A Heuristic Robust Approach for Real Estate Valuation in Areas with Few Transactions

Alexander DORNDORF, Matthias SOOT, Alexandra WEITKAMP and Hamza ALKHATIB, Germany

Key words: Real Estate Valuation, Areas with Few Transactions, Robust Estimation

SUMMARY

The German market transparency is mainly realized by results of analyzing purchase prices. Often, the purchases are analyzed in the context of a regression approach. The results are only reliable in areas with large numbers of purchases. However, in areas with only few transactions the solution of regression is not satisfactory. Furthermore, the purchase prices may contain outliers. Especially in areas with few transactions, the detection of outliers is a challenging task. This study presents three different estimation approaches which are dealing with outliers. The first approach uses the data snooping to detect the outliers. The second approach is based on a heuristic RANSAC (random sample consensus) algorithm. The third approach uses non–informative robust Bayesian regression techniques, in which the normal distribution of the likelihood data is replaced by a Student–distribution to ensure the robustness. The aim of this study is to investigate these three approaches in their efficiency to deal with outliers in areas with few transactions. For this purpose a closed loop simulation is carried. The results of the three robust approaches are compared based on the known regression coefficients and on the known observations. The results of the data snooping and RANSAC show that the estimation fail more often than the estimation by means of the robust Bayesian approach, which shows a suitable result for areas with few transactions.

ZUSAMMENFASSUNG

Die Markttransparenz in Deutschland wird hauptsächlich durch die Ergebnisse von analysierten Kaufpreisen realisiert. Meistens werden die Kauffälle mit einem Regressionsansatz untersucht. Die Ergebnisse sind nur für Gebiete mit einer großen Anzahl an Kauffällen zuverlässig. Allerdings ist die Lösung der Regression für Gebiete mit wenigen Transaktionen nicht zufriedenstellend. Außerdem können die Kaufpreise Ausreißer enthalten. Vor allem in Gebieten mit wenigen Transaktionen ist das Finden der Ausreißer eine herausfordernde Aufgabe. Diese Studie präsentiert drei verschiedene Schätzansätze, die mit Ausreißern umgehen können. Der erste Ansatz verwendet das Data Snooping um die Ausreißer zu finden. Der zweite Ansatz basiert auf dem RANSAC Algorithmus. Der dritte Ansatz verwendet eine nichtinformative robuste bayessche Regressionstechnik. Die Normalverteilung der Likelihood Daten ist durch eine Student–Verteilung ersetz um die Robustheit zu gewährleisten. Das Ziel dieser Studie ist es diese drei Ansätze auf ihre Effizienz mit Ausreißern in Gebieten mit wenigen Transaktionen umzugehen zu untersuchen. Hierfür wird eine Closed Loop Simulation mit den drei Ansätzen durchgeführt. Die Ergebnisse der drei robusten Ansätze werden mittels der bekannten Regressionskoeffizienten und Beobachtungen verglichen. Die Ergebnisse vom Data Snooping und RANSAC zeigen ein erhöhtes Risiko das die

Schätzung versagt. Nur die Ergebnisse vom robusten bayesschen Ansatz zeigen ein adäquates Ergebnis für Gebiete mit wenigen Transaktionen.

A Heuristic Robust Approach for Real Estate Valuation in Areas with Few Transactions

Alexander DORNDORF, Matthias SOOT, Alexandra WEITKAMP and Hamza ALKHATIB, Germany

1. MOTIVATION

The demand for reliable market values in real estate valuation has been increasing over the last decade. One reason is the last subprime crisis, which was caused by a false assessment of the real estate market. The German market transparency is mainly realized by results of analyzing purchase prices. Often, the purchases are analyzed with a multiple linear regression, which allows for an adequate examination of the real estate market. For an accurate estimate of the regression coefficients, the regression model normally needs 15 purchases per independent variable (Ziegenbein 2010, Kleiber et al. 2014), but in areas with few transactions only few prices are available. Hence, in areas with only few transactions the solution of regression is not satisfactory. A small number of purchase prices cannot represent the heterogeneity of the real estate market. Furthermore, the purchase prices may contain outliers. Especially in areas with few transactions, the detection of outliers is a challenging task.

Recently, property appraisers use their market expertise to determine market values in regions with few transactions. The few purchases are often not used methodically. Hence, a statistical approach would be preferable. In this context Alkhatib & Weitkamp (2013) and Weitkamp & Alkhatib (2014) suggested a robust Bayesian regression model to deal with the problems caused by outliers. They replaced the well-known normal distribution of the likelihood data by a Student-distribution that allows keeping outliers in the estimation but to down weight their influence on the estimated results. Furthermore, they used additional market information, e.g. results of an experts' survey, in this robust Bayesian approach. The additional market information support the few available purchases in the estimation.

The aim of this study is to investigate the efficiency of different robust estimation approaches to deal with outliers in areas with few transactions. One way of dealing with outlying observations is to apply data snooping with Baarda's or Pope's outlier test and to eliminate the detected outliers (Jäger et al. 2005). These tests are prone to fail; consequently, the results contain false and missed detections, especially when the outlier ratio is large. In this case the Random Sample Consensus (RANSAC) algorithm is an alternative, which allows up to 50% outliers in the data. This algorithm searches randomly for a given model the amount of observations with the largest number. Hence, RANSAC is a heuristic approach for outlier detection. However, the main disadvantage of data snooping or RANSAC to detect outliers is the reducing of the sample size. Another approach for dealing with outliers is the application of robust estimation approaches, which handle outliers by down-weighting their influence on the estimated values. Classical robust estimators are for example the L1–norm estimator or Huber's M–estimator (Koch 1999). An alternative to the abovementioned classical robust estimators is a Bayesian robust estimator, in which the normal distribution of the likelihood function is replaced by a longer-tailed distribution as the family of the Student-

distribution. The use of this Bayesian robust model that applies the Student-t in place of a normal distribution is on the one hand to down weight the influence of the occurring outliers, on the other side to assess sensitivity to the normal assumption by varying the degrees of freedom from large to small (Gelman et al. 2014).

In this study we investigate the data snooping, RANSAC algorithm and the robust Bayesian approach via a closed loop simulation. Real purchase prices are used to create the simulation data. This allows for reproducing the characteristics of the real data in the simulation. Furthermore, the simulation data includes also outliers. The results of the three estimators are compared based on the known regression coefficients and on the known simulated observations.

