Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

# Value estimation of the second home market
# CONFIDENTIAL

Note

Note no SAMBA/18/12

Authors Gunnhildur Högnadóttir Steinbakk
Magne Aldrin

Date May 21, 2012

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

## Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in information and communication technology and applied statistical-mathematical modelling. The clients include a broad range of industrial, commercial and public service organisations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have in us is given by the fact that most of our new contracts are signed with previous customers.

| | |
|---|---|
| **Title** | **Value estimation of the second home market** **CONFIDENTIAL** |

**Authors** **Gunnhildur Högnadóttir Steinbakk** <gunnhildur@nr.no>
**Magne Aldrin** <Magne.Aldrin@nr.no>

Date May 21, 2012

Publication number SAMBA/18/12

## Abstract

Eiendomsverdi AS has developed an automatic algorithm for estimating the market value of holiday residences in Norway. The current system of Eiendomsverdi provides up to two different basic value estimates for each property: One is based on computing an indexed estimate of previous sales prices, while the other estimate is a modelled value corrected for different characterstics of the property.

A value estimate of the price is computed as a weighted sum of the basic estimates, with weights depending different characteristics of each method. In addition, we develope a statistical model to quantify the uncertainty of the price estimates. This note describes the model for value estimates of second home properties and the uncertainty model for these estimated prices.

*Front page photo: ©2011 http://www.photos.com*



| | |
|---|---|
| Keywords | Residences, holiday home, market value, uncertainty |
| Target group | Eiendomsverdi AS |
| Availability | Confidential |
| Project | Forbedret verdiestimering av boliger - Fase 2 |
| Project number | 220410 |
| Research field | Finance, insurance and commodity markets |
| Number of pages | 22 |

# Contents

# 1 Introduction

Eiendomsverdi has developed automatic methods for estimating the value of second home properties in Norway. Their current system provides two different methods for estimating the value of a property in the second home market. One estimate is a modelled price corrected for different characteristics of the property. The other basic estimate is an indexed value of the previous sales price.

An estimate of the price value of a property is computed by a weighted sum of the two basic estimates, using a similar approach as applied in the Norwegian real estate market (Vårdal and Aldrin, 2008; Wilhelmsen and Aldrin, 2009). The weights depend on different characteristics of each method and on which of the basic estimate that exist for a particular property. For recently sold properties, the value estimate is in addition assumed to be more influenced by the indexed previous price estimate. Finally, we propose an uncertainty model for describing the value estimates uncertainty (Vårdal and Aldrin, 2008; Wilhelmsen and Aldrin, 2009).

Our approach for estimating the value of properties consist of two elements: One point estimate and an uncertainty distribution of the point estimate. The aim of this document is to describe and document the model for value estimation and its uncertainty. We have previously developed an uncertainty model for the estimates of the Norwegian second home market, based on the indexed value of previous sales prices only (Steinbakk and Aldrin, 2011). Here, the weighted value estimate also include the modelled basic estimate.

The outline of this report is as follows. Section 2 presents the data. The models for the value estimate and its uncertainty are given in Section 3, with general theory and mathematical details given in Appendix A. Section 4 shows the performance of the weighted value estimate and the uncertainty model.

# 2 Data

The data includes sales of 16 027 transactions in the second home market from 2nd of January 2003 to 12th of February 2012. The number of recorded sales in each geographical area are listed in Table 1, together with the number of each of the two basic price estimates (i.e., $P_1$ and $P_2$).

There are different variables recorded for each property, such as price and living area, which are used in the model for value estimation (see Section 3).

| | Area | No. of sales | No. of $P_1$ | No. of $P_2$ |
|---|---|---|---|---|
| 1 | Ø-Skog | 1886 | 382 | 1687 |
| 2 | Ø-Alpin | 2043 | 253 | 1928 |
| 3 | Ø-Langrenn | 3818 | 359 | 3682 |
| 4 | Oslofjord Øst | 1810 | 179 | 1738 |
| 5 | Oslofjord Vest | 1061 | 125 | 1016 |
| 6 | Skagerak | 942 | 123 | 886 |
| 7 | Agder | 2266 | 402 | 2047 |
| 8 | Bergen | 650 | 109 | 595 |
| 9 | Stavanger | 385 | 58 | 356 |
| 10 | Midt Norge | 971 | 173 | 862 |
| 11 | Nord Norge | 195 | 61 | 141 |
| | Total | 16027 | 2224 | 14938 |

Table 1. Number of recorded sales and basic estimates ($P_1$ and $P_2$) in each index area.

# 3 Method

The tools for estimating the value of properties consist of a point estimate and an uncertainty distribution of the point estimate. The basic estimates and the combined point estimate are described in Section 3.1 and 3.2, respectively, while the uncertainty model is presented in Section 3.3. Section 3.4 explains how the uncertainty distribution is grouped into different accuracy levels. The models (i.e. the value estimate and its uncertainty) are estimated from historical data (Section 2). Technical details of the general models in addition to the methods used for estimating the models are given Appendix A.

## 3.1 The basic estimates

Eiendomsverdi can compute two different basic estimates for a property, see Table 2. These basic estimates represents a logic, but different way of estimating sales prices, taking into account, amongst others, several characteristics of the property in question.

| Basic estimate | The basis for the estimates |
|---|---|
| $P_1$ | An indexed value of previous sales price of the target property |
| $P_2$ | A modelled value corrected for different characteristics of the target property |

Table 2. The two basic methods for value estimation of second home properties

The basic estimate $P_1$ exist for about 8% of the (historical) data in Section 2, while $P_2$ exist for more than 90% of the cases, see the two rightmost columns in Table 1. Even though if the model based estimate $P_2$ can be computed for a property, the $P_2$ will only exist if the following requirements are satisfied:

- The property has to be in a valid basic statistical unit[1].
- The property can not be sectioned.
- There has to be at least five second home properties within a radius of 100 meter
- The size of the property has to be larger than $30m^2$ (living area).
- The property has to be established 1930 or later.

