Measuring the Capital Stock of Dwellings by using Hedonic Regression

Christian Gysting
(Statistics Denmark)

Jonas Johansen Næsby
(Statistics Denmark)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

Session PS3: Measuring Capital and Wealth

Time: Wednesday, August 22, 2018 [17:30-18:30]

Paper Prepared for the 35th General Conference of

The International Association for Research in Income and Wealth

Copenhagen, Denmark, August 20 – 25, 2018.

Measuring the Capital Stock of Dwellings by using Hedonic Regression

Christian Gysting and Jonas Johansen Næsby

For additional information please contact:

Christian Gysting
Statistics Denmark, National Account division
Sejrøgade 11, DK-2100 Copenhagen, Denmark
E-mail: cgy@dst.dk
Telephone: +45 3917 3665

Jonas Johansen Næsby,
Statistics Denmark, National Account division
Sejrøgade 11, DK-2100 Copenhagen, Denmark
E-mail: jjn@dst.dk
Telephone: +45 3917 3477

Abstract

The aim of this paper is to compare results of Statistics Denmark's current estimations of capital stocks for dwellings for the household sector with an alternative measurement approach based on hedonic regression model. Statistics Denmark uses a variation of the PIM approach for measuring the capital stocks for dwellings. The PIM model differ from traditional PIM model because it also incorporate information number of square meters and price per square meter, which gives an estimate of the gross stock.

The results from the PIM model for dwellings are compared to the results from the hedonic regression model. The hedonic regression model use information at individual level for all Danish dwellings owned by private households (S.14) together with information on property price, type of dwelling, location (zip-code and municipality), (quality adjusted) square meters of buildings and square meters of land. Outcome of the hedonic regression model are prices per square meter which are matched with number of square meters of dwellings, resulting in an estimate for the total value of dwellings. Results by the two approaches are compared for the period 2010 – 2015.

Chapter 1: Introduction

The aim of this paper is to compare (partial) results of Statistics Denmark's current estimations of capital stocks for dwellings with an alternative measurement approach based on hedonic regression model. Using a hedonic regression model is an alternative to a traditional Perpetual Inventory Method (PIM) calculation of capital stocks for dwellings, although probably never used by statistical offices in large scale.

The hedonic regression model for measuring the value of buildings is basically built upon a regression model with a set of independent explanatory variables and a dependent variable. In this case, the square meters of land and (quality adjusted) square meters of buildings are the explanatory variables and the combined value of buildings and land is the dependent variable.

The idea behind the hedonic regression model is to estimate prices per square meter of building, and calculate the entire capital stock by doing a price (price per square meter) multiplied with quantity estimation (number of square meter of dwellings). For practical purpose, the calculation is limited to the dwellings owned by the household sector (S.14).

Using a hedonic regression model for real estate data would normally be seen as a step towards separation the combined value of land and buildings and give the value of land as a result. As a secondary result, the value of buildings is also output of the hedonic regression model, if properly specified. Measuring the value of buildings by using the hedonic regression model gives a valuable alternative to the traditional PIM calculation of the value of buildings. If the PIM-model produces questionable results, the hedonic regression model could be used as an alternative. Otherwise it could serve as a check on the results from the PIM model and maybe help identify potential problems within the PIM calculation.

The hedonic regression model in this paper is used for deriving the level of prices for existing dwellings, but with few adjustments of the model, one would also be able to measure changes in prices for existing dwellings. This price index could serve as an alternative measure for price changes in the PIM model. A present, Statistics Denmark use a price index for gross fixed capital formation for dwellings which measure price changes for new buildings.

Our work on a hedonic regression models for dwellings is inspired by research by Diewert et al. and writings from the Eurostat Handbook on Residential Property Prices Indices. The research by Diewert et al. and the Eurostat Handbook contain the pioneer research regarding hedonic regression models for subdividing the combined value of land and buildings into a building and land component or measure the price development for respectively the building and land component. Further, this paper builds upon earlier work done at Statistics Denmark on measurement of the value of land by using the hedonic approach. Eurostat financed part of this work through a grant support and Magnus B. Eriksen from Statistics Denmark did a significant contribution. The authors have received valuable support on applying the regression model from Maria R. Holm, Simon B. Halifax and Dan Knudsen, all from Statistics Denmark. Jesper Lauritzen (Statistics Denmark) has provided valuable support regarding the BBR-register.

Views expressed in this paper are those of the authors and do not necessarily reflect those of Statistics Denmark.

