of the continuous variables with the municipality size and the Greater Helsinki indicators. As in the model from the previous section a set of time dummy variables for each quarter in the years 2002-2007 is included separately for the structure and land unit prices.

First the nonlinear regression of the type (2.3) was estimated. Because heteroskedasticity is expected to be a serious problem for the model, we ran an auxiliary regression of the squared residuals on the predicted values of the total house price of the following form:

\[(2.4) \quad e^2 = a_0 + a_1 \hat{P} + a_2 \hat{P}^2 + \text{residuals}\]

The F-test of the hypothesis that the coefficients \(a_1\) and \(a_2\) are jointly zero was massively rejected\(^8\), indicating the presence of significant heteroskedasticity in the model. Under these circumstances, the estimates from the standard NLS, although consistent, may be very inefficient. In an attempt to improve estimation efficiency we re-estimated the model applying weights to take at least roughly into account the heteroskedasticity. In this case it was possible to use the fitted values of \(e^2\) from (2.4) since they were all positive. The weighted model is of the form:

\[(2.5)\]

\[
\frac{1}{\sqrt{e^2}} P^T_{mr} = \frac{1}{\sqrt{e^2}} \left[ \sum_{r=1}^{R} D_r + 1 \right] \exp ( \alpha' x^S_m ) p^S_{mr} S_{mr} + \left( \sum_{r=1}^{R} D_r + 1 \right) \exp ( \beta' x^L_m ) p^L_{mr} L_{mr} + \frac{\theta_m}{\sqrt{e^2}}
\]

Estimates based on (2.5) are likely to be more efficient than the ones from (2.3). However, since (2.5) accounts only roughly for the heteroskedasticity, heteroskedasticity robust covariance estimates are still needed for statistical inference.

We also estimated a restricted version of the model, where the land part of the model is dropped from the equation both by unweighted and weighted NLS. It is interesting to look at the results because the restricted version of (2.3) distance variable is marginal compared to the total number of observations - over 60 000, so it should not matter for the results.

\(^8\) The F-statistics value is 1843.94 which is extremely significant against the null-hypothesis
is consistent with the most common form used in the estimation hedonic indices in general and for house-price indices in particular, namely:

$$P_{nt}^r = \left( \sum_{t=1}^{T} \delta D_t + 1 \right) \exp(\alpha' x_{nt}) p^S_{nt} S_{nt} + \omega_{nt}$$

$$\ln \left( \frac{P_{nt}^r}{S_{nt}} \right) = \alpha_0 + \sum_{t=1}^{T} \beta D_t + \alpha' x_{nt} + \psi_{nt}$$

In the last formula the logarithm of the total price of a dwelling per structure unit (such as square metre of dwelling space) is regressed on a number of characteristics of the dwelling (the land plot size may be one of them) and a number of indicators for the transaction period (sales quarter). The price index is read from the exponential of the coefficients of the transaction period indicators. Of course the specifications (2.6) and (2.7) cannot be used for the purpose of separating the total price into land and structure price components.

We checked for possible local minima by using different algorithms and initial values and we found two different sets of parameters which satisfied the first order condition for minimising the SSE for the weighted model with different SSE (sum of squared errors). This means that we indeed found a local minimum. The possibility of local minima is a problem which obviously has to be taken seriously in the type of model we are working with. The estimates we present are based on the smallest SSE we’ve found. There is no theoretical guarantee that we have found the parameter values which globally minimise SSE.

In Table 2.3 we report only the results from the full weighted model. The indices computed using the different models are presented later on in the diagrams.

In that type of model it is difficult to draw conclusions about the relative importance of the structure and land components for the overall price directly from the estimated coefficients, because these will vary by region and according to other relevant characteristics. We will only note that the coefficients of most regional indicators, continuous variables and interaction terms appear to be statistically highly significant. What we are interested in is what kind of indices we derive from the different models and these are discussed further in the text.

As in the previous section we tested whether the assumption of no overall intercept in the model is justified against the data. The results show a dramatic improvement compared to the simple model from section 2.3.1. The absolute value for the average residual for the weighted model is less than 1% of total average price for all quarters. Also the unweighted model performs well in this respect. We can therefore conclude

---

9 One should be careful about the inferences. Even though the statistics are heteroskedasticity robust there remains the potential problem of spatial correlation - the error terms can be correlated for example by region.
that the models presented in this section, at least for practical purposes, are consistent with the central assumption of the model - that the total price of a dwelling can be in principle represented as the sum of two price components - land and structure.

Diagram 2.6: The average residual as fraction of the average total price by quarter: the models of this section compared to the model from section 2.3.1
### Sales quarter dummies

| Sales quarter dummies | coeff | str error | t | pr>|t| | coeff | str error | t | pr>|t| |
|-----------------------|-------|-----------|---|-----|-------|-----------|---|-----|
| 2002/1                |       |           |   |      |       |           |   |      |
| 2002/2                | -0.00330 | 0.01592 | -0.21 | 0.836 | 0.19698 | 0.14068 | 1.40 | 0.161 |
| 2002/3                | 0.00792 | 0.01676 | 0.47 | 0.636 | 0.07650 | 0.03732 | 2.08 | 0.039 |
| 2002/4                | 0.03131 | 0.01821 | 1.72 | 0.085 | 0.16477 | 0.12901 | 1.29 | 0.212 |
| 2003/1                |       |           |   |      |       |           |   |      |
| 2003/2                | 0.07586 | 0.01942 | 3.91 | 0.000 | 0.15890 | 0.12002 | 1.32 | 0.208 |
| 2003/3                | 0.08559 | 0.01751 | 4.89 | 0.000 | 0.13893 | 0.09487 | 1.50 | 0.135 |
| 2003/4                | 0.09984 | 0.01815 | 5.50 | 0.000 | 0.12951 | 0.10371 | 1.27 | 0.226 |
| 2004/1                |       |           |   |      |       |           |   |      |
| 2004/2                | 0.13127 | 0.02005 | 6.55 | 0.000 | 0.16052 | 0.12501 | 1.27 | 0.226 |
| 2004/3                | 0.17061 | 0.02080 | 8.20 | 0.000 | 0.17190 | 0.13202 | 1.27 | 0.226 |
| 2004/4                | 0.20167 | 0.01991 | 10.13 | 0.000 | 0.16458 | 0.12152 | 1.27 | 0.226 |
| 2005/1                |       |           |   |      |       |           |   |      |
| 2005/2                | 0.21015 | 0.01962 | 9.74 | 0.000 | 0.16201 | 0.12201 | 1.27 | 0.226 |
| 2005/3                | 0.23265 | 0.02117 | 11.00 | 0.000 | 0.17563 | 0.14068 | 1.27 | 0.226 |
| 2005/4                | 0.28728 | 0.02298 | 12.50 | 0.000 | 0.18239 | 0.14068 | 1.27 | 0.226 |

### Regional indicators

| Regional indicators | coeff | str error | t | pr>|t| |
|---------------------|-------|-----------|---|-----|
| Helsinki            | 0.17314 | 0.02288 | 7.57 | 0.000 |
| Espoo               | 0.31179 | 0.02246 | 13.88 | 0.000 |
| Greater Helsinki    | 0.51397 | 0.04515 | 20.24 | 0.000 |
| Jyväskylä           | 0.27725 | 0.02184 | 12.69 | 0.000 |
| Kuopio              | 0.18359 | 0.02039 | 9.74 | 0.000 |
| Lahti               | 0.22617 | 0.02059 | 10.33 | 0.000 |
| Oulu                | 0.25502 | 0.02039 | 12.69 | 0.000 |
| Pori                | -0.07160 | 0.02039 | -3.51 | 0.000 |
| Tampere             | 0.54535 | 0.02130 | 25.60 | 0.000 |
| Turku               | 0.42332 | 0.02246 | 17.44 | 0.000 |
| Municip. around Greater Helsinki | 0.45976 | 0.00884 | 52.02 | 0.000 |
| Uusimaa             | 0.28831 | 0.00942 | 30.59 | 0.000 |
| Eastern Finland     | -0.13751 | 0.00822 | -16.73 | 0.000 |
| Northern Finland    | -0.16135 | 0.00880 | -18.34 | 0.000 |
| Corid Finland       | -0.13891 | 0.00772 | -18.34 | 0.000 |
| Municip. over 50 000 inhabitants | -0.10702 | 0.02630 | -4.07 | 0.000 |
| Municip. less than 10 000 inhabitants | -0.21011 | 0.01529 | -13.74 | 0.000 |

### Characteristics of the house

| Characteristics of the house | coeff | str error | t | pr>|t| |
|-----------------------------|-------|-----------|---|-----|
| size of structure           | -0.00393 | 0.00007 | -60.22 | 0.000 |
| (size of structure)* (Municip. over 50 000 inh.) | 0.00026 | 0.00014 | 1.80 | 0.073 |
| (size of structure)* (Greater Helsinki) | 0.00011 | 0.00010 | 1.08 | 0.276 |
| Year of construction - 1900 | 0.00263 | 0.00007 | 4.11 | 0.000 |
| Year of construction missing | 0.00263 | 0.00007 | 4.11 | 0.000 |

### Reference unit-price

| Reference unit-price | coeff | str error | t | pr>|t| |
|---------------------|-------|-----------|---|-----|
| 1194.60             | 22.38 | 53.37     | 0.00 | 295.8192 | 40.8547 | 7.38 | 0.000 |

According to the weighted model the average land share in total house price for the period 2002-2007 is 7.4 percent. The corresponding estimate from the unweighted model is 9.1%. So the relative importance of the estimated land component for the overall price is at least somewhat higher than in the model from section 2.3.1 but is still lower.
very low compared to the estimate from section 2.2 obtained directly from the price statistics on single-family houses and land plots for construction of single-family houses.

Unfortunately, the models fail to produce interpretable land price indices, as seen from Diagram 7

![Diagram 2.7: The land price indices computed from the models of this section and the official land price index](image)

On the other hand all models discussed here – also the ones that omit the land-price component – produce estimates of the total house price index which are almost surprisingly close to the Finnish official price index for single family houses, although the latter is computed using a rather complicated method involving data stratification, separate regional regression models and aggregation by an index formula (log-Laspeyres).

![Diagram 2.8: The price indices for single-family houses computed from the models of this section and the official price index for single-family houses](image)
Without reasonable estimates of the index for the land price component one cannot have a separate price index for the structure component either.