## 3.2 The value estimate

There are different ways of combining the basic estimates in Table 2. One simple approach is, for example, to compute the mean value. Here, we use an approach called the *weighted model* (see Appendix A for details). Basically, the basic estimates are weighted together with weights depending on characteristics of the property in question. The idea is to find the combinations of weights so that the value estimate is closest to its actual sales price. In addition, we adjust the weighted estimate by giving the indexed estimate $P_1$ more weights for recently sold properties. The value estimate is assumed to be equal $P_1$ if less than one year since last sale.

Mathematically, the weighted estimate $\hat{P}_\omega$ can be written as

$$\hat{P}_\omega = \omega_1 P_1 + \omega_2 P_2 \tag{1}$$

where $P_1$ and $P_2$ are the two basic estimates, and $\omega_1$ and $\omega_2$ are the corresponding weights. The weights $\omega_1$ and $\omega_2$ will not be equal for all properties, but rather vary according to their specific characteristics. There will be one weight belonging to each of the two basic estimates, where the weights are zero if the basic estimate do not exist for a particular property.

The model for non-scaled weight for the Norwegian second home market is

$$\begin{aligned} v_i = I_i \cdot \exp\Big\{ & \beta_1 \texttt{time} + \beta_2 \texttt{subCouncilModelDeviation} \\ & + \beta_3 I_1 + \beta_4 I_2 + \beta_5 I_1 I_2 \Big\} \end{aligned} \tag{2}$$

for $i = 1, 2$, where `time` is the time since last sale, `subCouncilModelDeviation` is a standard deviation of (historical) differences between observed prices and their corresponding estimate $P_2$ in each basic statistical unit, and $I_i$ is an indicator for basic estimate $P_i$.

The weights are scaled to summarise to one by

$$\omega_1 = \frac{v_1}{v_1 + v_2} \quad \text{and} \quad \omega_2 = \frac{v_2}{v_1 + v_2}. \tag{3}$$

Hence, the weights are exactly one if only one of the weights exist.

1. The validity of a basic statistical unit is defined by Eiendomsverdi based on an analysis of differences between the $P_2$ and the actual price of historical data. Please contact Eiendomsverdi for the list of valid basic statistical units in the second home market of Norway.

Computing the weighted estimate is thus done in two steps. First, the weights $\omega_1$ and $\omega_2$ are computed based on the formulae for weights in (2). Then, the weighted value estimate is computed based on these weights and the basic estimates by using the formulae for $P_\omega$.

For recently sold properties, we want to put more weight on the indexed estimate $P_1$. Let $\alpha$ be a function of time since last sale by

$$\alpha = \begin{cases} 1 & \text{if} \quad \texttt{time} \geq 3 \\ \frac{\texttt{time}-1}{2} & \text{if} \quad 1 \leq \texttt{time} < 3 \\ 0 & \text{if} \quad \texttt{time} < 1, \end{cases} \tag{4}$$

where `time` is the number of days since last sale divided by 365. Furthermore, denote the indexed previous price estimate $P_1$ on logarithmic scale by $\hat{\mathrm{P}}_{\mathrm{log,pp}}$. The final value estimate of second home properties will be a combination the weighted estimate $\hat{P}_\omega$ and the index previous price estimate $\hat{\mathrm{P}}_{\mathrm{log,pp}}$ on logarithmic scale (see Wilhelmsen and Aldrin (2009)), i.e.,

$$\hat{P}_{\mathrm{log}} = \log(\hat{P}) = \alpha\hat{\mathrm{P}}_{\mathrm{log},\omega} + (1-\alpha)\hat{\mathrm{P}}_{\mathrm{log,pp}}. \tag{5}$$

If there are less than a year since previous sale, $\alpha = 0$ and we only use the indexed previous sales price $\hat{\mathrm{P}}_{\mathrm{log,pp}}$. If there are in between one and three years since previous sale, $\alpha$ is somewhere in between 0 and 1, depending on what `time` is, and we use a combination of the two estimates. If more than three years since last sale, $\alpha$ is 1 and we only use the weighted estimate $\hat{P}_\omega$. The latter includes situations when only $P_2$ exist (i.e., $P_1$ is missing).

### 3.3 The uncertainty of the value estimate

In this section we present the final uncertainty model for the second home market in Norway. Details of the general model and the approach for fitting the model to (historical) data are given in Appendix A.2.

Our aim is to find the distribution of the error which defines the uncertainty distribution, described by its, mean, its variance, and its shape. The error term is defined as the difference between the actual price and the estimate $\hat{P}_\omega$ on logarithmic scale. We estimate the variance (i.e. $\sigma^2$) of the error term, while the empirical distribution of the error terms defines the shape of the uncertainty distribution.

The model for the variance for the weighted estimate $\hat{\mathrm{P}}_{\mathrm{log},\omega}$ is given by

$$\begin{aligned} \hat{\sigma}^2_\omega = \exp\Big\{ & \gamma_0 + \sum_{i=2}^{i=11} \gamma_i I_{\texttt{region}_i} + \gamma_{12} \cdot \hat{\mathrm{P}}_{\mathrm{log},\omega} + \gamma_{13} \cdot \texttt{rankedArea} \\ & + \gamma_{14} \cdot \texttt{buildYear} + \gamma_{15} \cdot \omega_1 \texttt{Time} \\ & + \gamma_{16} \cdot \omega_2 \texttt{logAltitude1} + \gamma_{17} \cdot \omega_2 \texttt{logAltitude2} \\ & + \gamma_{18} \cdot \omega_2 \texttt{SubCouncilModelDeviation} \\ & + \gamma_{19}\omega_1 + \gamma_{20} I_2 + \gamma_{21} I_1 I_2 \Big\}. \end{aligned} \tag{6}$$