Chapter 2.1: The basic dataset and the main sources

The basic dataset behind the hedonic regression model is gathered by using information from different administrative records which Statistics Denmark can access. Some of the data has already been subject to processing by others inside Statistics Denmark before it involves in our work. Other data has not been subject to a significant amount of processing.

In the gathering of the basic dataset, 2 major sources are used. The first is statistics on Market Value for Households Real Estate. Statistics Denmark compiles statistics on Market Value for Households Real Estate by using estate data on individual level.

The other major source is the BBR-register. The BBR-register (Danish: Bygnings- og Boligregistret (BBR)) is a register for all dwellings and non-residential buildings in Denmark. The BBR-register contains among other things information on size, location and use of the buildings in Denmark. The register is founded in 1977. All new buildings and extension of old buildings should be registered in the BBR-register. A significant part of the data gathered for the hedonic regression model used comes from this register.

Because data often contains large amounts of records and is gathered from different sources, a great deal of complexity is involved with the managing of the data. A significant number of records with insufficient information often appear even though data is merged by using unique records. Noise from the merging data sets adds up and must be removed. Merging and handling large datasets are technically demanding.

In total, the basic dataset contain 2.363.867 records for the year 2010. For the years 2011-2015, the number of records is on similar levels.

The main variables in our basic dataset are the following; 1) Type of building, 2) Square meter of building, 3) Square meter of land, 4) Age of building, 5) Combined market value and 6) Location (zip-code/ municipality). In order to ease the running of the regression model, the variables are further grouped, which is explained in next chapter. The variables are described below.

Type of building (dwelling): The basic dataset contains a subdivision of type of buildings into 46 different types. Examples are cottages and condos. The variable can be traced back to BBR-register based on the “Benyt” variable which has been further processed before it is used.

Square meter of building: The variable measures the size of the building in square meters. The variable is based on information from the BBR-register.

Square meter of land: Measure the size of land for real estate. The variable is based on information from the Danish Cadastral Register.

Age of building: Measuring the year when the construction of the building is completed. If a building is expanded, the age of the building is the year the building is first reported constructed, not the year the extension is completed.

Combined market value: The combined market value measures the combined value of building and land for a real estate. The combined market value is compiled at individual level for each real estate. For most buildings, the measure is a constructed estimate based on tax assessment adjusted for potential bias. The bias is measured as the ratio between tax assessments for traded real estate and actual sales prices – within the same area and for the same type of building. For real

estate which has actually been sold during the reference period, the actual sale price is used as estimate for the combined market value of the real estate.

Zip-code (Location): From the BBR-register, Zip-code location of the real estate can be identified. At most detailed level, the zip-code for each real estate can be identified and this information is included in our basic dataset.

Municipality (Location): As an alternative to zip-code, information of municipality is collected. There is no unique connection between zip-code and municipality, which means that the municipality-variable must be collected independently. Information on Municipality is gathered from the BBR-register.

The basic dataset is derived for the years 2010-2015, and is subject to further processing, see chapter 2.2.

Chapter 2.2: Data analysis and processing

In order to have the best use of the basic dataset in a regression model, some adaptation of the data is needed. Aggregations are introduced for the variables "type of building" and "zip-code". For instance is the number of observations within some zip-codes too small and as a consequence some zip-codes are merged into larger groups. This chapter describes the additional adjustment made to the basic dataset, including some needed "cleaning" of the dataset because of missing information.

Quality adjustment of the square meters of buildings

We assume that one square meter of building depend on the age of the building. We introduce a quality adjustment of the square meters of buildings, implying older buildings has less quality compared to new buildings. The quality adjustment depends on the depreciation rate and age of the building. The exact calculation of the quality adjustment is shown in chapter 3.

Geography and location

The basic dataset has a subdivision on 935 zip-codes. The central Copenhagen area alone has about 350 of these zip-codes. Within some of these subgroups there are too few observations to run the hedonic regression model; therefore they are merged into fewer zip-groups. In total we end up with a total of 591 zip-groups.

Even though an aggregation of zip-codes reduces the number of sub-groups, higher levels of aggregation are needed. Together with a break down on Municipalities, aggregations on Province and Regions are introduced. A unique link between Municipality, Province and Region exist and is based on Statistics Denmark's classification. In contrast, no unique link can be established between zip-code and municipality.

Table 1 **Number of districts in each subgroup for location**

	Zip-code	Municipality	Province	Region	Denmark
Districts	591	98	11	5	1

Type of buildings

The basic dataset has 46 different types of buildings. The 46 different types of buildings are aggregated into 10 categories. The categories are:

- Single/double/triple family houses
- Residential buildings (combined with) and business property
- Farmland
- Summer houses
- Building sites
- Public property, campsites, summer camps etc.
- Condos (i.e. resident owned flats) and business owned flats
- Factory owned flats and condos in family houses etc.
- Private institutions and buildings on foreign land etc.
- Other property

The reduction in the number of buildings increased the number of observations in the subgroups, which make the use of hedonic regression model more durable. It is the expectation that the aggregation does not match buildings with significant differences.