Because the estimated land weight is very low, the “structure price index” estimated from the weighted model is almost the same as the total price index.

On *a priori* grounds it was expected that land size and structure size can be highly correlated – large houses are built on large plots. The existence of such proportionality relation would have given rise to multicollinearity problems in our models and would have explained the inability of the model to separate price developments of the structure and the land components. However, somewhat surprisingly it turned out that land size and structure size are actually almost not correlated. The correlation coefficient is -1%. We also run a regression of the structure size on the land size and the interactions of the land size with the regional and time indicators. The $R^2$ of this regression is 5.1% showing that the correlation between land and structure sizes is very low even when regional differences are taken into account.

A possible direction for further examination of the issue is to try to estimate even richer models. In practice that would mean separate models for several large regions with appropriate indicators for sub-regions such as municipalities. Another one is to regress on “building permit units” rather than directly on land plot size. Building permit units are likely to be better related to the land price than the land size. On the other hand it maybe the case that building permit units are also considerably more correlated with the structure size. In Finland building permit units are not even exactly defined for all real estates - it depends on what kind of construction plan covers the area. Roughly speaking building permit units are defined in towns and not elsewhere. In the data we have used roughly half of the observations contain such information. To our understanding also some additional checks on the validity of information on building permit units is necessary.
2.3.3 The adjacent period model

The method we applied in sections 2.3.1 and 2.3.2 is a variant of the so called time dummy approach. This approach is criticised mainly for two reasons. The first one is that in the approach it is assumed that the coefficients of all regressors are time-invariant. Changes in supply (due to technological change and or competition for example) and demand (consumer’s tastes) may result in changes in the price equilibrium relationship between supply and demand of goods with different bundles of characteristics described by the hedonic function. This means that over time the “relative prices” of the different characteristics which are reflected in the coefficients of the regressors may change. Actually it may be the case that even the hedonic function form changes over time.

Another argument against the time dummy approach is of more practical nature and is usually put forward by statistical agencies. An official index is produced regularly and the data set used for calculations of the index will be inevitably updated with more recent data. Under the time dummy approach, addition of new observations to an existing data set and subsequent re-estimation of the model will result in some changes in the whole history of the index series. Continuous revisions of an index history is problematic – an official index is likely to be used in political decision making, contracts and prices may be linked to it and it may be itself an ingredient in other indices.

A simple modification of the time dummy approach that allows both for regressor coefficients changing in time and avoids the necessity to update the index history is the so called adjacent period approach. In this approach the usual time dummy model is estimated only for subsequent (adjacent) periods. One first runs a regression using the data for periods 0 and 1 to obtain the relative price change between these two periods, $I_0^2$. Then the regression is run using the data for periods 1 and 2 to obtain $I_1^2$ etc. The index series is then obtained by chaining the adjacent period relative price changes. The price change between period 0 and $T$ is measured by $I_0^T = \prod_{i=0}^{T-1} I_i^{i+1}$.

The practical problem with the adjacent period model is that it requires enough data for each and every period to obtain reliable estimates for the relative price changes in adjacent periods. Because the data sets available for adjacent quarters is much smaller than the total data for the period 2002-2007, we simplified the model presented in table 3.2 by excluding from the adjacent period regressions the municipality-level indicator variables. We also estimated a model with the same explanatory variables and the time dummy approach to be able to compare the behaviour of the indices from the adjacent period models to the indices from the time dummy model.

To alleviate the serious heteroskedasticity problem estimation is based on a two-step procedure analogous to the one described in section 2.3.2. We first estimated the models without weights and used an auxiliary regression of the logarithms of the squared residuals on the predicted house price and its square. Weights were constructed from the
exponents of the predicted values from the auxiliary regressions and the model is re-estimated using these weights.

We tested for the presence of local minima and found parameter values that minimise locally the SSE on several occasions.

There is not really a difference between the total house price index estimated from the adjacent period regressions and the time dummy index from a model with the same structure. Both follow very closely the official house-price index.

![Diagram 2.10: The price index from the adjacent period models.](image)

The estimate for the share of land in total house price from the adjacent period model is 6.7%, a value comparable the one obtained from the model from section 2.3.3. The estimated land price index is very volatile and because of the low land share in total price it does not affect much the total price index.

![Diagram 2.11: The land index from the adjacent period models.](image)
Diagram 2.12: The behaviour of the total house-price index and the estimates of the subindices for “land” and “structure” from the adjacent period model
2.3.4 The Diewert specification

Although in chapter 1 we were critical about the original proposal made by Diewert (2007), it is naturally worthwhile testing the model suggested there. The model is of the form

\[
\ln(P_{mt}^T) = \ln \left( \sum_{t=1}^{T} \delta_t D_t + 1 \right) \frac{\alpha' \mathbf{x}_{mt}^S}{p_m^S(0)} S_{mt} + \sum_{t=1}^{T} \gamma_t D_t + 1 \left( \beta' \mathbf{x}_{mt}^L \right) L_{mt} + \xi_{mt}
\]

One should note that although the specification ensures that all predicted prices will be positive, it is possible for some observations to have negative estimates for the predicted share of either the land or the structure of total price.

In our data we have 10% of all observations for which the model predicted negative share of land in the total price. The proportion of such observations is stable for all quarters covered by the data. The average share of land in total price is positive but low – 5.3%.

The behaviour of the total price index and the subindices for land and structure follow qualitatively the general pattern observed for all models that we tested in this chapter. The total price index is much like the official index and the land price index is very unstable.

![Diagram 2.13: The behaviour of the total house-price index and the estimates of the subindices for “land” and “structure” from the model suggested by Diewert (2007)](image)
2.4 Summary of the findings and conclusions

In this chapter we tested several different methods and provided a number of different estimates for separating the structure price and the land price developments of a price index for single-family houses.

The first method was to compute an implicit structure-price index using the available indices on total price for single-family houses and land-plots for construction of single-family houses. Even though since the end of 1990’s prices of land plots in Finland have increased much faster than single-family house prices, the share of land in the overall single-family house price during 2002-2007 is on average rather modest, around 20%. As a consequence the implicit structure price index shows a slightly more moderate growth than the overall house price index.

However, the above conclusions apply only “on average” for the whole country. In the region of Greater Helsinki the share of land is half of the total price of a single-family house. Due to very few transactions per quarter the quarterly land price index is very volatile and this excess volatility is transferred to the structure price index for Greater Helsinki.

Calculating implicit structure price indices is objected on the obvious grounds that the land price index may be based on transaction prices of land plots which are very different in their characteristics (especially location) compared to the built land plots on which actual house transactions occur. For that purpose an econometric model was developed and tested using the transaction price data for single-family houses. The essence of the model is to break down the total price of a house into the sum of two components – land and structure.

In the first stage we estimated a simple version of the model. In this version the model does not include characteristics of the house transacted other than its size and the size of its plot. Separate price developments for the land and structure unit prices are estimated to obtain separate price indices for land and structure. Then the two indices are aggregated into a total house price index.

The statistical properties of the simple version of the model were not in all respects satisfactory; particularly it systematically underestimated total house prices on average by 9 to 12 % for each quarter of the data. Nevertheless, the total house price index computed from this model seems quite reasonable. However the estimated share of the land component in total price was low, 5.9 percent, and the behaviour of the land price index was not satisfactory.

In the second stage we estimated a model in which unit prices of structure and land were allowed to depend on different characteristics of the structure and the land plot – region, size of municipality, distance from the municipality centre, year of construction, size of the structure and the land plot. We estimated two versions of this model - unweighted and weighted. We used weights to correct for the heteroskedasticity of the model. The overall house price indices based on the these models are very close to the Finnish
official price index for single family houses, although the latter is computed using a rather different and more complicated method.

The relative importance of the estimated “land component” from these models is still low, 9.1% in the unweighted model and 7.4 for the weighted. These models also fail to produce interpretable land price indices.

The method we applied in sections 2.3.1 and 2.3.2 is a variant of the so called time dummy approach. This approach is criticised mainly for two reasons. The first one is that in this approach it is assumed that the coefficients of all regressors are time-invariant. Another argument against the time dummy approach is that updating the index for a more recent time period will result in some changes in the whole history of the index series. Continuous revisions of the index history are problematic from the point of view of statistical agencies. A way to tackle both this problems is to have separate regressions to estimate the relative price change between two adjacent time periods.

We tested the adjacent period method and arrived at the same conclusion as before – the overall index is very similar to the official house price index, but the estimated land share of total price is low and the estimated land price index is very unstable.

We arrived also at the same qualitative results after testing the model specification suggested by Diewert (2007).

Concerning the estimation of the models in sections 2.3.2 and 2.3.3, which are based on the approach we developed in chapter one, special care is needed because local minima for the criteria function (the sum of squared residuals) are possible and we have actually encountered such situations in practice.

Altogether we have to conclude that while the econometric model we have developed seems to produce a very reasonable total price index for single-family houses, the interpretation of its two additive terms as “structure” and “land” components of the total price does not fit the expectations of potential users especially in what concerns the land fraction. Strong correlation between structure size and land size is definitely not a problem in the data used here.

From the point of view of econometric modelling an interesting finding is that a restricted model in which total house price is regressed only on the “structure component” of the full model leads to a very similar house price index as of the full model. The specification of the restricted model is consistent with the most traditional version of the hedonic models, in which the logarithm of the total price of a dwelling per structure unit (such as square metre of dwelling space) is regressed on a number of characteristics of the dwelling (the land plot size may be one of them) and a number of indicators for the transaction period (sales quarter). Our findings therefore show that starting from a more general model, we end up with the traditional specification, which unfortunately is not applicable for the purpose of breaking down the overall house price and the corresponding index into structure and land price components.

A possible direction for further examination of the issue of obtaining separate price indices for structure and land using observed total house prices is to try to estimate even
richer models. In practice that would mean separate models for several large regions with appropriate indicators for sub-regions such as municipalities. Another one is to use some variable better related to the owner’s rights to use the land for construction purposes than simply the land size. Such “building permit units” are likely to be better related to the land price than the land size. On the other hand it may be the case that building permit units are also considerably more correlated with the structure size. In Finland building permit units are, roughly speaking, typically defined in towns and not elsewhere.