2. MATHEMATICAL BASICS

2.1 Multiple Linear Regression in Real Estate Valuation

Since decades, the multiple linear regression is used as tool in the real estate valuation. In the sales comparison approach the input quantities of a real estate (e.g. area of lot or standard land value) explain the purchase price by the regression model (Ziegenbein 1977). The functional model follows:

$$y_i = \beta_1 + x_{i,1}\beta_2 + \cdots + x_{i,u-1}\beta_u + e_i \;\; ; \;\; i = 1, \dots, n \;\; ; \;\; e_i \sim N(0, \sigma^2). \qquad \textbf{Eq. 1}$$

The dependent variable y_i (in our case: standardized purchase price) of n observed purchases is explained by a linear combinations of the independent variables $x_{i,1}, \dots, x_{i,u-1}$ and the u unknown regression coefficients $\boldsymbol{\beta}$. The residuals e_i are the differences between predictions and observations that arise as measure of the not explainable spread between model and reality. They have to obey the normal distribution with the mean value 0 and the variance σ^2. The unknown regression coefficients are usually estimated by means of the method of least squares (Fahrmeir et al. 2013, Koch 1999). Then $\boldsymbol{\beta}$ is given by:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y} \, . \qquad \textbf{Eq. 2}$$

Further discussion of regression analysis can be found in e.g. Fahrmeir et al. (2013).

2.2 Data Snooping

The classic approach of dealing with outliers is to apply the data snooping. The data snooping based on a hypothesis test statistics, e.g. Baarda test (Baarda 1968) or Pobe test (Pobe 1976). At first for test statistics a hypothesis H_0 is required:

$$\text{accept} \quad H_0 \;\text{ if } \quad T_i < c$$

$$\text{reject} \quad H_0 \;\text{ if } \quad T_i \geq c$$

where T_i is the test value of the observation i and c is a critical value from the test distribution for a given confidence level $1 - \alpha$. The hypothesis is accepted, if the test value is smaller than the critical value. In this case, the null hypothesis will be accepted and it results that the observation i contains no outlier. In this paper, we use the Pobe test, which allows to detect outliers at unknown variance factor. The test value is calculated as follow:

$$T_i = \sqrt{\frac{\hat{e}_i^2}{\hat{\sigma}^2 q_{\hat{e}_i}^2}} \sim \tau_{1-\alpha,1,n-(u+1)}. \quad \textbf{Eq. 3}$$

$$\text{with } \boldsymbol{Q}_{ee} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T. \quad \textbf{Eq. 4}$$

The diagonal values from the cofactor matrix of residuals $\boldsymbol{Q}_{ee}$ are the values $q_{\hat{e}_i}^2$ in Eq. 3. The determination of the residuals $\hat{e}_i$ and the variance factor σ^2 are realized by:

$$\hat{e}_i = \boldsymbol{x}_i\boldsymbol{\beta} - y_i. \quad \textbf{Eq. 5}$$

$$\hat{\sigma}^2 = \frac{\hat{e}^2}{n-u}. \quad \textbf{Eq. 6}$$

The distribution of the test value T_i given in Eq. 3 obey Pobe distribution τ. For a given significance level α, we can derive the critical value c for this test as follows:

$$c = \tau_{1-\alpha,r,n-(u+r)} = \left(\frac{(n-u)F_{1-\alpha,r,n-(u+r)}}{n-u-r+r\,F_{1-\alpha,r,n-(u+r)}}\right)^{1/2}. \quad \textbf{Eq. 7}$$

In the case that only one observation is detected as an outlier, the value r in the τ-distribution is one. We use here $\alpha = 5\%$ as significance level. The τ-distribution is approximated by means of the F-distribution, also known as Fisher distribution. Below are all important equations for implementation of the Pobe test and the data snooping:

Step 1: Estimate the regression parameters with Eq. 2.

Step 2: Calculate for each observation the test value T_i and approximate critical value c with the τ-distribution.

Step 3: If H_0 is rejected in at least one observation, delete the observation with the largest T_i value.

Step 4: Repeat step 1 to 3 until H_0 is accepted for each observation. Then reestimate the regression coefficients with the reduced coefficient matrix $\boldsymbol{X}_{\mathrm{DS}}$ and observations $\boldsymbol{y}_{\mathrm{DS}}$ as follow:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{DS}} = (\boldsymbol{X}_{\mathrm{DS}}^T\boldsymbol{X}_{\mathrm{DS}})^{-1}\,\boldsymbol{X}_{\mathrm{DS}}^T\,\boldsymbol{y}_{\mathrm{DS}}. \quad \textbf{Eq. 8}$$

Consequently, the data snooping is an iterative process for detection and elimination of outliers. A detailed discussion of data snooping can be found in Jäger et al. (2005).

2.3 Bayes Inference as Robust Approach

In contrast to classical statistical inference, the Bayesian inference uses probability distributions to determine the unknown parameters of a regression model. The Bayesian inference is based on the Bayes' theorem:

$$p(\boldsymbol{\beta}|\boldsymbol{y}) \propto p(\boldsymbol{\beta}) \cdot p(\boldsymbol{y}|\boldsymbol{\beta}). \quad \textbf{Eq. 9}$$

In this equation $p(\boldsymbol{\beta})$ is the prior density of the regression parameters and the $p(\boldsymbol{y}|\boldsymbol{\beta})$ is denoted as likelihood function. The likelihood function represents the information of the observations (in our case: purchases). All additional information about the unknown parameters are expressed and modeled in the prior density. In case of real estate valuation, additional information are for example offer prices (Soot et al. 2016). $p(\boldsymbol{\beta}|\boldsymbol{y})$ is called posterior density, from it's the posterior unknown regression parameters can be derived by given data $\boldsymbol{y}$. Detailed information about the Bayesian

inference can be found in, e.g., Koch (2007), Gelman et al. (2014). In Alkhatib & Weitkamp (2012) and Weitkamp & Alkhatib (2012) a Bayesian approach was developed to combine normal distributed prior information, which was generated by interviews with valuation expert, with normal distributed likelihood, which summarizes the given data (purchases).

The abovementioned assumptions are stated only if the quality of the data is ensured. So, e. g., if the data base contains some outliers, then the classical Bayesian regression approach can fail and an alternative robust Bayesian regression model should be used. The robustness can be realized, e.g., by assuming the Student-distribution for the likelihood function. The Student t-distribution has a longer tail than the normal distribution. Hence, observations with great dispersion have a smaller influence on the estimated parameters. For the data (here the purchases) in the context of the Bayesian multiple linear regression follow:

$$y_i|\boldsymbol{X} \sim t(\boldsymbol{x}_i \cdot \boldsymbol{\beta}, \sigma^2\ ;\ \nu\ ). \qquad \textbf{Eq. 10}$$

The observation y_i is conditioned by the independent variables $\boldsymbol{x}_i$ of the purchases. The observations are now assumed to obey the univariate Student-distribution. In Eq. 10 $\boldsymbol{x}_i \cdot \boldsymbol{\beta}$ is the center μ of the t-distribution. The variance σ^2 scales the t-distribution and ν is the unknown degree of freedom. The length of the tails of the t-distribution depends on the choice of ν. If $\nu > 30$, the t-distribution is nearly equivalent with the normal distribution. Hence, the degree of freedom must be less than 30 for a robust estimate. In Gelman et al. (2014) a value of 4 is suggested for ν. An alternative to fix ν is to estimate ν as unknown parameter (Geweke 1993). We use a fixed degree of freedom with 4 as suggested in Gelman et al. (2014). The term in Eq. 10 can be equivalent specified as follow:

$$y_i|\boldsymbol{X}, \boldsymbol{V}_i \sim N(\boldsymbol{x}_i \cdot \boldsymbol{\beta}, \boldsymbol{V}_i)\ \ ;\ \text{with}\ \boldsymbol{V}_i \sim inv\,\chi^2(\sigma^2, \nu). \qquad \textbf{Eq. 11}$$