Missing information

The basic dataset from 2010 has 2.363.867 observations, see the table below. Of the 2.363.867 observations, 516.032 observations have no information at all on the finest subdivision of location (zip-codes), type of building and combined market value. The high number of missing values means that a lot of noise has entered the dataset when extractions from the different sources have been collected and matched.

Table 2 **Records with mission information**

Variable	Information is available	Information is missing	Total number of records
Zip-code	1 847 835	516 032	2 363 867
Municipalities	2 363 582	285	2 363 867
Building types	1 847 835	516 032	2 363 867
Construction year	1 759 933	603 934	2 363 867
Combined market value	1 847 835	516 032	2 363 867
Building (square meters)	1 881 151	482 716	2 363 867
Land (square meters)	2 201 375	162 492	2 363 867

In the “cleaned” dataset, which contain a full set of information for all variables (zip-code, municipalities, building category, year of construction, combined market value, number of (quality adjusted) square meters of buildings and number of square meters of land) we end up with only 1.715.452 observations. The “cleaning” reduces the dataset significantly, 516.032 observations is removed for the year 2010.

Additionally we have a subsample of 132.383 records outside the cleaned dataset, which cannot be used for the hedonic regression model based on our requirement that a record has a full set of information. In the subsample we only have information on two of the following three variables; 1) combined market value, 2) square meters of land and 3) square meters of building. The subsample will be used as supplement, with the parameters added from the hedonic regression model.

Treatment of land without a building

For land without a building (approximately 99.000 of 132.383 observations), the split of the combined market value into a land and building component should all be attributed towards the land-component. In practice, these observations are not part of the cleaned dataset, and the split is not derived by using hedonic regression model. See more in the section on supplement no. 1.

This chapter has focused on the adjustments to the basic dataset for the year 2010, but the treatment and impact on the adjusted dataset for the years 2011 - 2015 is similar.

Chapter 3: The hedonic regression model

In this chapter is explained how the hedonic regression model is specified. The variables used for the regression model are square meters of land and buildings together with the combined value of land and building. The simplest hedonic regression model is

#1 $$P_i^P = P^B B_i + P^L L_i + \epsilon_i, \quad i = 1, \dots, n.$$

The goal from the model is to estimate the two parameter P^B and P^L which describe the value per square meter for building and land, respectively. The input for the model is the depended variable P_i^P; combined marked value for observation i. The independent variables are B_i building and L_i land sizes and an error term ϵ_i.

The simplest form does not take the age of building into account. This is important as one square meter of a new building may be expected to have a higher value than an older building. Therefore we formulate an expanding model

#2 $$P_i^P = P^B(1 - \delta A_i)B_i + P^L L_i + \epsilon_i, \quad i = 1, \dots, n.$$

The only thing that is changed is this term; $1 - \delta A_i$. A_i is the age of building i and δ is the yearly depreciation rate given as

#3 $$\delta = \frac{\text{Consumption of fixed capital}}{\text{Net capital stock} + ½ \cdot \text{Gross fixed capital formation}}$$

The whole term $(1 - \delta A_i)B_i$ can be interpreted as quality adjusted square meter for buildings.

It is the expanded model (equation #2) we use to derive estimates for square meters of land and (quality adjusted) square meters of buildings (P^B and P^L) . The estimated parameters are interpreted as the price for a square meter for building and land. These estimates are then multiply by the observed building and land size which match the model within building type and type of location.

Calculation for each level of location

From the expanding model above the hedonic regression will be done over groups of building types and type of location. That means that for every combination of building type and type of location a model has to be estimated.

Criteria for the model and parameters

For each model a check on the estimated parameters etc. are made. If the hedonic regression model meets our criteria the estimated parameters are used. Below are listed the three criteria which each hedonic regression model must be in accordance with:

1. **Parameters greater than zero**. A requirement which secure prices for land and buildings are positive. Negative values for land and building are only economically meaningful in rare cases. Intensive pollution could be an explanation for negative values of land.
2. **A minimum number of degrees of freedom of 30**. The criteria ensure that of minimum number of records are devoted to the estimations of the parameters. In our case it is set to at least 15 for each independent variable, which is equal to 30 in total.
3. **$R^2 \geq ½$**. The requirement ensures sufficient linearity between the hedonic regression model and the data.