Based on the evidence from this work, it seems that at least in the Finnish case, deriving an implicit structure price index from separate series on total house prices and land prices may be a better alternative than trying to estimate land and structure price indices from transaction data on houses. This statement is subject to the condition that the land price index is of good quality and does not exhibit excess volatility, since this volatility will be translated into the structure price index. Given the limited number of transactions of land plots for single-family houses in each quarter, for Finland and other small countries in general, it may be difficult to obtain good quality regional series on structure prices. Another important caveat is that this methodology may not be applicable to the case of apartments in blocks of flats. Land price indices for land plots intended for construction of blocks of flats may be much more difficult to construct than land price indices for plots allocated for single-family houses. Use of land prices for single-family house plots to evaluate price developments of land plots for blocks of flats is certainly a problematic alternative.
3. Apartments in blocks of flats in Helsinki

Compared to the case of single-family houses, constructing a model to separate structure price from land price and compute the respective indices for dwellings in multidwelling buildings (typically apartments in blocks of flats) involves the extra difficulty of allocating land share of the total land plot to a particular dwelling. We developed a theoretical framework for this in chapter 1. In this chapter we describe the empirical results obtained when the method was applied to data covering old apartments in blocks of flats in Helsinki for 24 quarters from the beginning of 2002 till the end of 2007.

3.1 Data sources and Definitions

The current data used for computing the Finnish official index for old dwellings in housing corporations (apartments in blocks of flats, terraced and semi-detached houses) uses transaction price data obtained from the Finnish Tax Administration. 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 data covers all transactions of that type.

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 coordinates). 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.

For the purpose of the current project we needed also information on the size of the land plot on which the buildings owned by a housing-share corporation are built as well as information covering the total dwelling space and the number of dwellings in the housing corporation. We obtained data for the size of the land plot from the register of the Finnish Tax Administration kept for purposes of calculating the municipal tax on real estate property. Then this information was linked to the data used to calculate Finnish official index for old dwellings in housing corporations using the real-estate identification code and the business identification code of the housing share corporation.

Information for the total dwelling space and number of dwellings belonging to a housing share corporation as well as some auxiliary information on the presence of other than residential buildings and dwelling space in the housing share corporation was obtained from the building and dwelling register.

According to the definitions used in the official index it covers transactions of old (used) dwellings in housing share corporations. 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.
The restriction is dictated to a large extent by the fact that that for new dwellings linking data from different registers has proved difficult because of the different timing of their updating. There is a separate index for new dwellings launched this year by Statistics Finland. This index is based on data collection from real-estate agents and major construction companies. Unfortunately the data on new dwellings could not be used in the project, because it contains no information on total land plot size, dwelling size and number of dwellings in the housing share corporations. It was not possible to link these data to registers either, because it contains no identification codes.

The official index covers only free market transactions thus excluding transactions of dwellings in housing share corporations built using certain forms of subsidy and subject to price controls and/or imposing criteria on prospective buyers such as sufficiently modest income and wealth.

In our research we follow on this part the definitions of the official index - we examine free-market transactions of old (as defined above) apartments in blocks of flats in Helsinki. However we excluded the apartments built on leased land. We present our findings regarding dwellings on leased land in the following section.

### 3.2. Dwellings on leased land plot

The Finnish official index does not in any way account for whether a dwelling is situated on a land plot owned or just leased by the housing share corporation as long as the land lease agreement does not entail in any form price restrictions and/or criteria which the prospective buyer of a dwelling should meet. This is because historically there have been no large price difference between apartments on owned or leased land after differences in the distribution of the two types of dwellings have been taken into consideration.

Housing corporations using leased land is a common phenomenon in Finland especially in the large towns. In the data for Helsinki 22.5% of the transactions covered an apartment built on leased land. We computed the price ratio of apartments on owned land and apartments on leased land for the period 2002-2007 as a weighted average of the price ratios of the two types of dwellings computed by quarter and city region (the city was divided into 4 regions in the same way as in the official index). The weights were constructed from the proportion of transactions of dwellings on leased land by quarter and region to all transactions of dwellings on leased land for the period 2002-2007. The result was that the price of apartments on leased land is on average 95 % of the price of apartments on dwellings of own land when location differences are roughly controlled for.

Controlling for location differences is crucial for the result. A simple comparison of the average price of apartments on own land to the average price of apartments on leased land would give a price differential as large as 25 % in favour of apartments on own land. However, this is very misleading, because in the most expensive areas in the Helsinki centre less than 1 % of all transactions cover apartments on leased land, while in the least expensive north / north-east region transactions of apartments on leased land account for half of all transactions.
The existence of a large fraction of apartments on leased land and the very small price difference between apartments on own land and apartments on leased land pose a certain conceptual problem concerning the treatment of land in the net-acquisition approach to be used in Eurostat’s HICP. In theory there should be no land price involved in a transaction of an apartment on leased land provided that the land rent (paid ultimately by the dwelling owner as part of the charges collected by the housing share corporation) adequately reflects the value of the land plot. If this were the case, one would expect a sizeable price difference between dwellings on own and dwellings on leased land. Dwellings on leased land are expected to be cheaper to offset the land rent which owners of an apartment built on own land do not pay. The evidence from Helsinki suggests that the situation in practice is likely to be more complicated. After controlling for location differences we observe a very modest price difference between apartments on own land and apartments on leased land. This may mean that land is not very important for total price or that the present value of land rent is much lower than the value of the land plot. Land rents are probably rather modest in Helsinki. This is because the municipality of Helsinki is the most important (if not the only) owner of land leased for the purpose of dwellings construction. At least in the past pursuing profit has not been the municipality’s main objective when leasing land.

In a situation where the (present value) of the land rent is much lower than the price of land, one can expect that prices of apartments on leased land would reflect also a land component – the difference between the value of the land plot and the present value of land rent.

In principle treating apartments on leased land in an adequate way is possible within the econometric framework we have developed. However, because we are not aware of the position of Eurostat on the treatment of dwellings on leased land in the net acquisition approach and because we wanted avoid possible further complications of the method we test here we excluded apartments on leased land from the calculations.

### 3.3 Data description

The data for apartments in blocks of flats for Helsinki used in the official index for the period 2002-2007 contains altogether 61,533 observations. Most likely due to some changes in the data processing that has taken part over the years and/or much delayed information on some transactions our set contained initially 61,907 observations, 374 more than the data used for the official index computations. From these data we excluded the transactions of apartments on leased land, 13,705 observations for the examined period. There were also 1,309 observations which we could not link to the registers to obtain information on land size, etc. and which we also had to exclude from the analysis. Thus the data we use have in the end altogether 47,163 observations on transactions of free-market apartments in blocks of flats located on land owned by the respective housing share corporation. All apartments in the data are in Helsinki and the transactions have taken place in the period between first quarter of 2002 and fourth quarter of 2007.

---

10 For each quarter of 2007 the number of observation in our data at this stage was identical to the number of observations reported in the official index
Table 3.1 summarises the correlations between the main variables of interest in the data. It is important to bear in mind that direct correlation coefficients between the variables can provide useful information but can also be misleading when considering a strategy for constructing a model. The direct correlation coefficient’s size and sign is not necessarily a good indicator of the potential importance of a particular variable in a multivariate regression framework. This is because in a regression analysis what matters is the partial correlation - the correlation between the dependant variable and a regressor which remains after the effects of all other regressors have been taken into consideration. The partial correlation coefficient’s size and sign can be different than the size and sign of the simple correlation.

The apartment size is very strongly correlated with the transaction price as can be expected. Unlike in the case of single-family houses there is strong positive correlation between land plot size and total dwelling space, but we are not interested directly in these variables here, because we model the relation between price of a single apartment, its size and the portion of land allocated to it. The result is interesting mainly because it indicates that in the case of apartments in blocks of flats land is more efficiently used than in the case of single-family houses.

We have constructed two variables which allocate land to a particular apartment. The first measure (L1) allocates land equally between all dwellings on the land plot. The second measure (L2) allocates land to apartments in proportion to dwelling space. We have argued in Chapter 1 that both L1 and L2 are relevant measures - it is an empirical matter to decide if one uses either or both in the model. There is strong correlation (0.87) between the two allocation measures. This is important, because as it will become clear later on we were not able to include both measures in our models in the way described in Chapter 1 but had to use an alternative solution. L2 is slightly positively correlated with total apartment price while L1 and total apartment price are practically not correlated. This fact would seem to support the use of L2 rather than L1 if either one should be chosen, but as we will see from the modelling results such
conclusion can be misleading. L2 is more correlated to apartment’s dwelling space than L1, but the correlation is still not very strong – 0.44. This observation is perhaps to a large extent a due to fact that L2 is by construction directly related to apartment’s size.

An issue related to the allocation of land to an apartment is the presence of other than dwelling space on the land plot. There can be premises for common use by the inhabitants (common recreation rooms saunas and/or swimming pool), commercial premises (shop, office) or other premises (kindergarten, garages, storage space) located on the same land plot as the dwellings. These can be located in within the blocks of flats buildings or in separate buildings on the plot. While the housing share corporation is owner of the land and the buildings, commercial and other non-dwelling space can be either owned by the corporation or it may be owned by separate shareholders.

The way that other than dwelling space may affect total apartment price and its structure and land components depends on the kind of space and its ownership. Common facilities should affect total apartment price positively. Commercial space is expected to affect apartment’s price positively, if it is owned commonly by the housing share corporation. This is because such space will be rented and the rent will be used by the housing share corporation to maintain the property. On the other hand if commercial and other non-dwelling space is owned by separate shareholders, one should take this into consideration when constructing the land allocation variables. For example in constructing L2 one should allocate land in proportion to all dwelling and non-dwelling space in separate ownership.

In the register data we have there are in principle fields for different types of premises, but in the overwhelming majority of the cases they contain no information. We have also information if there are buildings on the land plot whose main intended use is for example shop, office, manufacturing etc. but we have no information about the size of such buildings. We do not have information on whether other than dwelling premises are ultimately owned by the housing share corporation or by separate shareholders.

By analysing the information on the types of buildings on the plot and the information on the total size of different types of premises in the buildings (as discussed the latter contain mostly missing values), we found out that altogether 6.5 % of the observations in our data are such that there is some information for the presence of commercial space and 0.5 % are also such that cases there are some social, cultural or teaching premises. In 1.2 % of the cases some other space (garages, storage) is present.