In this equation the observation y_i is normal distributed under the condition of $\boldsymbol{X}$ and $\boldsymbol{V}_i$. The center of the normal distribution is the same as in the t-distribution. For Eq. 10 and Eq. 11 are equivalent, the variance of the normal distribution must follow an inverse χ^2-distribution. Now, we apply Eq. 11 on the functional model of the multiple linear regression in Eq. 1. Thus follows:

$$y_i = \beta_1 + x_{i,1}\,\beta_2 + \cdots + x_{i,u-1}\,\beta_u + e_i\ \ ;\ e_i \sim N(0, \sigma^2\omega_i). \qquad \textbf{Eq. 12}$$

$$\text{with}\ (\boldsymbol{e}) = \sigma^2\boldsymbol{\Omega}\,, \quad \boldsymbol{\Omega} = \begin{bmatrix} \omega_1 & & & 0 \\ & \omega_2 & & \\ & & \ddots & \\ 0 & & & \omega_n \end{bmatrix}. \qquad \textbf{Eq. 13}$$

The functional model in Eq. 12 is nearly identical with Eq. 1. The residuals $\boldsymbol{e}$ in Eq. 12 are normal distributed, but the variance σ^2 and weights $\boldsymbol{\omega}$ obey an inverse χ^2-distribution. The weights $\boldsymbol{\omega}$ are summarized in the weight matrix $\boldsymbol{\Omega}$. Due to the prior assumption that σ^2 and $\boldsymbol{\omega}$ are inverse χ^2-distributed, the prior density of the Bayes' theorem is now of the form $p(\boldsymbol{\beta}, \sigma, \boldsymbol{\omega}) = p(\boldsymbol{\beta})\,p(\sigma)\,p(\boldsymbol{\omega})$. In this model the prior density of the coefficient $\boldsymbol{\beta}$, standard deviation σ and weights $\boldsymbol{\omega}$ are assumed to be independent among each other. Detailed formula descriptions of the prior densities and the likelihood function is presented in Geweke (1993).
The Bayesian inference distinguishes between conjugate and nonconjugate prior distributions (Gelman et al. 2014). In case of conjugate prior, the posteriori density belongs to the same family of

distribution as the prior density. This has the advantage, that the posterior density can be solved analytically. Due to the use of the inverse χ^2-distribution as prior, the posterior distribution doesn't belong to the same family of prior. In this case the posterior density must be solved numerically with Markov chain Monte Carlo (MCMC) methods. We use in this paper the Gibbs sampler (one of the most famous MCMC algorithm) to calculate the posterior density due to its simplicity.
Another distinction in the Bayesian inference is the informative and noninformative prior distribution (Gelman et al. 2014). The reason for using noninformative prior distribution is the no existing of previous knowledge about the parameters, or we are interested in an estimate of our posterior parameters, which is not affected by external information to our current data. In the case of informative prior, it exists previous knowledge of the regression coefficients $\boldsymbol{\beta}$, e.g. offer prices or knowledge of experts. In this paper the noninformative Bayesian approach, presented in Geweke (1993), is used.

Initial Values:

$$\boldsymbol{\beta}^{(0)} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$$

$$[\sigma^{(0)}]^2 = \frac{[\boldsymbol{e}^{(0)}]^{\mathrm{T}}\boldsymbol{e}^{(0)}}{(n-u)} \quad \text{with } \boldsymbol{e}^{(0)} = \boldsymbol{X}\boldsymbol{\beta}^{(0)} - \boldsymbol{y}$$

$$\omega_i^{(0)} = 1 \; ; \; i=(1, \ldots, n)$$

$$\nu = 4$$

Results:

$$\hat{\boldsymbol{\beta}}_{\mathrm{Bayes}} = \frac{1}{s_2 - s_1} \sum_{j=s_1}^{s_2} \boldsymbol{\beta}^{(j)} \qquad \textbf{Eq. 14}$$

$$\mathrm{Var}\,[\,\hat{\boldsymbol{\beta}}_{\mathrm{Bayes}}^{(\#)}\,]$$

$$\text{with } \# = \{s_1, \ldots, s_2\}$$

for $j = 0, \ldots, s_2$

1. Draw $\omega_i^{(j+1)}$ conditional on $\boldsymbol{\beta}^{(j)}$ and $\sigma^{(j)}$ with:

$$\left[\frac{([\sigma^{(j)}]^{-2}[e_i^{(j)}]^2 + \nu)}{\omega_i}\right] \Big| (\boldsymbol{\beta}, \sigma) \sim \chi^2(\nu+1) \qquad \textbf{Eq. 15}$$

2. Draw $\sigma^{(j+1)}$ conditional on $\boldsymbol{\beta}^{(j)}$ and $\boldsymbol{\omega}^{(j+1)}$ with:

$$[\sum_{i=1}^{n} ([e_i^{(j)}]^2/\omega_i^{(j+1)})/\sigma^2\,] | (\boldsymbol{\beta}, \boldsymbol{\omega}) \sim \chi^2(n) \qquad \textbf{Eq. 16}$$

3. Draw $\boldsymbol{\beta}^{(j+1)}$ conditional on $\boldsymbol{\omega}^{(j+1)}$ and $\sigma^{(j+1)}$ with:

$$\boldsymbol{\beta}|(\sigma,\boldsymbol{\omega}) \sim N|[\,\hat{\boldsymbol{\beta}}(\boldsymbol{\omega}^{(j+1)}), [\sigma^{(j+1)}]^2\,(\boldsymbol{X}^{\mathrm{T}}[\boldsymbol{\Omega}^{(j+1)}]^{-1}\boldsymbol{X})^{-1}\,] \qquad \textbf{Eq. 17}$$

$$\text{with } \hat{\boldsymbol{\beta}}(\boldsymbol{\omega}^{(j+1)}) = (\boldsymbol{X}^{\mathrm{T}}[\boldsymbol{\Omega}^{(j+1)}]^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}[\boldsymbol{\Omega}^{(j+1)}]^{-1}\boldsymbol{y}$$

end

Figure 1: Calculation procedure for the Gibbs sampler.