If the hedonic regression model does not meet the criteria; the parameters are estimated on more aggregated level, see next section.

It can be discussed if R^2 make sense as a criteria for our model as we fit without any intercept. We have decided that it can and that it is a fine approximation for how linear our model is, despite the lack of intercept.

A problem can arise, if one tries to compare R^2 between a model without an intercept to a model with an intercept, as the calculation of R^2 is different in the two models.

A model without an intercept R^2 is calculated as $R^2 = 1 - \frac{\sum_i \left(P_i^P - \widehat{P_i^P}\right)^2}{\sum_i {P_i^P}^2}$, where P_i^P is the observed value and $\widehat{P_i^P}$ is the estimated value for observation i.

In contrast the model with an intercept $R^2 = 1 - \frac{\sum_i \left(P_i^P - \widehat{P_i^P}\right)^2}{\sum_i \left(P_i^P - \overline{P^P}\right)^2}$, where $\overline{P^P}$ is the mean of the observed data[1].
The denominator for the model without an intercept will therefore be larger and therefore R^2 smaller than for a model with an intercept.

Expanding the regression groups

Not all models and parameters fulfill the criteria on zip-code level, so we decided to increase the number of observation by choosing another type of location. For models which not are accepted, the model are estimated at a higher level of aggregation for location which contains more observation than the previously type of location did – this continues until parameters for all records are accepted.

We start with the finest location division which is zip-codes and moving to less divided location, as municipalities, provinces, regions and the whole country.
In the last model we leave out building types i.e. Denmark*.

The process goes as follows; each model are being estimated on the following groups of building types and locations, starting with building types × zip-codes, models which are not accepted are being estimated on the next group building types × municipalities, and so on. The following groups are building types × provinces, building types × regions, building types × Denmark and the last group which is Denmark*, where the type of buildings are left out of the equation.

[1] $\overline{P^P} = \frac{1}{n}\sum_{i=1}^{n} P_i^P$

Table 3 **Maximum numbers of models**

Type of location:	Zip-codes	Municipalities	Provinces	Regions	Denmark	Denmark*
Models (building × location)	5910	980	110	50	10	1

Normally, it would be considered optimal to estimate parameters at the most detailed level. In our case it means at the level of type of building and zip-code. However, for various reasons it may not always be possible, i.e. too few observations in the subcategory or the parameters are negative. In these cases, estimations at a higher level of aggregation for location must be done as an alternative. In total, 7.061 hedonic regression models are calculated for each year.

By applying this approach, the combined market value is distributed on buildings and land for 1.715.452 observations (2010). The table below shows at which level of aggregation for location each type of buildings is determined. It can be seen that most observations 1.603.930 out of 1.715.452 are determined by most detailed level of aggregation, this correspond to 93.5 per cent. For the largest group of buildings (Single/double/triple family houses) almost all observations are determined at most detailed level of aggregation, with a tiny group decided on the level of municipality, this must be considered very satisfactory.

Table 4 **Level of determination for different types of buildings. 2010**

Building types	zip-codes	Municipality	Province	Region	Denmark	Denmark*	Total
Single/double/triple family houses	1 118 945	1 049	0	0	0	0	**1 119 994**
Residential buildings (combined with) and business property	94 280	5 174	1 841	0	0	2 995	**103 991**
Farmland	84 172	452	161	0	0	0	**84 759**
Summer houses	201 307	2 596	640	0	1 029	0	**205 560**
Building sites	3 118	2 629	6 523	0	0	765	**7 770**
Public property, campsites, summer camps etc.	110	290	8 262	32	0	152	**1 021**
Condos (i.e. resident owned flats) and business owned flats	80 628	3 927	18 802	27 273	0	28 561	**158 423**
Factory owned flats and condos in family houses etc.	20 483	4 158	2 560	733	0	4 327	**31 780**
Private institutions and buildings on foreign land	887	463	1 733	0	0	546	**2 151**
Other property	0	0	0	0	0	3	**3**
Total	**1 603 930**	**20 738**	**40 522**	**28 038**	**1 029**	**37 349**	**1 715 452**

Weaknesses of the model

A disadvantage of the model is that a significant amount of the combined market value is not distributed on either land or building component. Two reasons are important explanations for this; 1) Intercept assumed to be zero, 2) A log-transformation of data fits the hedonic regression model better.

It has been assumed that the intercept should be zero. The reason for this assumption is an economic consideration; zero square meters of land and zero square meters of buildings must imply a combined value of zero. However, a test running of a hedonic regression model with an intercept suggest a model with intercept fits better with data. Even though this is the case we have decided to retain the requirement that the intercept should be zero for economic reasons.