Given the information available the best we could do was to construct an indicator variable indicating the presence of commercial premises on the land plot. The indicator is included as an explanatory variable in both structure and land unit prices in our models. The indicator gets value 1 if there is a separate commercial building on the land plot or if according to the information available the space of the commercial premises was at least 5% of the total dwelling space of the dwellings on the land plot. We put this threshold requirement for two reasons. First, the information on the size of other than dwelling space is highly unreliable. We thought that perhaps reporting is more accurate in cases when commercial space is of non-negligible size compared to the dwelling space. The second argument is that if commercial space is negligible in comparison to total dwelling space it would not affect the price of an apartment. The indicator constructed in this way got value 1 in 3.85% of all cases.
3.4 Empirical results for the Helsinki apartments on blocks of flats

Unlike the case of single family houses we do not have statistics on land prices for plots intended for construction of blocks of flats so we cannot compute an implicit structure price index in the way we did in section 2.2. Making certain assumptions one can still use the information on land prices for single-family houses to compute a structure price index for apartments in blocks of flats, but this exercise is not the main topic of this research. Moreover we do not have a good land price index for Helsinki but only for the whole of Finland as discussed in 2.2.

The starting point of our model is equation (1.9) repeated in (3.1):

\[ P_{nh}^{T} = \left( \sum_{i=1}^{T} \delta_i D_i + 1 \right) \exp(\alpha' x_{nh}) S_{nh} + \]

\[ + \left( \sum_{i=1}^{T} \gamma_i D_i + 1 \right) \exp(\beta' x_{nh}) \left( \frac{1}{M_j} L_{nh} + (1 - \tau) \frac{S_{nh}}{S_j} L_{nh} \right) + \theta_{nh}, 0 \leq \tau \leq 1 \]

the variable \( \frac{1}{M_j} L_{nh} \) corresponds to our land allocation variable L1 and \( \frac{S_{nh}}{S_j} L_{nh} \) corresponds to the variable L2. Our conjecture was that in practice land allocation is some combination of the two variables. When initially testing the models, it turned out that the condition \( 0 \leq \tau \leq 1 \) is not satisfied by unconstrained estimation. On the other hand, imposing the constraint leads to numerical problems in estimation.

It also turned out that unconstrained estimation of models of the type (3.1) could result in coefficient estimates for the land-fraction time-dummies \( \gamma_i \) smaller than -1. Such estimates are not admissible, because they lead to negative contribution of the land price share to total apartment price and to a negative value of the land price share index for period \( t \).

For the above reasons we re-specified (3.1) as:

\[ P_{nh}^{T} = \exp \left( \sum_{i=1}^{T} \delta_i D_i \right) \exp(\alpha' x_{nh}) S_{nh} + \]

\[ + \exp \left( \sum_{i=1}^{T} \gamma_i D_i \right) \left( \frac{1}{M_j} L_{nh} + (1 - k) \frac{S_{nh}}{S_j} L_{nh} \right) + \theta_{nh}, k \equiv \frac{1}{M_j} \text{ or } k \equiv \frac{S_{nh}}{S_j} \]

In plain words we tested models where all coefficients for the unit prices and time trends are within an exponential term and where the land allocation variable was either L1 or L2, but we included in all models both L1 and L2 as explanatory variables in the land unit price equations. Estimates of the parameters of (3.2) necessarily satisfy the conditions that the structure and land price indices are positive and that the structure and
land contributions to total apartment price are positive. However if there are serious specification problems, estimation of a model like (3.2) may not succeed - the numerical optimisation algorithms may fail to converge or no unique solution is found (the Hessian matrix is singular).

We were able to estimate a reasonable model of the type (3.2) when L1 (land is allocated equally among all dwellings) was used as the land-allocation variable but could not obtain estimates for the same or similar model when the L2 (land allocated in proportion to dwelling size) was used. This despite the fact that the direct correlation of L1 to apartment price was in practice zero while L2 was slightly but clearly correlated to apartment’s price (see section 3.3).

Two-stage procedure was used. We first estimated (3.2) without weights and used an auxiliary regression of the logarithms of the squared residuals on the predicted apartment price and its square to obtain weights. The weights were constructed by applying the exponential function to the predicted values of the auxiliary regression. Afterwards, the model was re-estimated using these weights. The procedure to obtain weights is a slightly modified version of the procedure we used for the single-family houses data and the reason for applying weighted NLS was the same as in the single-family houses case too: to alleviate the serious heteroskedasticity problem which is present here too.

Table 3.2 is a summary of the estimation for the weighted model. With the usual warning about the problems related to statistical inference in that type of models we note that the coefficients of most explanatory variables appear to be statistically highly significant. However the relationship between certain groups of coefficients is not easy to explain. We take as an example the regional dummy variables. Helsinki is divided into four broad regions in the same way as in the official statistics. Helsinki 1 covers the very city centre with the most expensive dwellings and Helsinki 4 covers mostly the north-eastern suburbs where the relatively cheapest apartments are situated. Three regional indicators are included in both the equation for unit prices of structure and land with Helsinki 3 as reference region. While the regional indicator for structure for Helsinki 1 is large and positive – an observation consistent with overall prices higher than in the reference region, the corresponding coefficient for land in Helsinki 1 is negative. The reverse situation is observed for Helsinki 4.

Looking at the standard errors it seems that the time dummy coefficients for structure are much more precise than the corresponding coefficients for land. In many cases the time dummy coefficient for land appears to be statistically insignificant.

The usual check for the behaviour of the average residuals by quarter shows that in practice the assumption that total apartment price can be represented as the sum of structure and land price components is plausible for practical purposes. The estimated land share in total house price for the period 2002-2007 is only 4.9 percent. The indices derived from the model summarised in table 3.2 are presented in diagram 3.1.

The situation is similar to the single-family houses case. The official index and the total index from the model are very close to each other and the land price index is rather volatile, although its long-term behaviour seems reasonable. The volatility of the land
price index does not really affect the behaviour of the total price index, because of its very low weight.

We estimated also the following models, all of them including the same set of explanatory variables related to the unit prices of land and structure as the one presented in table 3.2:

- A model according to Diewert’s specification
- An adjacent period model
- A restricted version of model (3.2) so that the land-fraction term is set to zero

We also tried a specification where the regional indicators were included only in the land-fraction part but we could not obtain estimates for it, indicating that the model was seriously misspecified.

The results from these models are presented in the following diagrams 3.2-3.5.
### Table 3.2: Estimation results for the weighted regression model for apartments in blocks of flats in Helsinki. Land allocated equally among all dwellings on the land plot.

<table>
<thead>
<tr>
<th>Sales quarter dummies</th>
<th>Structure component</th>
<th>Land component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coef</td>
<td>str.error</td>
</tr>
<tr>
<td>2002/1</td>
<td>reference period</td>
<td></td>
</tr>
<tr>
<td>2002/2</td>
<td>0.06222</td>
<td>0.00947</td>
</tr>
<tr>
<td>2002/3</td>
<td>0.06601</td>
<td>0.00954</td>
</tr>
<tr>
<td>2002/4</td>
<td>0.05701</td>
<td>0.00940</td>
</tr>
<tr>
<td>2003/1</td>
<td>0.06086</td>
<td>0.00933</td>
</tr>
<tr>
<td>2003/2</td>
<td>0.09307</td>
<td>0.00855</td>
</tr>
<tr>
<td>2003/3</td>
<td>0.11347</td>
<td>0.00815</td>
</tr>
<tr>
<td>2003/4</td>
<td>0.14360</td>
<td>0.00843</td>
</tr>
<tr>
<td>2004/1</td>
<td>0.14333</td>
<td>0.00868</td>
</tr>
<tr>
<td>2004/2</td>
<td>0.18007</td>
<td>0.00830</td>
</tr>
<tr>
<td>2004/3</td>
<td>0.16401</td>
<td>0.00874</td>
</tr>
<tr>
<td>2004/4</td>
<td>0.16291</td>
<td>0.00853</td>
</tr>
<tr>
<td>2005/1</td>
<td>0.18105</td>
<td>0.00846</td>
</tr>
<tr>
<td>2005/2</td>
<td>0.22203</td>
<td>0.00917</td>
</tr>
<tr>
<td>2005/3</td>
<td>0.23255</td>
<td>0.00829</td>
</tr>
<tr>
<td>2005/4</td>
<td>0.27359</td>
<td>0.00844</td>
</tr>
</tbody>
</table>

### Regional indicators

<table>
<thead>
<tr>
<th>Helsinki 1 (centre)</th>
<th>Helsinki 2</th>
<th>Helsinki 3</th>
<th>Helsinki 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>reference</td>
<td>reference</td>
<td>-0.14756</td>
</tr>
<tr>
<td>0.52263</td>
<td>0.00503</td>
<td>103.84</td>
<td>0.00</td>
</tr>
<tr>
<td>0.30834</td>
<td>0.00339</td>
<td>90.89</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.25982</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.03797</td>
</tr>
</tbody>
</table>

### Characteristics of the house

<table>
<thead>
<tr>
<th>size of structure</th>
<th>1 room apartment</th>
<th>at least 3 rooms apartment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00010</td>
<td>-0.02952</td>
<td>0.00999</td>
</tr>
<tr>
<td>0.00018</td>
<td>-0.00309</td>
<td>0.00433</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year of construction - 1900</th>
<th>(Year of construction - 1900)^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.00641</td>
<td>-0.00099</td>
</tr>
<tr>
<td>-34.78</td>
<td>38.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indicator for presence of community centre</th>
<th>L1: land allocation measure 1, equally among all dwellings, m²</th>
<th>L2: land allocation measure 2, same fraction as apartment dwelling size to total dwelling size of all dwellings on the plot, m²</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.01333</td>
<td>-0.00180</td>
<td>-0.03977</td>
</tr>
<tr>
<td>-1.73</td>
<td>-2.08</td>
<td>-23.06</td>
</tr>
<tr>
<td>0.08</td>
<td>0.04</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>unit-price equation intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.57468</td>
</tr>
<tr>
<td>0.01403</td>
</tr>
<tr>
<td>539.73</td>
</tr>
<tr>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of obs 47 163. R squared is 0.52</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard errors and t-values are computed by using robust covariance estimate</td>
</tr>
</tbody>
</table>
Diagram 3.1: Total price index for apartments in blocks of flats and its decomposition into land-price and structure-price indices based on the model from table 3.2

Diagram 3.2: Price indices for apartments in blocks of flats in Helsinki based on different models
Diagram 3.3: Land price indices for apartments in blocks of flats in Helsinki based on different models

Diagram 3.4: Structure price indices for apartments in blocks of flats in Helsinki based on different models

All methods produce similar results for the total price index and are at least reasonably close to the official price index for apartments in blocks of flats in Helsinki produced by Statistics Finland. The official index uses classification and hedonic method to adjust for quality change. It is based on the hedonic imputation method and not on time dummy approaches. The index based on Diewert’s model results in a somewhat faster
growth of the index but on the other hand the quarter-to-quarter price change correlation between it and the official index is 94%. Also the quarterly price changes measured by the adjacent period index and the model that excludes the land price share from the model have about 92% correlation to the quarterly price changes measured by the official index. The correlation between the quarterly price changes measured by time dummy model from table (3.2) and the official index is 88%.