The MCMC implementation of this approach based on Gibbs sampler is shown in Figure 1. Firstly, we select for the Gibbs sampler initial values for $\boldsymbol{\beta}$, σ, $\boldsymbol{\omega}$ and ν. Based on this initial values, the marginal distribution of $\boldsymbol{\beta}$, σ and $\boldsymbol{\omega}$ can be iteratively approximated (see Figure 1, the loop from 0

to s_2). In first step, the weights $\boldsymbol{\omega}$ are drawn under the condition of $\boldsymbol{\beta}$ and σ from the previous iteration (Eq. 15). After that, the variance σ^2 is drawn by means of Eq. 16 under the condition on $\boldsymbol{\beta}$ from the previous iteration and the new drawn weights $\boldsymbol{\omega}$. The weights and variance are generated from a χ^2-distribution. Because of the division of the drawn numbers through the numerator of Eq. 15 and Eq. 16 respectively, the weights and the variance are inverse χ^2-distribution as defined in Eq. 11. In third step the regression coefficients $\boldsymbol{\beta}$ are generated according to Eq. 17. The regression coefficients are normally distributed conditional on the new drawn weights and variance from step one and two. The repetition of the abovementioned three-steps results in the desired Markov chain. The result of the loop is the mean (Eq. 14) and the variance of the conditional regression coefficients $(\boldsymbol{\beta}|\sigma, \boldsymbol{\omega})$ for the samples from s_1 to s_2. The precision and accuracy of the Gibbs sampler results depends inter alia on the choice of the number of iterations s_2 and the warm-up period s_1. Further discussion of MCMC and Gibbs sampler can be found in, e.g. Gelman et al. (2014).

2.4 Random Sample Consensus

The Random Sample Consensus (RANSAC) algorithm is introduced by Fischler & Bolles (1981). This algorithm is developed to deal with a very large outlier percentage (e.g. 50%) in the data set. Typical application areas of RANSAC are the image analysis or the terrestrial laser scanning (Hartley & Zisserman 2004). The advantage of RANSAC is its simple implementation for different mathematical models. In this paper, the multiple linear regression (Eq. 1) is used as functional model. The RANSAC algorithm is implemented with following four steps:

Step 1: Randomly choose of u purchases from data set and compute the regression coefficient $\boldsymbol{\beta}^{(j)}$ (starting with $j = 1$), which defines the hypothetical model M_j.

Step 2: Compute for the calculate coefficient $\boldsymbol{\beta}^{(j)}$ the residuals of all purchases and determine the observations whose residuals are smaller than the error tolerance S. These observations define the consensus set Z_j of the model M_j.

Step 3: Go back to step 1 and repeat the process for another random choice of u purchases from the data set until a specified number N of iterations is reached.

Step 4: Determine the model with the greatest consensus set Z_{max}. Carry out an optimal estimation of the parameter vector $\boldsymbol{\beta}$ with the purchases in Z_{max}:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{RANSAC}} = \left(\boldsymbol{X}_{Z_{max}}^T \boldsymbol{X}_{Z_{max}}\right)^{-1} \boldsymbol{X}_{Z_{max}}^T \boldsymbol{y}_{Z_{max}}. \qquad \textbf{Eq. 18}$$

The efficiency of RANSAC algorithm depends on the choice of the error tolerance S and the number N of iterations. The error tolerance can be interpreted as Euclidean distance between the observations and a model M_j. The numerical specification of the error tolerance based on the prior knowledge of the dispersion of the observations and the used functional model. If S is selected too small, then good observations are interpreted as outliers. Otherwise, the outliers are interpreted as model conform if S is too large. The choice of the iteration number N can be approximated by the probability P of randomly choosing at least one model M_j, which has no outliers in the u observations. The calculation of the minimum number of iterations follows (Hartley & Zisserman 2004):

$$N = \frac{ln(1-P)}{ln(1-(1-\delta)^u)}. \qquad \textbf{Eq. 19}$$

The term δ is the outlier ratio in the data set. In the case that the outlier ratio is 25%, the model has 6 regression coefficients and we choose a probability P of 99%, the number of iterations N according to Eq. 19 is 24. Further discussion of RANSAC can be found in, e.g., Fischler & Bolles (1981), Hartley & Zisserman (2004).

3. DEVELOPMENT OF A STRATEGY FOR THE INVESTIGATION

In this paper, the focus lies on investigating the efficiency of the different estimation approaches to deal with outliers in areas with few transactions. Hence, a closed loop simulation is developed to validate the estimation approaches from Section 2. For this investigation strategy simulated purchases are required, which are derived from real purchases (see Subsection 3.2). The advantage of a closed loop simulation is, that we know the expected values of the regression coefficients and the purchase prices. This allows to compare the estimated results of the different approaches with the expected values. Detailed information about closed loop simulation and Monte Carlo are presented in Saltelli et al. (2008) and Kroese et al. (2011).

3.1 Closed Loop Simulation

The calculation process of the closed loop simulation is schematically illustrated in Figure 2. As input parameter the independent variables of purchases, which build the data set $\boldsymbol{X}$, are required. In order to select reliable expected value $\mathrm{E}(\boldsymbol{\beta})$ we use the estimated regression coefficient of the data set $\boldsymbol{X}$. The error model $\boldsymbol{\Psi}$ can be set up by using a known distribution, or a combination of different, from which random noise is generated. The used error model is presented in Subsection 3.3. If all input parameters are given, then the expected values of the observations $\mathrm{E}(\boldsymbol{y})$ are calculated by means of $\boldsymbol{X}$ and $\mathrm{E}(\boldsymbol{\beta})$. The expected values of the observations and the independent variables build the set $\boldsymbol{D}_A$.

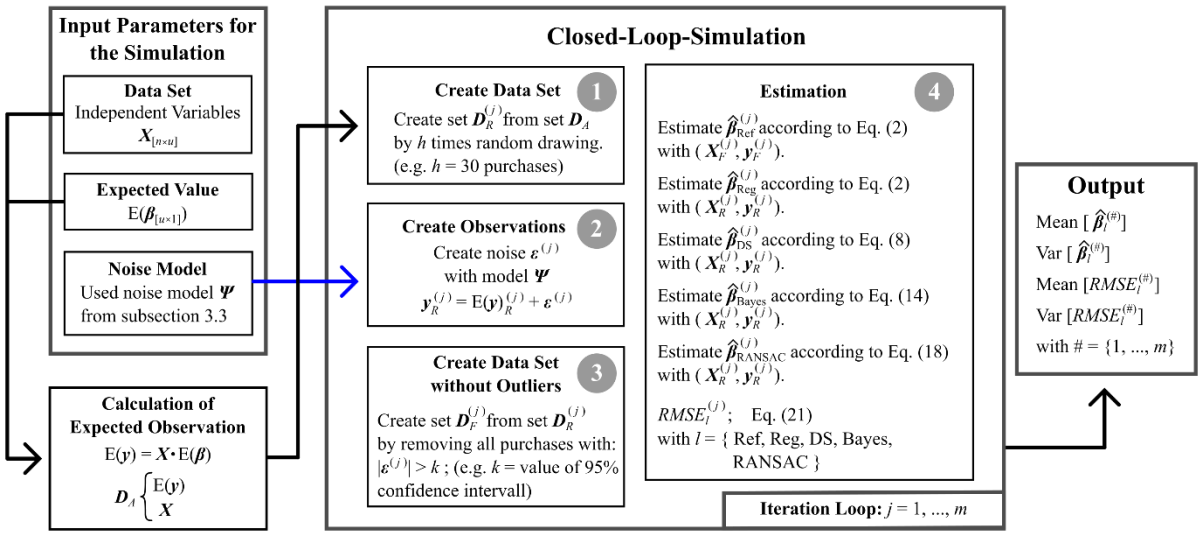

Figure 2: Schematically sequence of the closed loop simulation.