By studying the combined market value it could be argued that a log-transformation of that variable could deliver more stable results. A test running of the hedonic regression model with a log-transformation of the dependent variable supports this argument.

If a log-transformation on the depended variable is done the interpretation of the parameters changes. It is possible to calculate a square meter price for buildings and land, but we need to find some specific land or building sizes from where the price can be calculated. This means that the data has to be investigating for each building type for dense intervals of square meters and use it as an indicator for that type of building. Melichar et al. has done some research on this issue and could be used a as guideline. However, this process will increase the workload considerably and could not be achieved within our time limits. Future research in this direction is an opportunity.

Another option which maybe could ensure better match between hedonic regression model and the data is to investigate the geographical breakdown and breakdown on types of buildings. It could be investigated if population density is a better indicator than our chosen geographical location and if the best grouping of building vary across location/population density. No research has been done relating to these issues by the authors.

Supplement no. 1: Parameters applied for observations outside the cleaned dataset

The subsample of records which was left outside the cleaned dataset, can deliver some valid information by applying the parameters from the hedonic regression model at suitable level of aggregation. This process make it is possible to split the combined market value into a building and a land component. The table below shows at which level of aggregation the records from the subsample are assigned with parameters for square meters of land and buildings.

Most of the records in the subsample are records without information on square meters of buildings or with zero square meters of buildings.

Table 5 **Level of determination for different types of buildings, records outside cleaned dataset. 2010**

Building types	zip-codes	Municipality	Province	Region	Denmark	Denmark*	Total
Single/double/triple family houses	2 679	16	0	0	0	0	**2 695**
Residential buildings (combined with) and business property	11 076	883	299	0	0	395	**12 653**
Farmland	23 809	198	26	0	0	0	**24 033**
Summer houses	7 827	1 483	12	0	3 201	0	**12 523**
Building sites	12 631	20 860	5 265	0	0	3 691	**42 447**
Public property, campsites, summer camps etc.	336	2 228	7 825	1 081	0	1 832	**13 302**
Condos (i.e. resident owned flats) and business owned flats	3 555	423	768	746	0	1 613	**7 105**
Factory owned flats and condos in family houses etc.	5 538	948	481	22	0	508	**7 497**
Private institutions and buildings on foreign land	497	2 596	1 478	0	0	4 641	**9 212**
Other property	0	0	0	0	0	916	**916**
Total	**67 948**	**29 635**	**16 154**	**1 849**	**3 201**	**13 596**	**132 383**

In total, the value of stock of dwellings is estimated to be 1.679,8 billion DKK in 2010, and less than 3 billion DKK is determined by applying parameters from the hedonic regression model for observations outside the cleaned dataset. This corresponds to 0.2 per cent of the total value of dwellings.

Supplement no. 2: Distribution of non-distributed combined market value

The aim of using the hedonic regression model is to distribute the combined market value between land and buildings. But if the hedonic regression model does not distribute all of the combined market value towards land or buildings, a method is needed for distribute the non-distributed. For practical purpose, a simple proportional distribution is applied. At most detailed level (zip-code and type of building), any residual is distributed towards buildings and land; distribution keys are the relative shares of land and buildings estimated by using the hedonic regression model. This approach secure total market value equals land plus buildings.

Reasons for the existence of the residual are addressed in the "Weaknesses of the model" section.

Measured in values, the size of the residual is uncomfortably large. In 2010 it amount to 16.5 per cent of the total value of buildings and about 10.5 per cent of total value of land.

The hedonic regression model – the results

The hedonic regression model suggests the core results for the value of buildings equal to 1.399 billion DKK in 2010, whereas the value of land is 1.488 billion DKK. In the following years, the value of buildings decline until 2012, and then the value of buildings increase in 2013, 2014 and 2015. In contrast, the value of land

increases in all the years except 2012 and 2013. The combined value decrease in the first 2 years (2011 and 2012), and increase in the last 3 years, most significantly in 2015 with 5.1 per cent.

Some may anticipate in advance it will be the land component that will be most sensitive to changes in the total values, but the model shows that it is the building component taking the largest adjustments (measured in per cent). For example, the combined market value rises by 2.4 percent in 2014, where the land component "only" rises by 1.4 percent while the building component increases by 3.5 percent. The fact that the building component is most sensitive to changes in the combined market value surprises the authors.

The table below shows the results from the hedonic regression model together with the two supplements, one for data outside the cleaned dataset and one for the proportional distribution.