Thus, on the part of the total price index, we can draw very similar conclusions as in the case of single-family homes. The results seem rather robust to different specifications and modelling the land price share of total price does not matter for the total index. This is of course primarily because the land price share is according to the obtained estimates very low - about 5% for the time dummy and the adjacent period models and only 2% in the Diewert’s model specification.

The land price indices obtained from the time dummy and the adjacent period methods are close to each other showing no price trend in land prices for the period 2002-2006 and fast growing land prices after that. So their overall behaviour is plausible, but there seems to be still quite a lot of volatility in the quarter-to-quarter changes. The land price index constructed from the Diewert’s model exhibits a completely different behaviour.

In the situation where we have very low estimates for the land price share of the total price, the structure price indices are naturally very similar to the total price indices estimated from the respective models.

The main problem that we have with the models of the apartments in blocks of flats seems to be very low estimates of the land price share of total price. As a result it does not really matter whether one follows total apartment prices or structure prices. The specification in which the land fraction of total price is set to zero produces a very reasonable index for the total apartment price.
4. Summary and Conclusions

We developed a detailed econometric model for the purpose of separating overall dwelling price and its development into structure and land components using the idea of Diewert (2007) as a starting point. Although complicated in the details the model is intuitively very appealing, because it boils down to the obvious idea of representing the total price of a dwelling as the sum of two prices: the price of the dwelling structure and the price of land used for its construction. The representation makes possible both decomposing the total price of a dwelling into a structure and land components and constructing separate price indices for land and structure. Thus as a theoretical framework the model satisfies the requirements put forward by the wish to include owner occupied housing in the Harmonised Index of Consumer Prices on the basis of the net acquisition approach excluding land.

We tested the model using two Finnish data sets – data on single-family homes and data on old apartments in blocks of flats. The models for single-family houses covered the whole country. Models for apartments in blocks of flats covered transactions in Helsinki. Both the single-family house models and the models for apartments were estimated for 24 quarters from the beginning of 2002 till the end of 2007. We did not have data on new dwellings containing the necessary information to test our model.

We tested several methods:

1. The time dummy approach, where a single equation is estimated for all quarters and the price indices are constructed from the coefficients of the sales period indicators included in the regression.
2. Adjacent period models, where separate regressions are run for adjacent periods and the index series are constructed by chaining the quarterly price changes. This method allows for changing coefficients of the quality characteristics included in the regressions and avoids the problem of revising the index history as the index is updated for more recent periods.
3. The original model suggested by Diewert (2007).
4. Restricted version of the time dummy models, where the land fraction of total price is set to zero.

For the single-family houses we constructed also an implicit structure price index using the official indices produced by Statistics Finland for single-family houses and for land plots intended for construction of single-family houses.

For the apartments in blocks of flats we also tried a model where location variables were included only in the land price equation, but we were unable to obtain estimates for that kind of specification.

All models we experimented with resulted in qualitatively the same results for both data sets. We obtained overall dwelling price indices showing very similar price development as the official quality adjusted indices produced by Statistics Finland. The estimates for the land price fraction of total price are low – at most 9% for single-family houses and about 5% for the apartments in blocks of flats. The corresponding estimate for the importance of land in total price based on the use of a separate land price index.
which we constructed for the single-family case is considerably higher – 19 % for Finland on average.

The land price indices we estimated are very unstable but because the land fraction of total price is low they do not really affect the overall dwelling price indices. Correspondingly the behaviour of the overall dwelling price indices is dominated by the structure price component. Indeed, it appears that restricting the models we tested by setting the land fraction part of the equation to zero does not affect the estimated dwelling price index.

It seems that the assumption that the hedonic function relating total price of a dwelling to its characteristics can be expressed as the sum of two terms – one describing the structure price and the other one describing the land price – is more problematic than initially appears. Examined as a construction cost relationship it is indeed true that total cost can always be written as the sum of the land plot price and the production costs of the structure. It is also true that if a structure is destroyed (for example burns down in fire), the land plot will have a market price, so one can ex-post derive the price of the structure as the (likely) price of the dwelling before its structure were destructed minus the market value of the land after the destruction. But what we actually observe in our data is the market prices for dwellings and not costs and/or residual values. It is problematic to relate market prices of final goods to producers’ costs. That’s why we have separate consumer price and producer price indices in the first place.

The relation between land price, structure costs and house prices can be described according the following mechanism. Demand and supply of dwellings from the existing stock determines the overall house price level. The equilibrium prices for dwellings will depend in the short run mainly on macroeconomic factors, such as income levels (current and expected), employment and interest rates as well as current and anticipated demographic trends. Material prices, wages, planning expenses etc. are determined on the corresponding production factor markets. Demand for land is determined as a residual: an entrepreneur or a self-builder will pay for land at most the difference between the market price of the dwelling and the costs of structure (including profit margin for the entrepreneur). In the longer run one will of course have repercussions from the production factor markets to the house prices. Abundant supply of construction plots and sufficient labour supply will alleviate pressures on the dwellings prices and for example lack of land plots will push dwellings prices and land prices further up.

In the above framework it seems natural, that an econometric model may fail to separate land and structure price components from data on overall dwellings prices, because there is no separate final demand for structure and land, only final demand for dwellings. The model failure does not have any implications for construction costs. It is completely consistent for example with the fact that prices for construction land plots can be an important and growing fraction of the overall dwelling construction costs.

To recapitulate, our findings suggest that econometric models based on observed total dwelling prices attribute very limited importance to land in the total price of a dwelling. If one accepts the result, the implication is that one should not exclude land from the index measuring costs of owner occupied housing in HICP, since the estimated land price indices are very unstable. This is a problematic conclusion if there is strong evidence that land price is an important and perhaps growing fraction of total
construction costs and we consider it natural that other approaches to separate land and structure prices are examined.

References:


Owner-Occupied Housing
Econometric Study and Model to Estimate Land Prices

Discussion of the Q&A session in the HICP meeting at Vienna on the 8th and 9th of September 2008
Summary and conclusions on the OOH econometric land study presentation

Following the presentation, there was a Q&A session. The following issues were the most relevant that were raised.

Variables controlling the unitary prices

The basic model may be written as:

\[ P^T = p^S S + p^L L \]

Thus, the total house price is comprised of two components: the land price and the structure price. These prices are again the result of the product of unitary prices by the size of the structure size and land plot size, respectively. Unitary prices have an associated regression model. The explanatory variables of the unitary prices are the following.

For the single family houses:

- Regional indicators, on both components of the price equation
- Size of the structure, on both components
- Interactions between size of the structure and municipality size indicators, on the structure component
- Year of construction, on the structure side
- Interaction between year of construction and municipality indicators, on the structure side
- Size of the land plot, on the land component
- Interactions between the size of land plot and municipality size indicators, on the land side
- Distance from municipality centre, on the land side
- Interactions between distance and municipality size indicators, on the land side.

For apartments (in Helsinki):

- Regional indicators, on both components of the equation
- Size of the structure, on the structure component
- Number of rooms indicators, on the structure component
- Year of construction and its square, on the structure component
- Common spaces indicator, on both components
- Land plot size measures, on the land side

The low estimate of 9% for land share value (against an expected value of 19%, which is derived from real data)

There were concerns raised about the low estimate of land share value for the case of single family houses. The econometric model tries to forecast the land price (and value) by having information only on the total house price and the land plot size. It uses explanatory variables in order to better adjust for the differences among land plots (and houses) on the unitary prices (prices per square metre of the structure and of the land...
plot). The low estimate is considered to be due mainly to the fact that there is a very low
correlation between the total house price and the land plot size (on the other hand, there
is a high correlation, as expected, between the total house price and the structure size).
This estimate of 9% already takes into account the fact that prices differ among
locations (and other variables). This issue is again discussed in the extension to the
original report.

The use of linear models

The study is based solely on nonlinear models. There were doubts concerning the fact
that linear models were not presented. The fact is that there is a conceptual problem in
using linear models for this particular framework. The total house price is composed of
two parcels – one corresponding to the structure price and the other corresponding to the
land price. Prices must be positive and the model must ensure that. This can only be
achieved through the use of nonlinear functions.

Sometimes it is possible to relax this restriction because in practical terms it is
unbinding – that is, the linear model would forecast positive prices. Although not
conceptually sound, the advantage of an easier estimation (and, namely, the certainty of
convergence) would support the decision of considering the estimation of such models
and present some results in the study. However, in the present case, forecasts of
negative prices happened regularly due to the fact that the land share value had a low
estimate – meaning that the structure part of the equation was so “well defined” that it
was overestimated and the land part was consequently underestimated (see also the
latter point on the estimate of the land share value). The overall fit turned out better.
Therefore, restrictions on prices had to be included and the use of nonlinear models was
mandatory.

This issue is again dealt with in the extension to the report.

The use of location dummies only on the land part

It seems sensible to use location dummies only in the land part of the model. The use on
both parts of the model was questioned. There are two kinds of arguments for the use of
these dummies on both parts.

a) the empirical approach: empirical findings suggest that regional dummies are
indeed quite important on both parts of the model. T-statistics are quite high,
especially in the structure part. Again, this is probably also a consequence of a
higher correlation between structure size and total house price.

b) the theoretical approach: quality of the structure is often unmeasured but
intuitively it is quite important to define structure prices. Location dummies can
capture many of these unknown effects and so they are included in the model
with this aim. One can reason in the following way to support this view: when
one buys a house in an expensive location, one is sufficiently well off to demand
high quality houses. Therefore, contractors may often build high quality houses
on expensive land plots.
The problem of the land size variable on urban areas

On urban areas, the land plot size variable is not easy to define. In the study, two measures were proposed: the land plot size of the building would be equally divided by the total number of dwellings in the building or it would be proportionally divided, according to the size of each dwelling.

The model would estimate a linear combination of both these measures and so the land size variable would be empirically defined. In the end, the estimated model relied on the first measure of land plot size, due to convergence problems of the other models.