The data set $\boldsymbol{D}_A$ and the noise model $\boldsymbol{\Psi}$ are initial values for the closed loop simulation. At first, a set $\boldsymbol{D}_R^{(j)}$ is compiled from set $\boldsymbol{D}_A$ by h times random drawing in every iteration j (step 1). In step 2 the simulated purchase prices $\boldsymbol{y}_R^{(j)}$ are generated for the data set $\boldsymbol{D}_R^{(j)}$. For this, the random noise $\boldsymbol{\varepsilon}^{(j)}$ is generated from the noise model $\boldsymbol{\Psi}$ by h times random drawing. The simulated purchase prices consist of the expected observations and the noise $\boldsymbol{\varepsilon}^{(j)}$. The independent variables $\boldsymbol{X}_R^{(j)}$ and the simulated purchase prices $\boldsymbol{y}_R^{(j)}$ of the data set $\boldsymbol{D}_R^{(j)}$ are the transactions on the simulated submarket. In step 3 the outliers in resulting simulated purchases of data set $\boldsymbol{D}_R^{(j)}$ are removed. The resulting data set is denoted with $\boldsymbol{D}_F^{(j)}$. In this study, we define outliers in the purchase price as observation, which differs significantly from the normal distributed noise in the submarket. For the detection of the outliers the threshold k is specified. The selection of the threshold k depends on the noise model $\boldsymbol{\Psi}$. A further discussion of the selection of k are presented in Subsection 3.3. In step 4 the coefficients $\widehat{\boldsymbol{\beta}}$ of data set $\boldsymbol{D}_R^{(j)}$ are estimated using four estimation approaches introduced in Section 2. In addition, the regression coefficients $\widehat{\boldsymbol{\beta}}_{\text{Ref}}^{(j)}$ of the data set $\boldsymbol{D}_F^{(j)}$ are estimated by the multiple linear regression (Eq. 2). The coefficients $\widehat{\boldsymbol{\beta}}_{\text{Ref}}^{(j)}$ are used as reference values to the estimated coefficients $\widehat{\boldsymbol{\beta}}_{\text{Reg}}^{(j)}$, $\widehat{\boldsymbol{\beta}}_{DS}^{(j)}$, $\widehat{\boldsymbol{\beta}}_{\text{Bayes}}^{(j)}$ and $\widehat{\boldsymbol{\beta}}_{RANSAC}^{(j)}$ of the data set $\boldsymbol{D}_R^{(j)}$. This means that the estimated coefficients $\widehat{\boldsymbol{\beta}}_{\text{Ref}}^{(j)}$ are only influenced by normal distributed noise, which corresponds to the optimal case that all outliers are detected by the data snooping. Furthermore, the root mean square error (RMSE) is calculated for all estimated coefficients (Eq. 21). Therefore, the residuals $\hat{\boldsymbol{e}}_l^{(j)}$ are required. Note please that the expected purchase prices $\mathrm{E}(\boldsymbol{y}_l)^{(j)}$ and not the simulated purchase prices $\boldsymbol{y}_l^{(j)}$ are used in Eq. 20. Hence, the $RMSE_l^{(j)}$ describes how well the predicted purchase prices correspond to the true values in the mean. The true value is unknown in praxis, because it is not possible to determine all error influences of one purchase reliably. Examples for error influences are the imperfection of the functional model to describe the reality or the individual purchase negotiation between seller and buyer.

$$\hat{\boldsymbol{e}}_l^{(j)} = \boldsymbol{X}_l^{(j)}\widehat{\boldsymbol{\beta}}_l^{(j)} - \mathrm{E}(\boldsymbol{y}_l)^{(j)} \,. \qquad \textbf{Eq. 20}$$

$$RMSE_l^{(j)} = \sqrt{\frac{1}{n_l}\sum\left(\hat{\boldsymbol{e}}_l^{2\,(j)}\right)} \,. \qquad \textbf{Eq. 21}$$

$$l = \text{Ref, Reg, DS, Bayes, RANSAC}$$

These four steps are repeated m times in the loop. The number of iteration loops determines the precision of the simulation result. Repetitions of this simulation show that the different results are approximately equal for $m = 100'000$. Therefore we fixed the number of iterations to 100'000 iteration loops, which are sufficient for this study. The final result of the closed loop simulation is the mean and the variance of the m estimated regression coefficients and RMSE values for the five estimations.

3.2 Data Base

The data set which is used as input value for the simulation had been collected in Nienburg (Weser), a small city with approximately 30'000 inhabitants in the south of Lower Saxony. The spatial submarket of Nienburg (Weser) has a regular supply and demand situation. As the functional submarket, we use the market of one and two-family houses. In the period of 2011 to 2015 about 260 purchases are available. All these purchases have the following influence quantities: "area of lot" [sq. m], "standard land value" [EUR per sq. m], "construction year" [age – 1946], "living space" [sq. m] and "equipping standard" [without unit]. These independent variables are used as input parameter $\boldsymbol{X}$ for the simulation. In Table 1 the estimated regression coefficients for these purchases are presented, which are used as expected value E($\boldsymbol{\beta}$) in the simulation. Detailed information of the used spatial and functional submarket can be found in Soot et al. (2016).

Table 1: Result of the multiple linear regression for the used purchases.

β_1 Intercept	β_2 Living space	β_3 Area of lot	β_4 Construction year	β_5 Standard land value	β_6 Equipping standard
336.40	– 4.57	0.33	12.65	3.05	176.39

3.3 Noise Model

The used noise model $\boldsymbol{\Psi}$ is a mixed normal distribution, which consists of a combination of three normal distributions with different mean values and variances (Figure 3). The blue depicted normal distribution is used for the creation of the normal measurement noise, which correspond to the normal distributed residuals $\boldsymbol{e}$ in Eq. 1. The mean value of this distribution is **0** and the variance is derived from the scattering of the real purchase prices. In the used submarket the RMSE value is about 200 € per sq. m. Therefore, a standard deviation of 150 € per sq. m. is used for the creation of normal measurement noise. This value is smaller than for the real purchases, but together with the outliers the scattering of the noise model $\boldsymbol{\Psi}$ is approximately the same. The 95% confidence interval boarders of the normal measurement noise are illustrated in green. In this study we define all purchases as outlier, which noise is outside of this confidence interval. From this follows that the threshold k is 294 € per sq. m.