Table 6 **Results from the hedonic regression model, including supplements**

	2010	2011	2012	2013	2014	2015
	Mill. DKK					
Buildings (Dwellings)	**1 679 796**	**1 569 205**	**1 520 638**	**1 607 601**	**1 656 070**	**1 772 588**
Hedonic regression model	1 398 983	1 316 484	1 269 293	1 344 470	1 392 097	1 480 629
Outside cleaned dataset	2 865	2 421	2 107	2 262	2 324	2 419
Proportional distribution	277 947	250 300	249 238	260 869	261 648	289 541
Land	**1 680 828**	**1 689 974**	**1 671 986**	**1 657 506**	**1 680 736**	**1 743 031**
Hedonic regression model	1 488 354	1 514 212	1 502 992	1 495 924	1 516 962	1 577 541
Outside cleaned dataset	15 188	13 166	11 100	10 234	10 119	10 825
Proportional distribution	177 286	162 596	157 895	151 349	153 655	154 666
Combined market value	**3 360 623**	**3 259 179**	**3 192 625**	**3 265 108**	**3 336 806**	**3 515 620**
Hedonic regression model	2 887 337	2 830 697	2 772 285	2 840 394	2 909 059	3 058 169
Outside cleaned dataset	18 053	15 587	13 207	12 496	12 443	13 244
Proportional distribution	455 233	412 895	407 133	412 218	415 304	444 206

As mentioned in the previous chapter, we add supplements for observations outside the cleaned dataset and for non-distributed combined market value (proportional distribution). These 2 additions do not change the overall result; the overall result is correlated with the result from the hedonic regression model.

Chapter 4: The Capital Stock compilation for dwellings

This chapter describes Statistics Denmark's compilation of capital stock estimates for dwellings. Statistics Denmark use a PIM like approach, which also incorporate information on number of square meters of buildings and price per square meter.

The values of the gross stock of dwellings (at current prices) are derived by using register information for the number of square meters matched with prices per square meters. The net stock is derived by applying assumptions on service lives and depreciation profile (geometric).

A supplement for costs of ownership transfer costs is added.

The numbers of square meters are available with a subdivision into 9 subcategories. The numbers square meters are available from administrative records and the stock can be measured at any desired point in time, Statistics Denmark assembles

the figures by the end of the year. Examples of the subcategories are farmhouses, detached houses, townhouses, flats and cottages. The subcategories used in the hedonic regression model are slightly different.

Each category of square meter is match with a relevant price. The prices come from a benchmark, which has been taking forward in time by using the deflator from GFCF of dwellings. In practice is GFCF at constant prices of dwellings derived by using construction cost indices (until 2015).

For dwellings constructed after the year 1950, the service life is 75 years and the declining balance rate is 1.45. Dwellings constructed prior to the year 1950, the service life is varying, but always higher than 75 years. For costs of ownership transfer costs the service life is 30 years and the declining balance rate is 1.85.

Price indices are used to transform the figures into previous year's average prices and inflate end of the year's figures into next year's prices.

The figures for Capital Stocks for dwellings are published with a break down on 2 industries; "Renting of residential buildings" and "owner-occupied buildings". A dwelling is deemed "owner-occupied buildings" if the dwelling is owned by a person (and thereby not a corporation) and the same person "occupy" the dwelling. Administrative records are used to compile the split on industries.

The gross and net stock is subsequently split on institutional sector. Distribution keys are applied based on information from the BBR-register. The distribution keys are shared with the estimation of output by the "owner occupied dwelling"-industry.

The distribution keys are applied by the level of industries. It is assumed that owner occupied buildings are 100 per cent owned by households, whereas buildings attributed towards renting of residential buildings can be owned by both non-financial corporations (S11) and households (S14). As mentioned, information from the Building and Real Estate Register (BBR-register) are used for deriving this split on institutional sector.

As the final part of compiling capital stocks for dwellings, a balancing procedure is done. A link (restriction) between the value of the stock at the beginning of the period and the end of the period, gross fixed capital formation and consumption of fixed capital is secured. The link apply for all variables measured at the same level of prices, in practice either average price of the year and previous year's average price. As a starting point the restriction does not apply, and a manual balancing is needed. The balancing is necessary because gross fixed capital formation for dwellings and the capital stock for dwellings are compiled by using different sources.

Table 7 **Net Capital Stock, Dwellings**

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
	Net Stock, end of year Billion DKK											
Non-financial corporations (S.11)	663	733	781	814	737	807	837	873	899	925	954	963
Households (S.14)	1 484	1 641	1 732	1 796	1 534	1 610	1 692	1 753	1 792	1 853	1 923	1 944
Total	**2 147**	**2 374**	**2 513**	**2 609**	**2 270**	**2 417**	**2 529**	**2 626**	**2 691**	**2 778**	**2 877**	**2 907**

The table above shows the Net Capital Stock of dwellings (end-of-the-year, current prices) with a break down on institutional sector.

Risk of questionable results with the Perpetual Inventory Method

In some rare cases, using the Perpetual Inventory Method (PIM) could produce questionable results. If the PIM model is applied, assumptions on the development in prices for existing buildings (assets) are needed, and normally prices for construction of new buildings are used as a proxy. If the price development for existing buildings vary from the price development for new buildings (in particular if the prices goes in opposite directions); the PIM estimate may underestimate or overestimate the "true" development in the capital stock of dwellings.

It may be difficult to identify periods with movements in opposite directions in prices for existing buildings and new construction because estimates for the price development of existing buildings are usually not available. One way questionable results (for dwellings) by the PIM model can be revealed is by estimating the value of land by using the residual approach, assuming reliable figures for the combined value of land and building is known. In this case, any inaccurate in the PIM estimate will impact the estimate for the value of land and in extreme cases result in negative values for land. In less extreme cases, the estimate for the value of land is just not reliable. For other types of fixed asset than buildings (dwellings) this method to identify with problematic results is not possible.

The handbook "Eurostat-OECD compilation guide on land estimation" has more on this issue; see page 81-82 and page 96.

Chapter 5: Comparison

This chapter compares the results of the estimations of the value of buildings for the period 2010-2015 by the two different approaches; A PIM-like estimation capital stocks and the hedonic approach based on a hedonic regression model.

Table 8 **Comparison of results, Dwellings, household sector (S.14)**

	2010	2011	2012	2013	2014	2015
	billion DKK					
Capital Stock estimation	1 610	1 692	1 753	1 792	1 853	1 923
Hedonic regression model*	1 680	1 569	1 521	1 608	1 656	1 773
	index 2010=100					
Capital Stock estimation	100	105.1	108.8	111.3	115.1	119.4
Hedonic regression model*	100	93.4	90.5	95.7	98.6	105.5
	percentage, annual real growth					
Capital Stock estimation		5.1	3.6	2.3	3.4	3.8
Hedonic regression model*		-6.6	-3.1	5.7	3.0	7.0

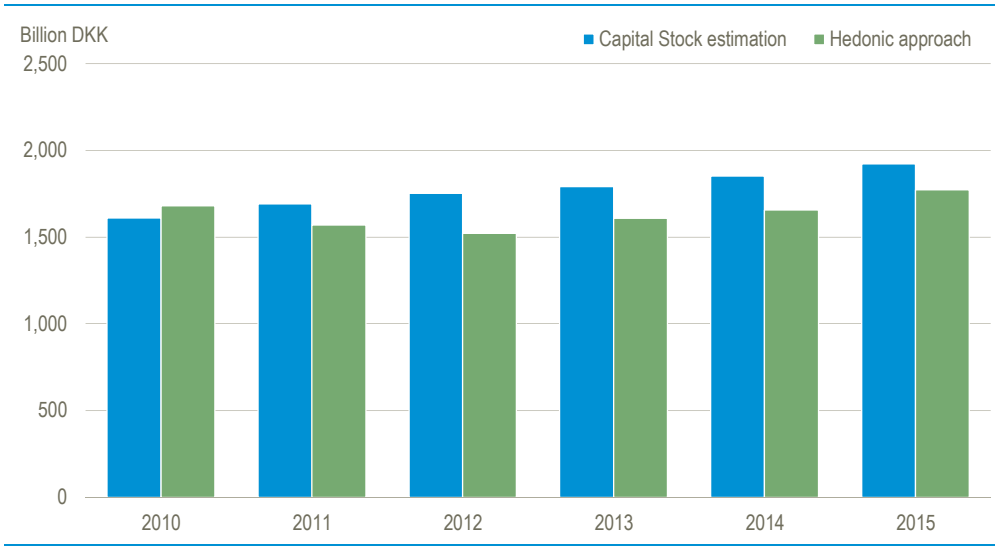
*Values for the hedonic regression model include supplements for and observations outside the cleaned dataset and for unallocated market values.

The methodologies behind these estimates are described in chapter 5 and chapter 4 respectively. Because the (cleaned) dataset used for hedonic regression model is limited to the household sector for dwellings, the results will be compared to corresponding part of the capital stock for dwellings. The value for the hedonic regression model includes supplements for unallocated market values and observations outside the cleaned dataset.

The nominal growth rates are very different for the 2 approaches for the years 2011 and 2012. The capital stock estimation shows increases for both of years, whereas the hedonic approach shows declines. In the next years (2013, 2014 and 2015) the growth rates are both positive for the 2 approaches, but on average the hedonic approach increases faster. Actually, the growth rates for the hedonic approach are more the a third higher than the growth rate for capital stock estimation.

In 2010, the 2 approaches are at almost similar level, with the hedonic regression model slightly larger than the capital stock estimates. This change in the subsequent years, as a result of the negative growth rates in 2011 and 2012 for the hedonic approach and positive growth rates for the capital stock estimates. The difference between the 2 approaches is largest in 2012 and decrease in the following years. In 2015 the result from the capital stock estimation is 7.8 per cent higher than the result from the hedonic regression model. The gap between the 2 approaches is largest for the years 2012, 2013 and 2014.

Figure 1 **Comparison: Capital Stocks vs. Hedonic approach**



Chapter 6: Conclusion

In some rare cases, the PIM model can produce questionable results. One dominant example is a period of time with divergent price trends between prices for construction of new buildings and prices of existing buildings. Normally, the prices for construction of new buildings are used as proxy for all prices in the PIM model for dwellings. In this case the PIM model can show strange results, PIM estimates can for example be too high if the prices of construction increase whereas prices for existing buildings decline. If the combined value of land and building can be derived with high degree of certainty, negative values for land (measured as a residual) could be the result of overestimation of the building component. This calls for an alternative to the classic PIM estimates for dwellings.

A hedonic regression model for estimating the value of buildings (dwellings) is an alternative. In this paper, estimations by the use of a hedonic regression model of the value of the net stock of dwellings for the household sector has been done with reasonable results. The differences between the results from the hedonic approach and the PIM model vary between 4.3 per cent (2010) and 13.2 per cent (2012). The directions of the year-to-year development are contradictory for the years 2011 and 2012 but consistent for the years 2013, 2014 and 2015. The hedonic regression model shows declining values for the years 2011 and 2012 whereas the PIM model shows increases for the years 2011 and 2012.

Using the hedonic approach is very time consuming. The input data can't be extracted from a single source in the Danish case, but has to be assembled from different sources and matched together with all the complications this may give. The (cleaned) dataset used for the hedonic regression model contains around 1.7 million records for each year, and dealing with a dataset at this size adds to the complexity. Together, assembling the large dataset and run around 7000 regressions for each year make the hedonic approach a very time consuming alternative compared to using the PIM model. However, it required because of the (lack of) quality of PIM estimates, the hedonic approach is a useful – but time consuming – alternative.

Some indications suggest that log-transformation of the data might fit the hedonic regression model better. However, log-transformation of data makes it difficult to interpret the parameters from the regression model as prices per (quality adjusted) square meter, and thus make it difficult to calculate the value of the stock of the dwellings which is the aim of the work. Or at least, the authors of this paper have not been successful in using parameters from log-transformed data to measuring prices per (quality adjusted) square meters; some kind of transformation of the parameters from the log-transformed data is required.

An alternative way forward is to calculate price indices for existing buildings by the use of hedonic regression model and use the price indices in the PIM model. The authors believe research in this direction could be useful. If time is allowed, the next step of work could be in this direction. A new price index for existing buildings may be another way to solve the rare cases with questionable results from the PIM model caused by misleading price index for new buildings.

Chapter 7: Bibliography

Diewert, W.E.: 'Alternative Approaches to Measuring House Price Inflation', Discussion Paper 10-10, Department of Economics, University of British Columbia, Vancouver, Canada, V6T 1Z1, December 2010.

Diewert, W.E., J. de Haan and R. Hendriks: 'The Decomposition of a House Price index into Land and Structures Components: A Hedonic Regression Approach', Discussion Paper 10-01, Department of Economics, University of British Columbia, Vancouver, Canada, V6T 1Z1, 2010.

Diewert, W. E. and Shimizu, C.: 'Residential Property Price Indexes for Tokyo', Discussion Paper 13-07, School of Economics, The University of British Columbia, Vancouver, Canada, V6T 1Z1, 2013.

Eurostat-OECD: 'Eurostat-OECD compilation guide on land estimation', Luxembourg, 2015.

Eurostat: 'Handbook on Residential Property Prices Indices (RPPIs)', Luxembourg, 2013.

Melichar, J., Vojáček, O., Rieger, P. and Jedlička, K.: 'Application on Hedonic Price Model in the Prague Property Market'.