Using this variable (or any other one, as a matter of fact) implies that it may be possible that two identical apartments have a different value for the land size. For instance, consider that there are two apartments in two different buildings, one next to the other, and assume that the apartments are indeed identical. The first building is two storeys high and the second building is five storeys high. Assume also that the land plot size is identical for both buildings. The values for the land variable of the apartments would therefore be different.

It was suggested that they should be different because of building permits, for instance. One might argue that the land size should still be equal and other variables would control for that effect (for instance, a variable relating to building permits or number of floors of the building). There is also another important factor coming into play here. If there are indeed two building with identical or very similar apartments, then the buyer chooses the cheapest one. The equilibrium state is to have equal prices for the dwellings. This leads to the conclusion that it is sensible to consider that the land prices of the dwellings are equal. It is difficult to estimate equal prices when the land size variable for each dwelling is different.

To conclude, there might be a problem with the definition of this variable. Conceptually, it is hard to argue otherwise. In practical terms, it may be negligible. More empirical studies (using data of different countries) are needed to confirm which part plays a more important role.
Owner-Occupied Housing
Econometric Study and Model to Estimate Land Prices

Extensions to the Draft Final Report
I. Discussion on the weight of the land component

One of the surprising results that we have obtained is that the land component has much less weight than expected. This has been regarded as a rather unsatisfactory outcome. However, there may be good reasons to this finding.

A plot of land has certainly very different values for different individuals. In particular, it is reasonable to assume that land is more valuable for professional developers and for individuals that buy it to build their own houses. This is what is captured by the data on transacted land plots – the source for the calculation of an average price per square metre. Nonetheless, when we construct a price index for transacted dwellings, the individuals involved in the transactions are very different. In this case, the relevant agents are the ones that decided to buy an already built house and for whom a plot of land by itself may not be very interesting.

Therefore, for the set of consumers relevant to the construction of house price indices, the value of land may be much lower than for those who are in the market for land plots. The problem here is that there are two sources of information, each one providing data on different purchases and their matching might just prove incoherent. If that is the case, when constructing a price index for structure by implicit derivation (deflation of the house price index by the land price index), the weight that is given to the land component is overestimated because it does not represent valuation of the land by the relevant set of individuals.

Hence it may well be the case that what we observe is not that the econometric model underestimates the weight of the land component but that the benchmark we are using is not appropriate.

II. Alternative econometric approaches: simpler models

The departure point to this chapter is the reasoning that often simpler models give way to better results and, in the process, present the advantage of easier and usually faster calculations. The report has relied heavily on nonlinear models to derive two distinct indices: a structure price index and a land price index. An obvious simplification of the process would be to consider linear models instead of nonlinear models. This section dwells on this issue.

Firstly, there are some clarifications to be made. The original report presents a simple model, nonlinear in its parameters, that is equivalent to a reparameterised linear model. This was thoroughly discussed in the theoretical chapter. Thus, the estimated model (2.2) is identical to a linear model. Diagrams 2.4 and 2.5 are exactly the same whether we estimate a simple nonlinear model or a linear one.

The hedonic model (2.3) is again nonlinear in its parameters. However, there is no identical linear model as in the previous case. One can simplify the model by letting the unit prices in period 0 to be linear, that is, by not using the exponential functions. That formulation was mildly tested in early attempts to decompose the total price into
structure and land price. Those attempts revealed that the linear specification would not conform to basic economic principles by having estimated parameters which implied a negative unit price for the land component. Hence, the formulation with exponentials was adopted.

Nevertheless, it has been discussed there is still a possibility of estimating a different, even simpler, linear model. The following paragraphs discuss this new approach. There were some remarks made stating that even if a linear model can be estimated, it would hardly lead to better results than those presented at the original report.

Anyway, at least to serve as a benchmark, a new linear model was developed. This model would simply have the prices of the dwellings regressed on its characteristics, including the size of structure and plot size. This should then take the following form

\begin{equation}
 p_{nt} = x'_{nt} \gamma_t + \phi_t S_{nt} + \psi_t L_{nt} + \epsilon_{nt}
\end{equation}

where the notation is as before, except that \( \phi_t \) and \( \psi_t \) are new parameters to be estimated.\(^\text{11}\) Estimation of this model can be performed by ordinary least squares but, as for the nonlinear models, it may be interesting to account for heteroskedasticity and use instead weighted least squares. This specification implies that the model is estimated in each period and constitutes therefore quite a different approach from the time dummy one presented at the report. Now, let

\begin{equation}
 p'_{n0} = x'_{n0} \hat{\gamma}_t + \hat{\phi}_t S_{n0} + \hat{\psi}_t L_{n0}
\end{equation}

denote the predicted price for period \( t \) of the dwellings sold in the base period.

The Dutot elementary index for the price change between periods 0 and \( t \) may be written as

\begin{equation}
 I^D_{0t} = \frac{N(0) \sum_{n=1}^{N(t)} p_{nt}}{N(t) \sum_{n=1}^{N(0)} p_{n0} \sum_{n=1}^{N(0)} p_{n0}} = \frac{\sum_{n=1}^{N(0)} \sum_{m=1}^{N(t)} p'_{n0}}{\sum_{n=1}^{N(0)} p_{n0} \sum_{n=1}^{N(t)} \sum_{m=1}^{N(t)} p_{nt}}
\end{equation}

where the second factor

\begin{equation}
 I^D_{0t} = \frac{N(0) \sum_{n=1}^{N(t)} p_{nt}}{N(t) \sum_{n=1}^{N(0)} p_{n0}} = \frac{N(0) \sum_{n=1}^{N(t)} (x'_{n0} \hat{\gamma}_t + \hat{\phi}_t S_{n0} + \hat{\psi}_t L_{n0})}{N(t) \sum_{n=1}^{N(0)} (x'_{n0} \hat{\gamma}_t + \hat{\phi}_t S_{n0} + \hat{\psi}_t L_{n0})}
\end{equation}

captures the effect of quality change. The equality stands because the vector of explanatory variables includes an independent term, thus making the average predicted prices in period \( t \) equal to the average actual prices in period \( t \). The denominator is equal by definition (cf. 4.2). Meanwhile, the first factor of (4.3)

\(^{11}\) Of course, \( S \) and \( L \) can be logged.
captures the pure price change. This is one of the traditional approaches of hedonic regression methods. The innovation is that we intend to break down this expression into two parcels whereas in the traditional approach this expression is final. Therefore, the index in (4.5) still includes the effect of changes in the price of land and the expression should be further decomposed. Let

\[
\tilde{I}_D^{(4.6)} = \frac{\sum_{n=1}^{N(0)} (x_{n00} \hat{y}_0 + \hat{\phi}_0 S_{n0})}{\sum_{n=1}^{N(0)} P_{n0}} + \frac{\sum_{n=1}^{N(0)} (\hat{\psi}_1 L_{n0})}{\sum_{n=1}^{N(0)} P_{n0}}
\]

be that decomposition. The first term would provide a price index net of the land component and the second term would constitute the price index for the land. This formulation encompasses some troubling features, namely the fact that all other regressors (including location indicators) are included in the part of the index that is net of the land component. This may not be appropriate if the location is considered a characteristic of the land and not of the structure. Alternatives may be contemplated but some ambiguity will prevail in the end.

Now, this is a mathematical sound way of decomposing the total price. However, the two terms in (4.6) will not constitute price indices for a very simple reason: the numerators are a fraction of a price whilst the denominator is a full price. As the structure component dominates the decomposition, the first term will numerically look like an index while the second term will hardly resemble an index as estimates of \(\psi\) are very different from period to period and only capture a small fraction of the full price.

Manipulating the basic formula in (4.1), using logs of the price for the dependent variable or exponentials for the right hand side of the equation, will not lead to a formulation that allows a meaningful decomposition. Finally, one might try a more daring approach, in the sense that economic interpretation of the decomposition would be difficult. Let us consider separate denominators in (4.6). In doing that, we are setting aside the decomposition approach – this may be seen as a separate index approach. That is, let us write

\[
\tilde{I}_S^{(4.7)} = \frac{\sum_{n=1}^{N(0)} (x_{n00} \hat{y}_0 + \hat{\phi}_0 S_{n0})}{\sum_{n=1}^{N(0)} \hat{y}_{n0}} , \quad \tilde{I}_D^{(4.7)} = \frac{\sum_{n=1}^{N(0)} (\hat{\psi}_1 L_{n0})}{\sum_{n=1}^{N(0)} \hat{\psi}_n L_{n0}}
\]

In this case, we are apparently comparing like with like. An immediate objection to this approach is that the indices would be completely based on predicted prices. The main problem, besides the feeble economic and index theory framework, is again the fact that parameter estimates are not necessarily stable. Thence there is the possibility of

---

\[12\] This is easily demonstrated, in an alternative way, by considering for example the loglinear case. If we assume that in period \(t\) the prices of land and structure have gone up by \(\alpha\%\) and \(\beta\%\) respectively, we will not be able to identify these parameters apart and thus one cannot estimate two different series for prices evolution.

60
compiling a highly volatile series, which will not even remotely resemble an index series\textsuperscript{13}.

To sum up, the only viable way to have a linear model that is decomposable into two index series is by the time dummy variable method. The experiments with the time dummy method have already been commented on the original report as well as in this chapter, if only in a brief way.

III. Generalisation of the models

3.1 A simple generalisation

The model we have been using (for single-family houses) can be written as

\begin{equation}
(4.8) \quad p_{nt} = \left(\sum_{r=1}^{R} \delta_r D_r + 1\right) \exp(\alpha'x_{nt}^S + \ln S_{nt}) + \left(\sum_{t=1}^{T} \gamma_t D_t + 1\right) \exp(\beta'x_{nt}^L + \ln L_{nt}) + \theta_{nt}
\end{equation}

where the unit prices for the base period are included in the vectors of the regressors ($p^S$ in $x^S$ and $p^L$ in $x^L$). That is we are forcing the regressors $\ln S_{nt}$ and $\ln L_{nt}$ to have coefficients equal to 1. Of course, this restriction can be easily relaxed by estimating

\begin{equation}
(4.9) \quad p_{nt} = \left(\sum_{r=1}^{R} \delta_r D_r + 1\right) \exp(\alpha'x_{nt}^S + \eta S_{nt}) + \left(\sum_{t=1}^{T} \gamma_t D_t + 1\right) \exp(\beta'x_{nt}^L + \lambda t \ln L_{nt}) + \theta_{nt}
\end{equation}

The attractive feature of this model is that, not only it is more general than what we have been assuming, but it also addresses the point that land plot size might not matter at all. In that case, location is the only relevant factor in the land part of the equation and hence the parameter $\lambda_t$ should have an estimate statistically equal to 0.