The outliers are generated from the mixture distribution (Figure 3, red line). The mixture distribution consists of three normal distributions ($\mu = [545, 555, 570]; \sigma = [15, 30, 60]$). The selection of the parameters for the mixture distribution depends on several conditions. For the simulated submarket we assume that 20% of the objects are traded significantly above the average prices. This assumption is possible for areas with few transactions, because the market transparency in these areas is not guaranteed. Hence, some buyers pay more as the usual market price. This is the reason why the mixture distribution lies only on the right side of the confidence interval. A further condition for the construction is to the fact that no outliers are created inside of the confidence interval. The probability of the used mixture distribution is nearly zero to generate a random number, which is smaller than the threshold k. We assume that the most probably value of the mixture distribution can be maximally of factor two larger than the threshold and it is improbable that an outlier is three times larger than the threshold. Furthermore, it is more probably that an outlier is close to the usual market price than far away. The mixture distribution satisfies all these conditions with the used parameters. The value with the greatest probability is about 546 € per sq.

m. and the probability is nearly zero to generate a value larger than 800 € per sq. m. The used mixture distribution has a shorter tail on the left side as on the right side. Thus the probability is higher to generate a random number closer to the normal distributed noise.

The generation of the random noise $\boldsymbol{\varepsilon}^{(j)}$ is realized as follow. From data set $\boldsymbol{D}_R^{(j)}$ 80% of the purchases are randomly selected. For these purchases random numbers are drawn from the normal distributed noise (Figure 3, blue line). The rest of the purchases obtain their noise from the mixture distribution. Hence, the outlier ratio δ of the noise model $\boldsymbol{\Psi}$ is between 20% and 25%. The random drawing from the mixture distributions is realized with the Acceptance-Rejection-Approach, for more details, refer to, e.g., Gelman et al. (2014), Koch (2007).

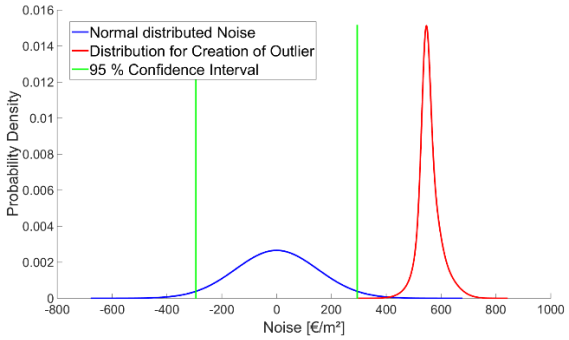

Figure 3: Probability density of the noise model $\boldsymbol{\Psi}$. Blue: Normal distributed noise of the purchase prices. Green: 95% confidence interval of the normal distributed noise. Red: Mixture distribution of three normal distributions for outlier creation.

Then, we check the noise model $\boldsymbol{\Psi}$ for plausibility. For this purpose 100 purchases are randomly drawn from data set $\boldsymbol{D}_A$ and simulated purchase prices are created as given in Subsection 3.1. Hence, the original purchase prices can be compared with the simulated purchase prices. Therefore, the regression coefficients are estimated by Eq. 2 and the predicted prices are determined for both data sets. The results are presented in Figure 4. The left figure shows the result of the original purchase prices and the right figure shows the result of the simulated purchase prices. Both results show similarities. In the low price segment the predicted prices (red crosses) are larger than the prices (blue line) and in the high price segment the predicted prices are smaller than the prices. However, the RMSE of the simulated prices is with about 250 € per sq. m. a little higher than about 200 € per sq. m. for the real purchases. This follows from the outliers in the simulated prices, which results in a larger spread than with the real prices. But the RMSE of the simulated purchases is not so large that the noise model can exist for a real estate market. We can conclude that the used noise model $\boldsymbol{\Psi}$ is realistic. In future studies alternative noise models should be used for the closed loop simulation. The models should be derived from future investigation of noise characteristics in areas with few transactions.

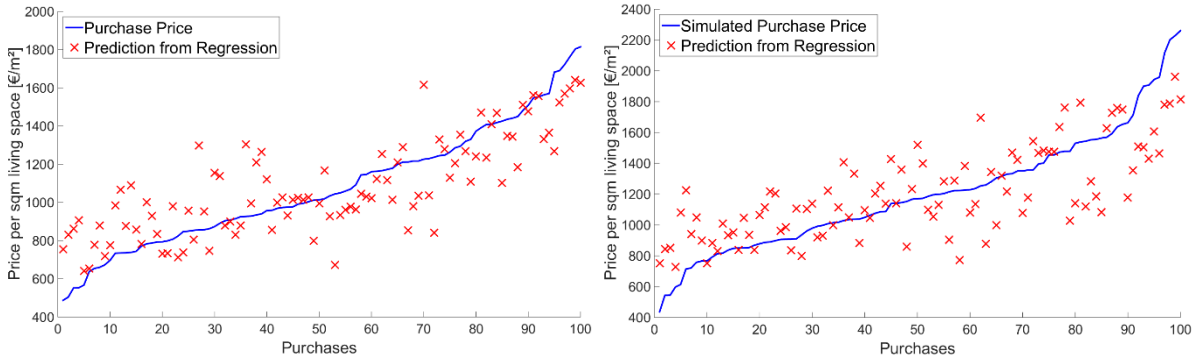

Figure 4: Comparison between original purchase prices and simulated purchase prices. On the left side: Prediction result of regression for 100 random purchases from $\boldsymbol{D_A}$. On the right side: Prediction result of regression for the same 100 purchases, whose prices are generated with the noise model.

4. RESULTS OF THE INVESTIGATION

The different estimation approaches of Section 2 are investigated with the presented closed loop simulation of Section 3. We simulate two different scenarios. The first scenario is a regular supply and demand situation on the market with $h = 100$ purchases. The second scenario is an area with few transactions with only 30 purchases. Furthermore, additional setting values for the Gibbs sampler and RANSAC are necessary. We found out that the length of warm-up period and the total number of samples in the Gibbs sampler are equal to $s_1 = 500$ and $s_2 = 5'000$; these estimates are derive from the whole repetitions of about 100 runs. The minimum number of iterations N for RANSAC can be calculated with Eq. 19. In this case RANSAC requires at least 24 iteration with an outlier ratio of about 25%. However, to ensure that a good model M is selected with a probability of 99%, a larger number of iteration is necessary. Hence, we increase the number of iteration to $N = 1000$. The error tolerance S is selected via the normal noise distribution of the noise model. We use the value 200 € per sq. m. for S, which is equal to the 80% confidence interval of the normal noise distribution. In the reality the error distribution is unknown. For this reason, the choice of the error tolerance with real data is not trivial in this case. The determination of the error tolerance for real data is investigated in future studies.

Table 2: Results of RMSE from the closed loop simulation.

	RMSE for 100 Purchases		RMSE for 30 Purchases	
	Mean [€/m²]	σ [€/m²]	Mean [€/m²]	σ [€/m²]
Reference	35.66	10.50	68.23	21.27
Regression	125.57	14.09	154.46	25.69
Data Snooping	70.36	34.83	158.77	57.04
Robust Bayesian Approach	97.83	18.92	142.58	32.26
RANSAC	75.25	31.71	163.44	89.10

The result of the closed loop (RMSE estimation) using 100 purchases is depicted in Figure 5. The histograms show the 100'000 calculated RMSE values for the different approaches. The mean value and the standard deviation are numerically derived from the distributions and presented in Table 2.

As expected, the RMSE result of reference (the optimal case without outliers in the measurements) has the smallest mean value and standard deviation. In contrast, the histogram of the regression approach shows the influence of the outliers on the estimated parameter. The mean value of regression is about 90 € per sq. m. larger than the mean value of the reference. But the spreading of both histograms is nearly identical. The result of the robust Bayesian approach is on average 25 € per sq. m. and smaller than the mean value of regression. Hence, the robust Bayesian approach affect the outliers by down weighting their influence on the resulting RMSE. Unfortunately, this effect is not large enough to make the Bayesian estimation better that the reference solution. The solution with the nearest mean value to reference is the solution by means of the data snooping approach. The RMSE is equal to approximately 70 € per sq. m. However, Figure 5 shows that two peak values exist within data snooping. The larger peak value is very close to the result of reference. The smaller peak value corresponds to the mean value of regression. In these cases, the data snooping fails and some of outliers are not detected or many good observation are wrongly deleted. The reason for this is the masking effect in which an outlier may show up in another residual or may hide behind another outlier. The long tail up to over 200 € per sq. m. of the data snooping shows the worst case which can happen by the masking effect. These results are worse than the RMSE of regression with outliers. The mean value of RANSAC is about 5 € per sq. m. larger than the mean value of data snooping, but the histogram of RANSAC has only one peak value. However, the histogram of RANSAC has a long tail like data snooping. The reason for this may be that the random choice of the u purchases isn't optimal or the used error tolerance S is too large. In contrast to data snooping and RANSAC, the robust Bayesian approach has no long tail.

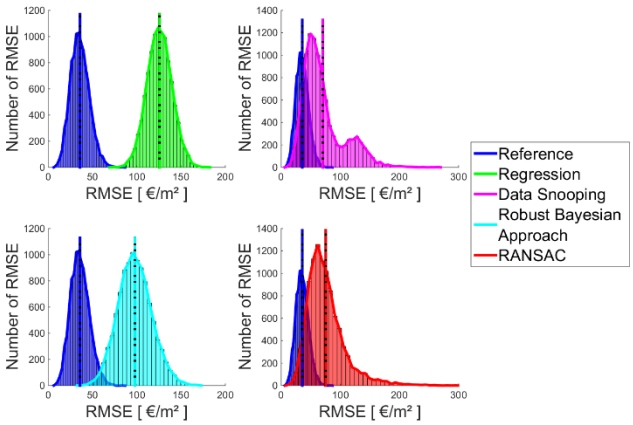

Figure 5: Histograms of RMSE values for the simulation with 100 simulated purchases and 100'000 iterations. The mean values of the histograms are shown as dashed line.

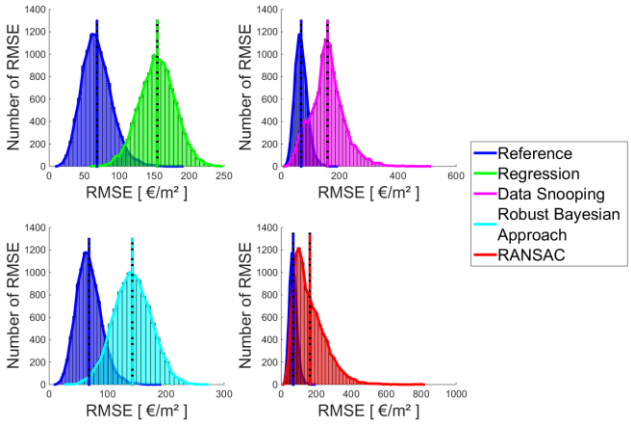

Figure 6: Histograms of RMSE values for the simulation with 30 simulated purchases and 100'000 iterations. The mean values of the histograms are shown as dashed line.

The simulation result of RMSE for the 30 purchases is depicted in Figure 6 and the determined moments of the histograms are presented in Table 2. The mean value and standard deviation of the reference are (as in case one) the smallest, but these values are twice larger than the values of the simulation with 100 purchases. The reason for this is smaller redundancy (with factor three) in the simulation with 30 purchases. This fact is also shown in the RMSE histogram of the reference, which has now a tail up to about 190 € per sq. m. The mean value and the standard deviation of the regression increase by about 10 € per sq. m.. Hence, the RMSE values of regression are closer to the RMSE values of reference in the simulation with 30 purchases. The mean value of the robust Bayesian approach is on average 10 € per sq. m. smaller than the regression result. This difference is smaller than the corresponding result of the simulation with 100 purchases. In the simulation with 30 purchases the mean value of robust Bayesian approach is thus not significantly better than the mean value of regression. In Figure 6 the RMSE histogram of data snooping has now one peak value at about 158 € per sq. m. From this result, it has become clear that the data snooping is not appropriate approach in case of only small database. The same result is shown also for RANSAC. In the simulation with 30 purchases RANSAC has the largest mean value and standard deviation. Furthermore, the histograms of data snooping and RANSAC have longer tails in the simulation with 30 purchases than in the simulation with 100 purchases. The results for the simulation with 30 purchases show, that all estimation approaches to deal with outliers lose their efficiency of the reliably detection of the outliers.

5. CONCLUSION AND OUTLOOK

In general, it can be noted for the close loop simulation that outlier detection approach (data snooping) is only reliable in data sets with an adequate number of purchases. The RANSAC is almost as good as the data snooping. However, due to masking effects, the failure probability to estimate reliable results is not negligible. Furthermore, the RANSAC need a suitable error tolerance S; its derivation from the real data is not an easy task. The optimal choice of the error tolerance for real estate data should be investigated in a future study. In contrast to the simulated submarket with 100 purchases the results of data snooping and RANSAC for the simulated submarket with 30 purchases are on average worse than the result of the regression. Hence, the failure probability in both approaches is relatively high in areas with few transactions. The results of the noninformative

robust Bayesian approach show that this approach isn't as efficient in outlier detection as data snooping and RANSAC in areas with an adequate number of transactions. In contrast, in areas with few transactions the robust Bayesian approach has in average better results than the other aforementioned two approaches. Furthermore, the robust Bayesian approach does not fail as often as data snooping or RANSAC. These simulation results show that the reliable outlier detection in data set with few observations is a challenging task. The robust Bayesian approach has the greatest potential to deal with outlier in areas with few transactions. In future studies the robust Bayesian approach should be improved to deal more efficient with outliers. Furthermore, the noise model of the closed loop simulation should be extended in future studies. For this purpose, the areas with few transactions should be investigated in more detail.

ACKNOWLEDGEMENT

The investigations published in this article are granted by the DFG (German Research Foundation) under the sign <60451047>. The authors cordially thank the funding Agency. Besides, we thank the surveying administration of Lower Saxony for the provided data.

REFERENCES

Alkhatib, H. & Weitkamp, A. (2012). Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung, Teil 1. ZfV, 137(02), pp. 93–102.

Alkhatib, H. & Weitkamp, A. (2013): Robust Bayesian Regression Approach for Areas with Small Numbers of Purchases. Royal Institution of Chartered Surveyors (Ed.): Proceedings of COBRA Conference, New Delhi.

Baarda, W. (1968). A testing procedure for use in geodetic networks. PhD thesis, Delft University of Technology, Publications on Geodesy, volume 2(5), Netherlands Geodetic Commission, Delft, Netherlands.

Fahrmeir, L., Kneib, T., Lang, S. & Marx, B. (2013). Regression: Models, Methods and Applications. Springer-Verlag Berlin, Heidelberg.

Fischler, M. A. & Bolles, R. C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM, Vol. 24, Issue 6, pp. 381-395.

Geweke, J. (1993). Bayesian Treatment of the Independent Student-t Linear Model. Journal of Applied Econometrics (8): 19–40.

Gelman, A. Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2014). Bayesian Data Analysis. Third Edition, CRC Press, Taylor & Francis Group, Boca Raton.

Hartley, R. & Zisserman, A. (2004). Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge.

Jäger, R., Müller, T., Saler, H. & Schwäble, R. (2005). Klassische und robuste Ausgleichungsverfahren. Ein Leitfaden für Ausbildung und Praxis von Geodäten und Geoinformatikern. Wichmann Verlag, Heidelberg.

Kleiber, W., Fischer, R., & Werling, U. (2014). Verkehrswertermittlung von Grundstücken: Kommentar und Handbuch zur Ermittlung von Marktwerten (Verkehrswerten) und Beleihungswerten sowie zur steuerlichen Bewertung unter Berücksichtigung der ImmoWertV. 7. vollst. neu bearb. Aufl., Bundesanzeiger Verlag, Köln.

Koch, K.-R. (1999). Parameter Estimation and Hypothesis Testing in Linear Models. 2. Edition, Springer-Verlag Berlin, Heidelberg.
Koch, K.-R. (2007). Introduction to Bayesian Statistics. 2. Edition, Springer-Verlag Berlin, Heidelberg.
Kroese, D. P., Taimre, T. & Botev. Z. I. (2011). Handbook of Monte Carlo Methods. John Wiley & Sons, Inc., Hoboken, New Jersey.
Pope, A. J. (1976). The Statistics of Residuals and the Detection of Outliers. NOAA Technical Report NOS65 NGS1, US Department of Commerce, National Geodetic Survey, Rockville, Maryland.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J. Gatelli, D., Saisana, M. & Tarantola, S. (2008). Global Sensitivity Analysis. The Primer. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England.
Soot, M., Weitkamp, A., Alkhatib, H., Dorndorf, A. & Jeschke, A. (2016). Analysis on Different Market Data for Real Estate Valuation – Investigations on German Real Estate Market. FIG Working Week 2016, Christchurch, New Zealand.
Weitkamp, A. & Alkhatib, H. (2012). Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung, Teil 2. ZfV, 137(02), pp. 103–114.
Weitkamp, A. & Alkhatib, H. (2014). Die Bewertung kaufpreisarmer Lagen mit multivariaten statistischen Verfahren: Möglichkeiten und Grenzen robuster Methoden bei der Auswertung weniger Kauffälle. AVN (Allgemeine Vermessungs-Nachrichten), 121(01), pp. 3–12.
Ziegenbein, W. (1977). Zur Anwendung multivarianter Verfahren der mathematischen Statistik in der Grundstückswertermittlung. PhD thesis, Technische Universität Hannover, Hannover.
Ziegenbein, W. (2010). Immobilienwertermittlung. In: Kummer, K. & Frankenberger, J. (Hrsg.): Das deutsche Vermessungs- und Geoinformationswesen 2010, Wichmann Verlag, Heidelberg, pp. 421-468.

BIOGRAPHICAL NOTES

Alexander **Dorndorf** received his Master of Science in "Geodesy and Geoinformatics" at the Leibniz Universität of Hannover in 2014. Since then he has been at the Geodetic Institute at the Leibniz Universität of Hannover. His main research interests are: Bayesian statistics, Monte Carlo simulation and modelling of measurement uncertainty.

Matthias **Soot** received his Master of Science (M.Sc.) in Geodesy at the Technische Universität Dresden in 2014. For half a year, he worked as a valuation expert in free economy. Since March 2015, he is working at the Geodetic Institute of the Technische Universität Dresden at Chair of Land Management. His research focus is on statistical analysis of market information and development of purchasing price databases.

Prof. Dr.-Ing. Alexandra **Weitkamp** received her diploma (Dipl.-Ing.) in Geodesy at the University of Hanover in 1999. She passed the highest-level state certification as "Graduate Civil Servant for Surveying and Real Estates" in Lower Saxony in 2001. After two-year experience at Bayer AG, she returns to Leibniz Universität Hannover. In 2008, she received her Ph.D. in "Geodesy and Geoinformatics" at the University of Bonn. Until 2014, she has been postdoctoral fellow at the Geodetic Institute at the Leibniz Universität Hannover. Since October 2014, she became Chair of

Land Management at Technische Universität Dresden. Her main research interests are: adaption of innovative evaluation methods for valuation, stakeholders in rural and urban development, and decision-making methods.

Dr. Hamza **Alkhatib** received his Dipl.-Ing. in Geodesy and Geoinformatics at the University of Karlsruhe in 2001 and his Ph.D. in Geodesy and Geoinformatics at the University of Bonn in 2007. Since 2007 he has been postdoctoral fellow at the Geodetic Institute at the Leibniz Universität of Hannover. His main research interests are: Bayesian statistics, Monte Carlo simulation, modeling of measurement uncertainty, filtering and prediction in state space models, and gravity field recovery via satellite geodesy.

CONTACTS

Alexander Dorndorf (M.Sc.) Geodetic Institute of Leibniz Universität Hannover – Evaluation methods Nienburger Str. 1 Hannover D – 30167 GERMANY Tel. +49 (0) 511-762 5194 Fax +49 (0) 511-762 2468 Email: dorndorf@gih.uni-hannover.de Web site: http://www.gih.uni-hannover.de/	Dr.-Ing. Hamza Alkhatib Geodetic Institute of Leibniz Universität Hannover – Evaluation methods Nienburger Str. 1 Hannover D – 30167 GERMANY Tel. +49 (0) 511-762 2464 Fax +49 (0) 511-762 2468 Email: alkhatib@gih.uni-hannover.de Web site: http://www.gih.uni-hannover.de/
Matthias Soot (M.Sc.) TU Dresden – Geodetic Institute Helmholzstraße 10 Dresden D - 01069 GERMANY Tel. +49 (0) 351-46331653 Fax + 49 (0) 351-46337190 Email: matthias.soot@tu-dresden.de Web site: http://www.tu-dresden.de/gi/lm	Prof. Dr.-Ing. Alexandra Weitkamp TU Dresden – Geodetic Institute Helmholzstraße 10 Dresden D - 01069 GERMANY Tel. +49 (0) 351-46337115 Fax + 49 (0) 351-46337190 Email: alexandra.weitkamp@tu-dresden.de Web site: http://www.tu-dresden.de/gi/lm