Here, $I_i$ is the indicator for the basic estimate $P_i$, which is one if $P_i$ exist and zero otherwise. Similarly, the area specific variable $I_{\texttt{region}_i}$ is one if the property is in area $i$ and zero otherwise. The areas are numbered according to Table 1. The intercept $\gamma_0$, and the terms $\sum_{i=2}^{i=11} \gamma_i I_{\texttt{region}_i}$, $\gamma_{18} I_2$, and $\gamma_{19} I_1 I_2$ are constant terms for different combinations of methods and regions. Note that $\gamma_1$ is missing in the sum, where $\gamma_1 = 0$ corresponds to Ø-Skog. The variable $\hat{\mathrm{P}}_{\log,\omega}$ is the weighted estimate on logarithmic scale, and $\omega_1$`Time` is time since last sale multiplied by the weight $\omega_1$. The $\omega_2$`LogAltitude` is altitude on logarithmic scale multiplied by the weight $\omega_2$, where $\omega_2$`logAltitude1` includes lower heights and $\omega_2$`logAltitude2` includes higher heights. Moreover, `rankedArea` is living area imputed for missing values (i.e. by utility floor space, gross area, or the average of historical living areas), and `buildYear` is the year of building. Finally, $\omega_2$`SubCouncilModelDeviation` is a historical standard deviation for $P_2$, multiplied by the weight $\omega_2$.

Similarly to the model for the weighted estimate $\hat{P}_\omega$, we define an uncertainty model for the indexed previous price estimate $\hat{\mathrm{P}}_{\log,\mathrm{pp}}$ by

$$\sigma_{\mathrm{PP}}^2 = \exp\left\{\gamma_0 + \gamma_1 \cdot \texttt{buildYear} + \gamma_2 \cdot \texttt{time}\right\}. \tag{7}$$

The model for the variance in the Norwegian second home market is then computed as a combination of the variances $\hat{\sigma}_{\mathrm{PP}}$ and $\hat{\sigma}_\omega$ ,

$$\hat{\sigma}^2 = \alpha^2 \hat{\sigma}_\omega^2 + (1-\alpha)^2 \hat{\sigma}_{\mathrm{PP}}^2 + 2\alpha(1-\alpha)\hat{\sigma}_\omega \hat{\sigma}_{\mathrm{PP}} \rho. \tag{8}$$

where $\rho$ is the empirical correlation between the error terms $\epsilon_\omega = (\log P - \hat{\mathrm{P}}_{\log,\omega})$ and $\epsilon_{\mathrm{PP}} = (\log P - \hat{\mathrm{P}}_{\log,\mathrm{pp}})$.

The final distribution of the prices is

$$\mathbf{P} = \exp\left\{\hat{P}_{\log} + \boldsymbol{\epsilon^*}_{\alpha=1}\hat{\sigma}\right\}, \tag{9}$$

if $\alpha = 1$, and

$$\mathbf{P} = \exp\left\{\hat{P}_{\log} + \boldsymbol{\epsilon}^*_{\alpha<1}\hat{\sigma}\right\}, \tag{10}$$

if $\alpha < 1$. The latter one is the distribution for recently sold properties (less than three years since last sale), while the first one is the distribution for properties with more than three years since last sale (inclusive those properties where only the basic estimate $P_2$ exist).

## 3.4 Accuracy levels

The sales are classified into accuracy levels 1 to 7. These uncertainty levels depend on how much the estimated sales price need to be adjusted in order to reach the 25% quantile of the sales price distribution. The accuracy level 1 consist of properties with lower downward adjustment than 4 %. The limits for downward adjustments for the other levels 2-7 are 6%, 8%, 10%, 14%, 20%, and 100%, respectively.

| | |
|---|---|
| Prop. | Proportion of sales in a category compared to the total sales |
| Est.value | The sum of the estimated prices in a category (in MNOK) |
| Sales price | The sum of the actual sales prices in a category (in MNOK) |
| Dev. | The relative difference between Est.value and Sales price |
| | Negative: the estimated price is less than the sales price |
| | Positive: the estimated price is higher than the sales price |

Table 3. Overview of some of the quantities in the tables used for validating the model performance of $\hat{P}_\omega$ in Section 4.1.

# 4 Results

The aim of this section is to quantify and illustrate the performance of the value estimate and the uncertainty model in the holiday house market.

## 4.1 Model performance

This section describes to which degree Eiendomsverdi's estimated market values fit the real sales price. The estimates are categorised according to the deviation in per cent from the real sales price using 13 categories. The first interval, "< -30 %", contains estimates which are underestimated and deviates more than 30% from the sales price. That is, reducing the sales price by 30%, the estimates in interval "< -30%" are less than this value. The next interval contains all estimates less than the sales price which deviates 20-30% from the sales price, and so on. The interval "-10, 10" contains all estimates which deviates less than 10% from the sales price in both positive and negative direction. Categories with negative values underestimate the sales prices, while the positive ones overestimate the prices. Table 3 explains some other quantities used in the summary tables in this section.

The error intervals vary between the counties as is expected, see Table 6. The estimated prices in Ø-Langrenn fit the actual prices best (31% have less than 10% error), while only 24% of estimated prices in Oslofjord Øst is within in 10% error interval.

The deviation according to market value bracket (Table 5) shows that the 31% of sales between 0.5 and 1 MNOK and between 1 and 2 MNOK deviate less than 10% from the actual sales price. The overall picture is that the most expensive properties are underestimated while the cheaper properties are overestimated.

The percentages for estimates that overestimate more than 30% are sometimes relatively large (see, for example, the column ">30" in Table 6). Table 4 shows, however, that the percentage based on the number of properties that overestimate properties is larger than if measured in amount of kroner. Furthermore, the percentage that underestimate the properties is smaller measured in countt than based on kroner. We also see that the total portfolio of estimates that underestimate more than -30% is more in kroner (i.e., 19%) than the total portfolio of estimates that overestimate more than 30% (i.e., 8%). This means that the estimates that underestimate affect the difference in the total portfolio more than the estimates that overestimate the prices, so that the total portfolio in sum is underestimated by 7.3%.

| | <-30 | [-30,-20] | [-20,-10] | [-10,10] | [10,20] | [20,30] | >30 | Total |
|---|---|---|---|---|---|---|---|---|
| Based on number | 13 % | 12 % | 16 % | 29 % | 9 % | 7 % | 13 % | 100 % |
| Based on kroner | 19 % | 14 % | 16 % | 28 % | 9 % | 5 % | 8 % | 100 % |
| Est. value (MNOK) | 4 970 | 2 147 | 2 949 | 7 569 | 3 282 | 2 155 | 2 170 | 25 243 |
| Sales price (MNOK) | 7 883 | 2 678 | 3 381 | 7 556 | 2 802 | 1 629 | 1 295 | 27 225 |
| Dev. (%) | -36.9 | -19.8 | -12.8 | 0.2 | 17.1 | 32.3 | 67.6 | -7.3 |

Table 4. Percentage of estimates falling into different error interval based on counts (upper row) and kroner (second row). The row names in the last three rows are explained in Table 3.

## 4.2 Quantils results

Here, we check if the uncertainty distribution captures the observed uncertainty in different subsets of the data set. Table 7 shows the percentage of observed sales prices that are below different quantiles in the estimated price distributions for model (9), while Table 8 shows the results for the model for recently sold properties in (10). We have considered different subsets of data; those with previous sale less than two years ago and more than two years ago, and the properties estimated to be less uncertain (named as smallSigma) and those being more uncertain. Finally, we consider each area for the indexed estimate.

For each subset, we check if the percentage of the observed sales prices that are below the quantiles are close to (10%, 25%, 50%, 75%, 90%). The column $q_P$ shows the percentage of the observed sales prices that are below the point estimates. The column 80%CI shows the percentage of the observed sales prices that are within the 80% confidence interval. The rightmost column of the table shows the number of observations in each subset.

## 4.3 Accuracy classes

The uncertainty of each sale is classified into seven different uncertainty classes, depending on how much adjustment is needed in order to reach the 25% quantile of the sales price distribution (see Vårdal and Aldrin (2008)). Most of the sales end up in the higher accuracy classes 5 and 6, see Table 9. Sales where the previous sales price exist (i.e. basic estimate $P_1$) are, however classfied into lower accuracy levels, see Table 10.

## 4.4 The mean deviations

We will in this section show the mean deviations between different price estimates and their associated real sales prices, illustrated on the weighted estimate $\hat{P}_\omega$. If we let $P_i^{\text{sale}}$ be the real sales price and $\hat{P}_i$ be the associated price estimate, we define the mean deviation `MD` to be

$$\texttt{MD} = \frac{1}{N}\sum_{i=1}^{N}\frac{|P_i^{\text{sale}} - \hat{P}_i|}{P_i^{\text{sale}}}, \tag{11}$$

where $N$ is the total number of sales.

Table 11 shows that the mean deviation for $\hat{P}_\omega$ is 25% for all areas. The mean deviation is 18% for $P_1$ and 16% for $\hat{P}_\omega$ in cases when both $P_1$ and $P_2$ exist.

| Price (MNOK) | <-30 | [-30,-20] | [-20,-10] | [-10,10] | [10,20] | [20,30] | >30 | Total sales | Prop. (%) | Est.value (MNOK) | Sales price (MNOK) | Dev. (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| <0.5 | 1 % | 3 % | 8 % | 25 % | 11 % | 12 % | 40 % | 1 258 | 8 | 618 | 485 | 27.5 |
| 0.5-1 | 6 % | 9 % | 14 % | 31 % | 12 % | 10 % | 18 % | 4 298 | 27 | 3 427 | 3 184 | 7.6 |
| 1-2 | 14 % | 14 % | 18 % | 31 % | 8 % | 6 % | 9 % | 5 677 | 35 | 7 809 | 8 239 | -5.2 |
| 2-3 | 17 % | 16 % | 18 % | 29 % | 9 % | 4 % | 6 % | 2 898 | 18 | 6 440 | 7 033 | -8.4 |
| 3-4 | 21 % | 15 % | 17 % | 29 % | 8 % | 5 % | 6 % | 1 080 | 7 | 3 265 | 3 635 | -10.2 |
| >4 | 36 % | 17 % | 13 % | 21 % | 7 % | 3 % | 3 % | 816 | 5 | 3 684 | 4 649 | -20.8 |
| Total | 13 % | 12 % | 16 % | 29 % | 9 % | 7 % | 13 % | 16 027 | 100 | 25 243 | 27 225 | -7.3 |

Table 5. *Deviation per market value bracket*: Percentages of estimates from different market value bracket (i.e. Price) falling into the different error intervals. The column names in last four columns are explained in Table 3.

| Area | <-30 | [-30,-20] | [-20,-10] | [-10,10] | [10,20] | [20,30] | >30 | Total sales | Prop. (%) | Est.value (MNOK) | Sales price (MNOK) | Dev. (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ø-Langrenn | 9 % | 12 % | 18 % | 33 % | 10 % | 7 % | 11 % | 3 818 | 24 | 5 667 | 5 949 | -4.7 |
| Agder | 13 % | 13 % | 16 % | 29 % | 10 % | 6 % | 12 % | 2 266 | 14 | 3 146 | 3 357 | -6.3 |
| Ø-Alpin | 11 % | 13 % | 17 % | 32 % | 9 % | 7 % | 12 % | 2 043 | 13 | 4 199 | 4 464 | -5.9 |
| Ø-Skog | 13 % | 10 % | 14 % | 29 % | 10 % | 8 % | 16 % | 1 886 | 12 | 1 478 | 1 576 | -6.2 |
| Oslofjord Øst | 17 % | 13 % | 16 % | 24 % | 9 % | 7 % | 14 % | 1 810 | 11 | 3 272 | 3 676 | -11.0 |
| Oslofjord Vest | 19 % | 13 % | 11 % | 25 % | 8 % | 6 % | 18 % | 1 061 | 7 | 2 237 | 2 503 | -10.6 |
| Midt Norge | 14 % | 15 % | 17 % | 31 % | 8 % | 6 % | 8 % | 971 | 6 | 1 118 | 1 256 | -11.0 |
| Skagerak | 17 % | 10 % | 13 % | 27 % | 9 % | 7 % | 17 % | 942 | 6 | 2 292 | 2 499 | -8.3 |
| Bergen | 12 % | 13 % | 16 % | 31 % | 9 % | 5 % | 13 % | 650 | 4 | 1 007 | 1 058 | -4.8 |
| Stavanger | 17 % | 11 % | 14 % | 26 % | 10 % | 6 % | 16 % | 385 | 2 | 657 | 713 | -7.9 |
| Nord Norge | 11 % | 9 % | 19 % | 26 % | 12 % | 7 % | 15 % | 195 | 1 | 169 | 173 | -2.5 |
| All areas | 13 % | 12 % | 16 % | 29 % | 9 % | 7 % | 13 % | 16 027 | 100 | 25 243 | 27 225 | -7.3 |

Table 6. *Deviation per county*: Percentages of estimates from different counties falling into the different error intervals. The column names in last four columns are explained in Table 3.

| | 10% | 25% | 50% | 75% | 90% | 80%CI | $q_P$ | N |
|---|---|---|---|---|---|---|---|---|
| All data set | 0.100 | 0.250 | 0.500 | 0.750 | 0.900 | 0.800 | 0.416 | 15069 |
| smallSigma | 0.092 | 0.245 | 0.495 | 0.748 | 0.905 | 0.813 | 0.418 | 3767 |
| mediumSigma | 0.095 | 0.244 | 0.493 | 0.743 | 0.896 | 0.801 | 0.404 | 7534 |
| largeSigma | 0.119 | 0.268 | 0.520 | 0.766 | 0.904 | 0.785 | 0.439 | 3768 |
| Ø-Skog | 0.124 | 0.302 | 0.569 | 0.789 | 0.917 | 0.793 | 0.465 | 1728 |
| Ø-Alpin | 0.097 | 0.245 | 0.484 | 0.741 | 0.903 | 0.805 | 0.408 | 1943 |
| Ø-Langrenn | 0.097 | 0.249 | 0.506 | 0.755 | 0.909 | 0.812 | 0.420 | 3654 |
| Oslofjord Øst | 0.087 | 0.234 | 0.485 | 0.740 | 0.884 | 0.797 | 0.411 | 1728 |
| Oslofjord Vest | 0.111 | 0.251 | 0.496 | 0.737 | 0.901 | 0.789 | 0.417 | 1016 |
| Skagerak | 0.118 | 0.273 | 0.526 | 0.753 | 0.887 | 0.769 | 0.455 | 896 |
| Agder | 0.094 | 0.243 | 0.479 | 0.742 | 0.892 | 0.799 | 0.398 | 2083 |
| Bergen | 0.088 | 0.217 | 0.490 | 0.797 | 0.920 | 0.832 | 0.385 | 600 |
| Stavanger | 0.112 | 0.246 | 0.513 | 0.739 | 0.905 | 0.793 | 0.437 | 357 |
| MidtNorge | 0.078 | 0.204 | 0.432 | 0.693 | 0.869 | 0.791 | 0.349 | 899 |
| NordNorge | 0.127 | 0.285 | 0.521 | 0.764 | 0.897 | 0.770 | 0.430 | 165 |

Table 7. The percentage of observed sales prices that are below different quantile in the estimated price distributions for different subsets of the data.

| | 10% | 25% | 50% | 75% | 90% | 80%CI | $q_P$ | N |
|---|---|---|---|---|---|---|---|---|
| All recently sold | 0.100 | 0.249 | 0.500 | 0.749 | 0.900 | 0.800 | 0.505 | 958 |
| < 2 years | 0.077 | 0.233 | 0.483 | 0.726 | 0.882 | 0.805 | 0.491 | 532 |
| 2-3 years | 0.129 | 0.270 | 0.521 | 0.779 | 0.923 | 0.793 | 0.523 | 426 |
| smallSigma | 0.083 | 0.229 | 0.525 | 0.758 | 0.896 | 0.812 | 0.533 | 240 |
| mediumSigma | 0.098 | 0.262 | 0.513 | 0.745 | 0.904 | 0.805 | 0.517 | 478 |
| largeSigma | 0.121 | 0.246 | 0.450 | 0.750 | 0.896 | 0.775 | 0.454 | 240 |

Table 8. The percentage of observed sales prices that are below different quantile in the estimated price distributions for different subsets of the data for recently sold properties.

| Area | $P_\omega$ | No. of sales | $P_1$ when both $P_1$ and $P_2$ exist | $P_\omega$ when both $P_1$ and $P_2$ exist | No. of sales |
|---|---|---|---|---|---|
| Ø-Skog | 26 % | 1 886 | 20 % | 16 % | 183 |
| Ø-Alpin | 23 % | 2 043 | 16 % | 15 % | 138 |
| Ø-Langrenn | 21 % | 3 818 | 16 % | 14 % | 223 |
| Oslofjord Øst | 30 % | 1 810 | 20 % | 17 % | 107 |
| Oslofjord Vest | 30 % | 1 061 | 24 % | 22 % | 80 |
| Skagerak | 29 % | 942 | 18 % | 15 % | 67 |
| Agder | 25 % | 2 266 | 20 % | 18 % | 183 |
| Bergen | 24 % | 650 | 20 % | 18 % | 54 |
| Stavanger | 28 % | 385 | 23 % | 16 % | 29 |
| Midt Norge | 25 % | 971 | 13 % | 13 % | 64 |
| Nord Norge | 23 % | 195 | 11 % | 11 % | 7 |
| All areas | 25 % | 16 027 | 18 % | 16 % | 1 135 |

Table 11. The mean deviations of the errors for different areas for all sales and in cases when both $P_1$ and $P_2$ exist.

| Levels | Max down scaling | Portion |
|---|---|---|
| 1 | 4% | 0.0 |
| 2 | 6% | 0.0 |
| 3 | 8% | 1.8 |
| 4 | 10% | 7.5 |
| 5 | 14% | 61.7 |
| 6 | 20% | 28.7 |
| 7 | | 0.3 |

Table 9. Portion of properties (%) in each accuracy class

| Levels | Max down scaling | Portion |
|---|---|---|
| 1 | 4% | 0.0 |
| 2 | 6% | 5.3 |
| 3 | 8% | 26.6 |
| 4 | 10% | 27.5 |
| 5 | 14% | 35.9 |
| 6 | 20% | 4.7 |
| 7 | | 0.0 |

Table 10. Portion of properties (%) in each accuracy class for all sales where indexed previous sale estimate $P_1$ exist.

## 4.5 Weights according to time since last sale

Table 12 shows how the weights $\omega_1$ and $\omega_2$ varies by the time since last sale from one to seven years, for fixed values of `SubCouncilModelDeviation` when both $P_1$ and $P_2$ exist.

| Time since last sale | $\omega_1$ | $\omega_2$ |
|---|---|---|
| 1 | 0.821 | 0.179 |
| 3 | 0.725 | 0.275 |
| 5 | 0.603 | 0.397 |
| 7 | 0.466 | 0.534 |

Table 12. Weights for different length of time (in years) since last sale.

## 4.6 Haircut scenario

In order to minimise the probability of overestimating individual properties, Eiendomsverdi applies haircuts to the value estimates. Here, we show the model performance when all estimates are downgraded to the 30%, 25%, and 20% percentile of the value estimate distribution (see the tables 13-15).

| Area | <-30 | [-30,-20] | [-20,-10] | [-10,10] | [10,20] | [20,30] | >30 | Total sales | Prop. (%) | Est.value (MNOK) | Sales price (MNOK) | Dev. (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ø-Langrenn | 16 % | 17 % | 19 % | 29 % | 8 % | 4 % | 6 % | 3 818 | 24 | 5 240 | 5 949 | -12 |
| Agder | 23 % | 15 % | 18 % | 25 % | 6 % | 4 % | 7 % | 2 266 | 14 | 2 879 | 3 357 | -14 |
| Ø-Alpin | 19 % | 17 % | 20 % | 26 % | 7 % | 4 % | 7 % | 2 043 | 13 | 3 869 | 4 464 | -13 |
| Ø-Skog | 20 % | 14 % | 16 % | 27 % | 8 % | 4 % | 10 % | 1 886 | 12 | 1 344 | 1 576 | -15 |
| Oslofjord Øst | 27 % | 16 % | 14 % | 23 % | 7 % | 4 % | 9 % | 1 810 | 11 | 2 944 | 3 676 | -20 |
| Oslofjord Vest | 30 % | 13 % | 16 % | 20 % | 6 % | 4 % | 12 % | 1 061 | 7 | 2 006 | 2 503 | -20 |
| Midt Norge | 23 % | 18 % | 18 % | 26 % | 6 % | 2 % | 6 % | 971 | 6 | 1 029 | 1 256 | -18 |
| Skagerak | 25 % | 14 % | 15 % | 24 % | 7 % | 5 % | 10 % | 942 | 6 | 2 069 | 2 499 | -17 |
| Bergen | 20 % | 17 % | 21 % | 23 % | 6 % | 5 % | 7 % | 650 | 4 | 916 | 1 058 | -14 |
| Stavanger | 26 % | 13 % | 16 % | 24 % | 8 % | 4 % | 10 % | 385 | 2 | 594 | 713 | -17 |
| Nord Norge | 17 % | 16 % | 15 % | 29 % | 8 % | 6 % | 8 % | 195 | 1 | 155 | 173 | -10 |
| All areas | 21 % | 16 % | 18 % | 26 % | 7 % | 4 % | 8 % | 16 027 | 100 | 23 044 | 27 225 | -15 |

Table 13. *Deviation per county for haircut scenario*: Percentages of estimates from different counties falling into the different error intervals, after haircuts 30%.

| Area | <-30 | [-30,-20] | [-20,-10] | [-10,10] | [10,20] | [20,30] | >30 | Total sales | Prop. (%) | Est.value (MNOK) | Sales price (MNOK) | Dev. (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ø-Langrenn | 20 % | 20 % | 19 % | 26 % | 7 % | 3 % | 5 % | 3 818 | 24 | 5 049 | 5 949 | -15 |
| Agder | 27 % | 18 % | 17 % | 23 % | 6 % | 4 % | 5 % | 2 266 | 14 | 2 762 | 3 357 | -18 |
| Ø-Alpin | 24 % | 19 % | 19 % | 24 % | 6 % | 3 % | 6 % | 2 043 | 13 | 3 722 | 4 464 | -17 |
| Ø-Skog | 24 % | 16 % | 17 % | 25 % | 6 % | 3 % | 8 % | 1 886 | 12 | 1 286 | 1 576 | -18 |
| Oslofjord Øst | 32 % | 17 % | 15 % | 20 % | 6 % | 3 % | 7 % | 1 810 | 11 | 2 800 | 3 676 | -24 |
| Oslofjord Vest | 35 % | 14 % | 15 % | 18 % | 6 % | 4 % | 9 % | 1 061 | 7 | 1 905 | 2 503 | -24 |
| Midt Norge | 28 % | 19 % | 19 % | 23 % | 5 % | 3 % | 4 % | 971 | 6 | 990 | 1 256 | -21 |
| Skagerak | 29 % | 15 % | 15 % | 23 % | 6 % | 6 % | 7 % | 942 | 6 | 1 972 | 2 499 | -21 |
| Bergen | 27 % | 19 % | 18 % | 21 % | 5 % | 3 % | 6 % | 650 | 4 | 875 | 1 058 | -17 |
| Stavanger | 29 % | 16 % | 14 % | 23 % | 7 % | 3 % | 8 % | 385 | 2 | 567 | 713 | -21 |
| Nord Norge | 21 % | 18 % | 12 % | 31 % | 6 % | 7 % | 5 % | 195 | 1 | 149 | 173 | -14 |
| All areas | 26 % | 18 % | 17 % | 24 % | 6 % | 3 % | 6 % | 16 027 | 100 | 22 077 | 27 225 | -19 |

Table 14. *Deviation per county for haircut scenario*: Percentages of estimates from different counties falling into the different error intervals, after haircuts 25%.

| Area | <-30 | [-30,-20] | [-20,-10] | [-10,10] | [10,20] | [20,30] | >30 | Total sales | Prop. (%) | Est.value (MNOK) | Sales price (MNOK) | Dev. (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ø-Langrenn | 26 % | 20 % | 19 % | 23 % | 5 % | 3 % | 4 % | 3 818 | 24 | 4 844 | 5 949 | -19 |
| Agder | 33 % | 19 % | 17 % | 20 % | 4 % | 3 % | 4 % | 2 266 | 14 | 2 637 | 3 357 | -22 |
| Ø-Alpin | 29 % | 21 % | 18 % | 21 % | 4 % | 3 % | 4 % | 2 043 | 13 | 3 564 | 4 464 | -20 |
| Ø-Skog | 29 % | 17 % | 17 % | 22 % | 5 % | 4 % | 6 % | 1 886 | 12 | 1 224 | 1 576 | -22 |
| Oslofjord Øst | 39 % | 17 % | 14 % | 18 % | 3 % | 3 % | 5 % | 1 810 | 11 | 2 649 | 3 676 | -28 |
| Oslofjord Vest | 41 % | 16 % | 13 % | 16 % | 5 % | 4 % | 6 % | 1 061 | 7 | 1 799 | 2 503 | -28 |
| Midt Norge | 34 % | 20 % | 18 % | 20 % | 2 % | 2 % | 3 % | 971 | 6 | 948 | 1 256 | -24 |
| Skagerak | 36 % | 15 % | 15 % | 20 % | 6 % | 3 % | 5 % | 942 | 6 | 1 869 | 2 499 | -25 |
| Bergen | 33 % | 23 % | 14 % | 19 % | 4 % | 2 % | 5 % | 650 | 4 | 833 | 1 058 | -21 |
| Stavanger | 34 % | 18 % | 14 % | 21 % | 5 % | 3 % | 6 % | 385 | 2 | 538 | 713 | -25 |
| Nord Norge | 25 % | 21 % | 15 % | 25 % | 7 % | 4 % | 3 % | 195 | 1 | 143 | 173 | -18 |
| All areas | 32 % | 19 % | 17 % | 21 % | 4 % | 3 % | 4 % | 16 027 | 100 | 21 047 | 27 225 | -23 |

Table 15. *Deviation per county for haircut scenario*: Percentages of estimates from different counties falling into the different error intervals, after haircuts 20%.

# References


Steinbakk, G. H. and Aldrin, M. (2011). Uncertainty model for the value estimate of the sectond home market. Technical Report SAMBA/20/11, Norwegian Computing Center.

Vårdal, J. and Aldrin, M. (2008). Forbedret verdiestimering av boliger. Technical Report SAMBA/45/08, Norwegian Computing Center.

Wilhelmsen, M. and Aldrin, M. (2009). Some changes of the value estimation – Volume 3. Technical Report SAMBA/41/09, Norwegian Computing Center.

# A Methodological details

The approach for estimating the value estimate consist of (1) a point estimate and (2) an uncertainty distribution of the point estimate. Here, we present general models and methods for the weighted value estimate and its uncertainty. The definitions of variables specific to the Norwegian second home market including the actual formulae for weights and uncertainty, is given in Section 3.

The actual sales price of a property and the corresponding estimate are denoted by $P_{\text{act}}$ and $P_\omega$, respectively. Following Vårdal and Aldrin (2008), the error of the value estimate is defined as the difference between the actual sales price and the estimate on logarithmic scale,

$$\epsilon = \log P_{\text{act}} - \log P_\omega. \tag{A.1}$$

Technically, all analyses are done on logarithmic scale and then properly transformed to originally scale. The error on logarithmic scale is, amongst others, better describing that a 1 million kroner difference is more for a property sold for 1,5 million kroner than a property sold for 11 millions kroner. If we rearrange the variables in (A.1), the (total) model for property prices may be written as

$$\log P_{\text{act}} = \log P_\omega + \epsilon, \tag{A.2}$$

where $\log P_\omega$ is the weighted point estimate and $\epsilon$ is describing the uncertainty of the point estimate. The latter quantity is a distribution, rather than a point, defined by its mean, its shape and its standard deviation.

In the following sections A.1 and A.2, we will describe how the weighted point estimate $P_\omega$ and the uncertainty distribution $\epsilon$, respectively, are computed.

## A.1 The weighted model

Let $P_i$, $i = 1, \ldots, n$, be the different basic estimates for a property. The weighted value estimate $P_\omega$ can be written as

$$P_\omega = \sum_{i=1}^{n} \omega_i P_i,$$

where $\omega_i$ is a weight corresponding to each basic estimate $P_i$.

The weights $\omega_i$s are not equal across the properties, but will vary according to specific characteristics of the methods and properties. These specific characteristics, denoted by $x_k$, $k = 1, \ldots, m_1$, are so-called known, explanatory variables. The model for the (non-scaled) weights are given by

$$v_i = \exp\left\{\sum_{k=1}^{m_1} \beta_k x_k\right\}, \qquad i = 1, \ldots, n, \tag{A.3}$$

where the $\beta_k$s are parameters belonging to each known characteristic $x_k$. For simplicity, we have suppressed the index $i$ from $\beta_k$ and $x_k$. The $\beta_k$s are unknown and will be estimated based on historical data. Note that the exponential function in (A.3) is forcing the weights $v_j$ to be positive. There will be one weight belonging to each of the basic estimates. The weight equals zero, $v_i = 0$, if the basic estimate $P_i$ does not exist for a particular property. We let the weights summarise to one by computing

$$\omega_i = \frac{v_i}{\sum_{j=1}^{n} v_j}, \qquad i = 1, \ldots, n.$$

Thus, the weight $\omega_i$ is equal one if only one of the basic estimates $P_i$ exist.

The weights (i.e. the unknown parameters $\beta_k$s) are estimated so that the error (A.1) between the actual sales price and the estimated price is as small as possible using historical data. In other words, our approach finds the set of unknown parameters $\beta_k$, $k = 1, \ldots, m$, for each weight $\omega_i$ such that the sum of quadratic errors over all $N$ properties

$$\sum_{l=1}^{N} (\log P_{\mathrm{act},l} - \log P_{\omega,l})^2 \tag{A.4}$$

is minimised. This sum can be minimised numerically, for example by using the R-package `nlminb()`. When we have found the estimates for the $\beta_k$s, we may use the model (A.3) to compute the weights and compute the weighted price estimate, $P_\omega$, for new properties in the data base of Eiendomsverdi.

## A.2 The uncertainty model

Here, we describe the model of uncertainty for the weighted estimate $P_\omega$. The uncertainty distribution of the error in the model for prices in (A.1) is described by its standard deviation and its shape.

To find the standard deviation, we assume for the moment that the errors $\epsilon$ have zero means and standard deviations $\sigma$. More specifically, the variance $\sigma^2$ is estimated by a gamma linked generalised linear model (GLM) using the square of the error $\epsilon^2$ as responses.[2] Thus, the model for the variance on logarithmic scale can be written as a regression model with mean

$$\log \sigma^2 = \gamma_0 + \sum_{k=1}^{m_2} \gamma_k z_k,$$

where the $z_k$s are known explanatory variables and the $\gamma_k$s are unknown parameters. Once we have found those $\gamma_i$s that fit the GLM model best, the variance can be computed by

$$\sigma^2 = \exp\left\{ \gamma_0 + \sum_{k=1}^{m_2} \gamma_k z_k \right\}. \tag{A.5}$$

2. Note that the variance of $\sigma^2$ of the error $\epsilon$ is given as the mean of the squared error $\epsilon^2$ since $\sigma^2 = \mathrm{Var}(\epsilon) = \mathrm{E}(\epsilon^2) - (\mathrm{E}(\epsilon))^2 \approx \mathrm{E}(\epsilon^2) - 0 = \mathrm{E}(\epsilon^2)$. The approximation is reasonable since $\mathrm{E}(\epsilon)$ is small compared to $\mathrm{E}(\epsilon^2)$.

The $z_k$ can either be the same as the explanatory variables $x_k$ in (A.3) or other known variables that affects the uncertainty of $P_\omega$.

The next step is to define the distribution of the error terms. Instead of using a parametric distribution (such as the Gaussian distribution), we assume that the error has the same shape as the empirical distribution for the error $\epsilon$ in (A.1). We want this distribution to be the same for all properties (i.e. within a specific property type such as detached homes). Thus, we define a new variable $\epsilon_l^\star$ for each property $l$, by

$$\epsilon_l^\star = \frac{\epsilon_l}{\sigma_l}, \qquad l = 1, \ldots, N, \tag{A.6}$$

which will have standard deviation 1 for all $l$. This means that $\epsilon_l^\star$ has the same variance for all properties in the data set.

The distribution of $\epsilon_l^\star$ is decided from the (historical) data. The $\epsilon_l$ is computed by (A.1) using the known actual sales prices from the historical data set, while $\sigma_l$ is computed from (A.5). The set of all $\epsilon_l^\star$ forms the empirical distribution of the stochastic variable $\boldsymbol{\epsilon}^\star$:

$$\boldsymbol{\epsilon}^\star = \{\epsilon_1^\star, \epsilon_2^\star, \ldots, \epsilon_N^\star\},$$

where all $\epsilon_l^\star$s are assumed to have equal probability of occurance.

Using the empirical distribution of $\boldsymbol{\epsilon}^\star$ and that $\epsilon_l = \sigma_l \epsilon^\star$ in the model for property prices (A.2), the empirical distribution for property $l$ may be written as

$$\begin{aligned} \mathbf{P}_l &= \exp\{\log P_{\omega,l} + \boldsymbol{\epsilon}^\star \sigma_l\} \\ &= P_{\omega,l} \exp\{\boldsymbol{\epsilon}^\star \sigma_l\}. \end{aligned} \tag{A.7}$$

This distribution will typically be skew, with an upper heavier tail. Several conclusions may be drawn from the uncertainty distribution than just a point estimate, for example a value telling that we are 75% certain that the sales price would be higher.