Estimation of model (4.9) has not been possible due to the singularity of the Hessian matrix. Several distinct initial values for the parameters were used, ultimately with no success. Nonetheless, the software package produces consistent estimates for most parameters and results are presented in the following tables and diagrams for the sake of exposition. A wide range of variations of model (4.9) has been tried out – including the time dummy variables inside the exponential function, grouping some location dummies, dropping some variables, etc. – always with the same outcome in what concerns the singularity of the Hessian matrix.

\textsuperscript{13} Both (4.6) and (4.7) were tested empirically to have a feel of what the indices might look like. The structure component, as it captures most of the price, leads to series resembling a price index in both cases. On the other hand, the land component leads to series highly volatile, presenting even negative values in some instances.
Assuming the results are completely interpretable, which they are not due to singularity issues, there is a very troubling result – the bigger the land plot, the lower the price of the house. While statistically this is a natural result, economically it is not acceptable. Of course, this is only troubling if the outcome stands to full estimation of the model. We will present a simpler model based on the same principles, fully estimated, also presenting a negative sign for the land plot area parameter.
Hence, the results for the total price index are to be examined very carefully. The estimated weights depend a great deal on the initial values used in the estimation process as the model is not fully estimated.

The following diagram presents the results in the form of indices, based on the results presented in table 4.1. The estimated weight for the structure component is in this case 88.34% and consequently the weight for the land component is estimated to be 11.66%.

Diagram 4.1: The total price index and the indices for the components computed from the generalised model (4.9)

One of the main arguments devised when studying the current problem was that the generalised model (or any alternative model thought up for those particular circumstances) would tend to have an estimated parameters for the land plot area variable very close to 0, thus implying that the size of the land plot does not matter in the price formation of the house. To test this hypothesis, we have estimated a model very similar to (4.9) but forcing the parameter $\lambda_t$ to be zero, that is

$$ p_{mt} = \left( \sum_{i=1}^{T} \delta_i D_i + 1 \right) \exp(\alpha' x_{mt}^s + \eta_t \ln S_{mt}) + \left( \sum_{i=1}^{T} \gamma_i D_i + 1 \right) \exp(\beta' x_{mt}^l) + \theta_{mt} $$

Again, this model is not fully estimated due to the singularity of the Hessian matrix. The troubling features of model (4.9) remain in this case and assuming results are interpretable, the precision of estimates has improved.

For the same model, we have obtained vectors of weights differing as much as from 90%-10% to 60%-40% for the structure and land components respectively.
### Table 4.2: Estimation results for the weighted regression model (4.10)

The next diagram shows the indices computed from this model. The estimated weights for structure and land are now 88.2% and 11.8% respectively.

<table>
<thead>
<tr>
<th>Sales quarter dummies</th>
<th>Structure component</th>
<th>Land component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coeff</td>
<td>str.error</td>
</tr>
<tr>
<td>2002/1</td>
<td>0.020</td>
<td>0.016</td>
</tr>
<tr>
<td>2002/2</td>
<td>0.028</td>
<td>0.017</td>
</tr>
<tr>
<td>2002/3</td>
<td>0.043</td>
<td>0.018</td>
</tr>
<tr>
<td>2003/1</td>
<td>0.094</td>
<td>0.019</td>
</tr>
<tr>
<td>2003/2</td>
<td>0.095</td>
<td>0.017</td>
</tr>
<tr>
<td>2003/3</td>
<td>0.110</td>
<td>0.017</td>
</tr>
<tr>
<td>2003/4</td>
<td>0.122</td>
<td>0.017</td>
</tr>
<tr>
<td>2004/1</td>
<td>0.175</td>
<td>0.018</td>
</tr>
<tr>
<td>2004/2</td>
<td>0.202</td>
<td>0.017</td>
</tr>
<tr>
<td>2004/3</td>
<td>0.229</td>
<td>0.018</td>
</tr>
<tr>
<td>2004/4</td>
<td>0.299</td>
<td>0.019</td>
</tr>
<tr>
<td>2005/1</td>
<td>0.341</td>
<td>0.017</td>
</tr>
<tr>
<td>2005/2</td>
<td>0.342</td>
<td>0.017</td>
</tr>
<tr>
<td>2005/3</td>
<td>0.376</td>
<td>0.017</td>
</tr>
<tr>
<td>2005/4</td>
<td>0.405</td>
<td>0.016</td>
</tr>
<tr>
<td>2006/1</td>
<td>0.351</td>
<td>0.018</td>
</tr>
<tr>
<td>2006/2</td>
<td>0.341</td>
<td>0.017</td>
</tr>
<tr>
<td>2006/3</td>
<td>0.342</td>
<td>0.017</td>
</tr>
<tr>
<td>2006/4</td>
<td>0.376</td>
<td>0.017</td>
</tr>
<tr>
<td>2007/1</td>
<td>0.395</td>
<td>0.016</td>
</tr>
<tr>
<td>2007/2</td>
<td>0.399</td>
<td>0.016</td>
</tr>
<tr>
<td>2007/3</td>
<td>0.427</td>
<td>0.017</td>
</tr>
<tr>
<td>2007/4</td>
<td>0.415</td>
<td>0.017</td>
</tr>
</tbody>
</table>

| Regional indicators | coeff | str.error | t     | pr>|t| |
|---------------------|-------|-----------|-------|--------|
| Helsinki            | -0.219| 0.080 | -2.76 | 0.006 |
| Espoo               | 0.167 | 0.038 | 4.41 | 0.000 |
| Greater Helsinki    | 1.661 | 0.078 | 21.34 | 0.000 |
| Jyväskylä          | 0.536 | 0.073 | 7.30 | 0.000 |
| Kuopio              | 0.588 | 0.071 | 8.23 | 0.000 |
| Lahti               | 0.295 | 0.055 | 3.11 | 0.000 |
| Oulu               | 0.450 | 0.081 | 5.58 | 0.000 |
| Pori              | 0.406 | 0.070 | 5.78 | 0.000 |
| Tampere           | 0.945 | 0.068 | 13.85 | 0.000 |
| Turku             | 0.746 | 0.072 | 10.36 | 0.000 |
| Uusimaa         | 0.454 | 0.013 | 35.94 | 0.000 |
| Eastern Finland    | -0.189 | 0.008 | -22.38 | 0.000 |
| Northern Finland  | -0.251 | 0.014 | -18.14 | 0.000 |
| Central Finland  | -0.170 | 0.009 | -19.87 | 0.000 |
| Municip. over 50 000 inhabitants | -0.789 | 0.073 | -10.85 | 0.000 |
| Municip. less than 10 000 inhabitants | -0.350 | 0.019 | -18.69 | 0.000 |

| Characteristics of the house | coeff | str.error | t     | pr>|t| |
|-----------------------------|-------|-----------|-------|--------|
| size of structure          | -0.001 | 0.000 | -8.41 | 0.000 |
| (size of structure)*(Municip. over 50 000 inh.) | 0.001 | 0.000 | 4.86 | 0.000 |
| (size of structure)*(Municip. lt 10 000 inh.) | 0.001 | 0.000 | 5.77 | 0.000 |
| (size of structure)*(Greater Helsinki) | -0.001 | 0.000 | -6.51 | 0.000 |
| Year of construction - 1900 | 0.002 | 0.000 | 19.74 | 0.000 |
| Year of construction missing | 0.175 | 0.020 | 8.65 | 0.000 |
| (Year of construction - 1900)*(Greater Helsinki) | -0.002 | 0.000 | -6.06 | 0.000 |
| size of land plot           | -0.001 | 0.000 | -8.49 | 0.000 |
| (size of land plot)*(Municip. over 50 000 inh.) | 0.001 | 0.000 | 5.19 | 0.000 |
| (size of land plot)*(Municip. lt 10 000 inh.) | 0.001 | 0.000 | 4.97 | 0.000 |
| (size of land plot)*(Greater Helsinki) | 0.000 | 0.000 | 2.14 | 0.032 |
| distance from municipality centre | 0.119 | 0.027 | 4.37 | 0.000 |
| (distance from municipality centre)*(Municip. over 50 000 inh.) | -0.008 | 0.002 | -4.53 | 0.000 |
| (distance from municipality centre)*(Municip. lt 10 000 inh.) | -0.138 | 0.019 | -7.15 | 0.000 |
| In (area structure) | 0.417 | 0.016 | 26.88 | 0.000 |
| In (area land plot) | 9.480 | 0.063 | 149.78 | 0.000 |

Number of obs 61 377. R squared is 0.132
Standard errors and t-values are computed by using robust covariance estimate
Another criticism to the models in the original report was that each one of them included location dummies on both components. This issue has been discussed thoroughly in the original report and commented on the small discussion paper following the presentation of first results in the HICP meeting at Vienna in the beginning of September. Nonetheless, we have estimated a model with location variables only on the land component. The new model is as (4.9) but the vector $x^3$ does not include those location variables.

Results are presented in the following table. This time, the model is fully estimated. Two main features result from the estimation. Firstly, the location variables determine to a large extent the formation of prices, thus making the land component the most important to that purpose. The precision of the estimates of the parameters associated to the land component improves greatly while, simultaneously, the precision of the estimates referring to the structure component deteriorates. The second relevant feature is that, despite the model being fully estimated and convergence has been achieved, the land plot area parameter is again estimated to be negative. The outcome is economically uninteresting, to say the least. Therefore, analysis based on this model should still be made carefully.
### Table 4.3: Estimation results for the weighted regression model (4.9) with location variables appearing exclusively on the land component

The following diagram presents the indices associated to the model. The weights for the components appear reversed as has been suggested previously – 9% refer to the structure and 91% to the land component.
Diagram 4.3: The total price index and the indices for the components computed from the generalised model (4.9) with location variables only on the land component

The following diagram compares all three previous models. Despite the fact that the first two are not fully estimated, it still stands out that results are robust to minor changes of specification. The third model relies on a completely different specification. The series are now very different and prove an important point: location variables are fundamental to determine the price of a house. When using these variables only on the land part of the equation, the erratic movement of the land price series is now attributable to structure instead. Thus, the estimated weight is the opposite of what has been so far. Finally, whatever the models, the estimated total price index always follows closely the official house price index, reinforcing the idea that the main difficulty is not to come up with a good model for final prices but to be able to decompose those prices into two reliably estimated components.

Diagram 4.4: The models compared
3.2 A more complex generalisation

A further generalisation of model (4.9) is possible. Let us consider that

\[(4.11) \quad E(p_{nt} \mid x^S_{nt}, x^L_{nt}, S_{nt}, L_{nt}) = m\]

where \(m\) is defined as in equation (4.9), that is

\[(4.12) \quad m = \left( \sum_{i=1}^{T} \delta_i D_i + 1 \right) \exp(\alpha' x^S_{nt} + \eta_i \ln S_{nt}) + \left( \sum_{i=1}^{T} \gamma_i D_i + 1 \right) \exp(\beta' x^L_{nt} + \lambda_i \ln L_{nt})\]

If we now let the function to be maximised by maximum likelihood take the following form

\[(4.13) \quad \log l = -\frac{p_{nt}}{m} - \ln(m)\]

we are estimating what is usually known as a general linear model. There are obvious advantages to this model. Firstly, the fact that is a generalisation to the previous models and secondly, the fact that the model accommodates automatically the inherent heteroskedasticity.

Once again, there is a problem of singularity of the Hessian matrix. The model is not fully estimated but results are anyway presented in table 4.4. The troubling features mentioned earlier repeat themselves in the present case. The land plot area parameter is estimated to have a negative sign.
| Sales quarter dummies | coeff | str.error | t | pr>|t| | coeff | str.error | t | pr>|t| |
|-----------------------|-------|-----------|---|-------|-------|-----------|---|-------|
| 2002/1                | 0.003 | 0.032     | 0.08 | 0.937 | 0.016 | 0.054     | 0.30 | 0.763 |
| 2002/2                | 0.057 | 0.034     | 1.66 | 0.097 | -0.072 | 0.063     | -1.15 | 0.252 |
| 2002/3                | 0.071 | 0.037     | 1.92 | 0.055 | -0.055 | 0.066     | -0.83 | 0.404 |
| 2003/1                | 0.115 | 0.033     | 3.46 | 0.001 | 0.012 | 0.060     | 0.20 | 0.838 |
| 2003/2                | 0.182 | 0.036     | 4.98 | 0.000 | 0.074 | 0.065     | 1.14 | 0.256 |
| 2003/3                | 0.082 | 0.033     | 2.49 | 0.013 | -0.072 | 0.063     | -1.15 | 0.252 |
| 2004/1                | 0.167 | 0.035     | 4.72 | 0.000 | 0.140 | 0.061     | 2.30 | 0.021 |
| 2004/2                | 0.157 | 0.034     | 4.58 | 0.000 | 0.211 | 0.058     | 3.67 | 0.000 |
| 2004/3                | 0.171 | 0.033     | 5.17 | 0.000 | 0.174 | 0.057     | 3.06 | 0.002 |
| 2005/1                | 0.260 | 0.038     | 6.85 | 0.000 | 0.250 | 0.063     | 4.74 | 0.000 |
| 2005/2                | 0.259 | 0.037     | 6.70 | 0.000 | 0.264 | 0.059     | 4.48 | 0.000 |
| 2005/3                | 0.241 | 0.034     | 7.16 | 0.000 | 0.257 | 0.056     | 4.56 | 0.000 |
| 2006/1                | 0.279 | 0.036     | 7.84 | 0.000 | 0.258 | 0.060     | 4.29 | 0.000 |
| 2006/2                | 0.313 | 0.038     | 8.32 | 0.000 | 0.385 | 0.061     | 6.30 | 0.000 |
| 2006/3                | 0.332 | 0.034     | 9.45 | 0.000 | 0.368 | 0.056     | 6.63 | 0.000 |
| 2006/4                | 0.315 | 0.034     | 9.27 | 0.000 | 0.366 | 0.057     | 6.45 | 0.000 |
| 2007/1                | 0.359 | 0.036     | 9.93 | 0.000 | 0.386 | 0.058     | 6.67 | 0.000 |
| 2007/2                | 0.352 | 0.036     | 9.65 | 0.000 | 0.466 | 0.058     | 8.08 | 0.000 |
| 2007/3                | 0.392 | 0.034     | 11.67 | 0.000 | 0.425 | 0.056     | 7.64 | 0.000 |
| 2007/4                | 0.383 | 0.034     | 11.28 | 0.000 | 0.355 | 0.058     | 6.17 | 0.000 |
| 2008/1                | 0.418 | 0.034     | 12.15 | 0.000 | 0.377 | 0.059     | 6.35 | 0.000 |

| Regional indicators | coeff | str.error | t | pr>|t| |
|---------------------|-------|-----------|---|-------|
| Helsinki            | -0.410 | 0.160     | -2.56 | 0.011 |
| Espoo               | 0.289 | 0.017     | 17.16 | 0.000 |
| Greater Helsinki    | 1.966 | 0.183     | 10.76 | 0.000 |
| Jyväskylä           | 0.457 | 0.067     | 6.46 | 0.000 |
| Kuopio              | 0.547 | 0.085     | 6.45 | 0.000 |
| Lahti               | 0.263 | 0.123     | 2.15 | 0.032 |
| Oulu                | 0.334 | 0.107     | 3.13 | 0.002 |
| Pori                | 0.515 | 0.126     | 4.07 | 0.000 |
| Tampere             | 1.008 | 0.111     | 9.07 | 0.000 |
| Turku               | 0.828 | 0.106     | 7.85 | 0.000 |
| Municipal. around Greater Helsinki | 0.529 | 0.042     | 12.52 | 0.000 |
| Uusimaa             | 0.197 | 0.044     | 4.51 | 0.000 |
| Eastern Finland     | -0.042 | 0.035     | -1.21 | 0.227 |
| Northern Finland    | -0.084 | 0.027     | -3.13 | 0.002 |
| Central Finland     | -0.257 | 0.039     | -6.64 | 0.000 |
| Municipal. over 50 000 inhabitants | -0.285 | 0.114     | -2.49 | 0.013 |
| Municipal. less than 10 000 inhabitants | 0.097 | 0.048     | 0.23 | 0.042 |
| Characteristics of the house | -0.001 | 0.000     | -3.19 | 0.001 |
| (size of structure)* (Municip. over 50 000 inh.) | 0.000 | 0.000     | 0.04 | 0.972 |
| (size of structure)* (Municip. lt 10 000 inh.) | 0.000 | 0.000     | -3.09 | 0.002 |
| (size of structure)* (Greater Helsinki) | -0.002 | 0.000     | -6.40 | 0.000 |
| Year of construction - 1900 | 0.003 | 0.000     | 9.85 | 0.000 |
| Year of construction missing | 0.270 | 0.036     | 7.40 | 0.000 |
| ln (area structure) | 0.634 | 0.050     | 12.63 | 0.000 |
| ln (area land plot) | -0.144 | 0.025     | -5.73 | 0.000 |

Number of obs 61 377.
Standard errors and t-values are computed by using robust covariance estimate

Table 4.4: Estimation results for the generalised linear model 1

The indices are shown in diagram 4.5 ahead. Due to the initial values chosen and the fact that the location parameters do not have reliable estimates\(^\text{15}\), the estimate for the land component is a lot larger than in previous models. The weight for the land component is 36.5% while for the structure component it is 63.5%.

\(^\text{15}\) As it is possible to confirm when having total distinct estimates as different initial values are used in the estimation process.
Diagram 4.5: The total price index and the indices for the components computed from the generalised linear model (4.13)

Another generalisation of (4.9) was tried out. The equations (4.11) and (4.12) stand along with

(4.14) \[ \log l = -m + p_{\text{w}} \cdot \ln(m) \]

The results follow but the conclusions are exactly the same as with model (4.13).
### Table 4.5: Estimation results for the generalised linear model 2

The weight of the structure component is estimated to be 58.4% and of the land component 41.6%.
Diagram 4.6: The total price index and the indices for the components computed from the generalised linear model (4.14)

Despite the fact that the models were not fully estimated, it is remarkable how robust the results appear to be, as can be inferred in the diagram below.

Diagram 4.7: The generalised linear models compared
IV. Conclusions

The extension to the original report was decided after a discussion where the parties involved agreed that some more tests were in need. One of the main points was the fact that linear models were not tested appropriately to begin with. Moreover, some relevant specifications were also not tested, namely a specification having location variables attributed only to the land part. Finally, generalisations of the studied models might be tried out, just in case it might improve a lot the estimation and consequent outcome.

As for the linear models, we discuss that they were indeed tried out in the original report revealing some very negative aspects. A simple model was estimated and is part of the report but when one tried to generalise the model to include regressors relating to characteristics of the house and land the model failed to produce economic sound results. The estimates implied a negative unit price for land. For the extension, one of the main points was to present simpler linear models. However, after a thoughtful analysis, we were not able to come up with an identifiable linear model allowing the decomposition of house prices into two distinct index series. To our understanding, this is not possible to do unless we use the time dummy variable approach, which was already discussed in the original report.

The generalisation of models in chapter 2 of the original report was pursued in this report. Unfortunately, this was done without great success. Most generalised models have singularity problems with the Hessian matrix rendering the full estimation unviable. Results were presented nonetheless so that some of the problems the researchers face are made clearer to readers. Not having full estimations, no conclusions can really be derived from these models.

There is however a generalised model that was possible to estimate. Allowing the structure and land plot size variables to have an associated parameter to be estimated along with a specification that dropped the location variables from the structure component made the estimation possible. This model highlights one important point: location is to a large extent one of the main determinants of prices. Dropping the location variables from the structure part has improved the precision of estimates on the land component variables while deteriorating the estimates of the structure component variables. An important consequence of this specification is that the estimated weight for the land price index is around 90% of the total price index. Another troubling feature of the estimated model is that the land size parameter is estimated to be negative – this result collides with what is traditionally accepted in economic theory.

Finally, two points are worth mentioning. Many trials and specifications were tried out. They are not presented here because they do not add relevant information to what has been already discussed. Adding to that, all estimated models in the present report refer to single-family houses. The same procedures were applied to flats, some more variations were tried out (as for the flats case there are additional features to account for) but the results are essentially in line with what has been written before in the original report and in this extension. Lastly, it is unfortunate that the new experiments with the data have not led to more satisfactory results.
References:

