Propagating the Uncertainty of the Market Value by the use of a Bayesian Regression Approach

Sebastian ZADDACH and Hamza ALKHATIB, Germany

Key words: market value, uncertainty, sales comparison approach, Bayesian regression analysis

SUMMARY

The real estate and finance crisis has shown the importance of real estate valuation: The market value has to satisfy high objective requirements. Besides, the German jurisdiction demands a maximum dispersion of ± 20 % of the market value. The sales comparison approach as one of the valuation methods is from a mathematical-statistical point of view based on a multiple linear regression analysis. Since decades it has been considered as a standard procedure for analysing the real estate market and to determine the current market value. Nevertheless, the method has not been enhanced significantly since its introduction. The estimated comparative value is in particular depending on the number and the type of value influencing characteristics which are considered within the regression model.

The aim of this research is to enhance the use of regression analysis in real estate valuation by the use of a recursive Bayesian regression approach, which is able to consider the uncertainty of the value affecting characteristics as a prior information and thus to quantify its impact to the market value. Furthermore, it enables the propagation of uncertainty through subsequent periods of evaluation. For this purpose, prior information of the data has to be derived empirically from the data itself using empirical Bayes method. After initialization, the estimates of the regression coefficients and their uncertainty are used as a prior information for following analysis. Besides the propagation of uncertainty, the Bayesian regression models allow to predict comparative values, which dispose of a lower uncertainty than the results of the classical approach.The methodology is tested on a real data set. It is proven, that this approach provides more precise and appropriate uncertainty of predicted values and the use of information that are not yet included in the regression analysis.

Propagating theUncertainty of the Market Value by the use of a Bayesian Regression Approach

Sebastian ZADDACH and Hamza ALKHATIB, Germany

1. INTRODUCTION

In Germany, the Federal Building Code provides the main legal framework for the determination of market values for real estate. Subsidiary,statutoryregulations definethe individual, standardized methods for the determination of the market value. The focus of this contribution lies on the enhancement of the sales comparison approach, one of the prescribed methods, or rather its mathematical-statistical point of view. Applying the sales comparison approach, the market value of a real estate is derived from purchase prices of properties that match the valuation object in a sufficient manner concerning the main value-influencing characteristics. From the mathematical-statistical point of view the approach is based on hedonic price models, representing the most frequently applied models in regard to matters of real estat valuation. They have proven themselves capable for multifunctional purposes, amongst others, the determination of market values, in applications for mass appraisals and for the deduction of relevant data like index series or conversion factors. One of the hedonic models is the multiple linear regression analysis, which is highly attractive for valuation purposes because the results are directly interpretable in terms of the characteristics involved (Op't Veld,Bijlsma, and van de Hoef 2008). Since the 1980's, the regression analysisis considered as a standard method in Germany. Ziegenbein (1977)adopted the method for valuation purposes and it was put into practice by many expert committees in the following years.

The main idea of the regression analysis is the consideration as an iterative process, which results in the determination of an appropriate optimal regression function describing a specific issue of valuation, and in this case enables the estimation of comparative values.The estimated comparative valuesare dependent solely on the number and the type of value-influencingvariables, which are considered within the regression model. Besides the assumption, that the value-influencing characteristics have a linear effect on the comparative value, they also provide individual uncertainty, which are not be able to be considered within the regression equation. The Bayesian approach as an innovative method enables to quantify the uncertainty of single variables and to reduce it concerning the estimation of the parameters. In this way, the estimated parameters dispose of a higher certainty. Furthermore, the integration of Bayesian analysis into a sequential process leads to a recursive Bayesian regression approach that enables the propagation of uncertainty throughout subsequent periods of analysis. The main advantage of this process is that acurrent evaluation is not only based onthegiven data, but also depends on theresults of thepreviousperiods.This type of evaluationcorresponds to thereality of aconstantly changingmarket, whichof courseis not based onconstant boundsforanalysis periods.

Concerning real estate valuation, the Bayesian approach has already been adopted successfully: First attempts can be found in Alkhatib and Weitkamp (2012a), Weitkamp and Alkhatib (2012b), Alkhatib and Weitkamp (2012), Zaddach and Alkhatib (2013) and Weitkamp and Alkhatib (2014). The studies of the present contribution are a further development of the approaches presented in Zaddach and Alkhatib (2013a).

2. BAYESIAN REGRESSION APPROACH IN REAL ESTATE VALUATION

In general, Bayesian inferenceis based on the assumption, that all occurring uncertainty should be modeled by the use of probabilities and that resulting statistical inferences should be logical conclusions resulting from the laws of probability. It follows from this, that the underlying Bayes Theorem is defined as a process of fitting a certain probability model to a data set(Gelman et al. 2004). In contrast to the classical multiple regression analysis, the application of Bayesian inference enables to consider prior information about the main value-influencing variables and thus the parameters of the model. The consideration of prior knowledge is the main advantage compared to the classical regression approach. The following sections give a short insight into the theory of both the classical and the Bayesian approach.

2.1 Theory of Classical Multiple Regression Analysis

The regression analysis as an approximation method gives the possibility to explain variations of aresponse variable y (observation, e. g. purchase price per square meter) by the variability of $m$predictor variables $x_1, \dots ,x_m$,to indicate a functional relationship. The functional relationship is given by the unknown regression coefficients$\boldsymbol{\beta}$, resulting in the linear combination

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im} + \varepsilon_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} + \varepsilon_i, \quad [1]$$

where $i = 1, \dots , n$ with n = number of data sets and $\beta_0, \dots ,\beta_m$ indicate the m regression coefficient.The general trend model is overlain with random noise, the residual ε_i, since the relationship between the response variable and the predictor variables are often known only vaguely and knowledge about an exact function is not given. (Fahrmeir et al. 2009). The use of matrix-vector notation allows to rewrite the n equations in Eq. 1 to

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}. \quad [2]$$

Bold symbols indicate vectors and matrices, respectively. In the context of real estate valuation the residuals can be assumed as independent and identically distributed (i.i.d.) with $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I})$, where σ^2 is the unknown variance of unit weight and $\boldsymbol{I}$ is the identity matrix (weight matrix of the observations). All observations y_i are assumed to be independent from each other with equal variance. The common approach for estimating the unkown regression coefficients is the Ordinary Least Squares (OLS) technique: this method enables the estimation of the unknown parameters, which fit the sample data best in the specific sense of minimizing the sum of squared residuals (Fahrmeir et al. 2009). In this way, the regression coefficients and the corresponding cofactormatrix can be estimated as follows:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{y}\right), \quad Q_{\hat{\beta}\hat{\beta}} = \boldsymbol{V} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \quad .$$

[3]

Thevariance-covariance matrix (VCM) is then given by

$$\Sigma_{\hat{\beta}\hat{\beta}} = \hat{s}^2 Q_{\hat{\beta}\hat{\beta}} \qquad \text{with } \hat{s}^2 = \frac{\varepsilon'\varepsilon}{n-m}. \qquad [4]$$

If the whole model is statistically tested on its validity and the significance of the coefficients, the regression function can be used to predict the value of an object within the chosen submarket. A more detailed introduction to the theoretical background can be found in subject-specific or statistical literature (e. g. Ziegenbein 1977, Fahrmeir et al. 2009, Chatterjee 2013).

2.2 Theory of Bayesian Inference

As mentioned before, the main advantage of Bayesian inference lies in the fact, that prior information about the parameters of interest (here: regression coefficients) can be included in the mathematical model. In order to make probability statements about the unknown parameters given a collection of data, the Bayes theorem has to be evaluated:

$$p(\boldsymbol{\beta} \mid \boldsymbol{y}) \propto p(\boldsymbol{\beta})\, p(\boldsymbol{y} \mid \boldsymbol{\beta}). \qquad [5]$$

The *prior density* $p(\boldsymbol{\beta})$ contains the prior knowledge about the parameters and describes the current state of knowledge. The prior density can be distinguished by two types: On the one hand, the *informative* prior density is used if specific prior knowledge is available. On the other hand, the *non-informative* prior density is used if there is no knowledge about the parameters at all.As shown in the next section, the results of the Bayesian approach and the classical regression are identical if the deterministic context between the predictor variables and the response variable is linear and there is no prior knowledge, respectively. The term $p(\boldsymbol{y} \mid \boldsymbol{\beta})$ in Eq. 5, called*likelihood function*, is the probability density of the observations conditional on the model parameters. The likelihood function summarizes the information from real data (in this case the purchaseprices per m² of comparable objects) before knowledge about the parameters is collected. The term $p(\boldsymbol{\beta} \mid \boldsymbol{y})$ is proportional to the product of the prior density and the likelihood function and indicates the *posterior density*, the empirical estimated probability of $\boldsymbol{\beta}$conditional on $\boldsymbol{y}$. This density function can now be used, to receive estimates of the parameters and to predict new (comparable) values.In summary, the Bayes theorem starts by using prior probabilites to describe acurrent state of knowledge. It then incorporates information by the collection of new data which finally results in new posterior probabilities to describe the state of knowledge after combining the prior probabilities with the data. Gelman points out, that in "*(...)Bayesian statistics, all uncertainty and all information are incorporated through the use of probability distributions, and all conclusions obey the laws of probability theory*." (Gelman et al. 2004, p 18). For a more detailed introduction in Bayesian statistics, the reader is referred to standard literature (e. g. Gelman et al. 2004, Koch 2007).

2.3 Application of Bayesian Regression Approach

Concerning the Bayesian regression approach, the main interest lies in learning about the unknown, stochastic regression coefficients$\boldsymbol{\beta}$based on the data $\boldsymbol{y}$. The trend model based on the coefficientsthen reflects the goodness of fit between the real observations and the deterministic trend. For applications in real estate valuation, a normal distribution of the response variable as well as functional linear model is assumed. The evaluation of the Bayes theorem provide in this case an analytically solvable solution for the posterior density. The advantage of this is the avoidance of a computationally intensive derivation of the posterior density according to Eq. 5. Analytically not solvable solutions require numerical computations based on Monte Carlo techniques and will be an issue of future research as far as applications of real estate valuation are concerned. Under the assumption of a normal distribution for the response variable and the residuals as well as a general linear deterministic context within the model, Koch (2007) derives the mathematical procedure for estimating the coefficients. The main results are assembled in Table 1.

Table 1: Comparison of classical regression approach and Bayesian regression approach.

Classical regression approach	*Bayesian regression approach*	
$\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y})$	$\overline{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X} + \underline{\boldsymbol{V}}^{-1}\right)^{-1}(\boldsymbol{X}'\boldsymbol{y} + \underline{\boldsymbol{V}}^{-1}\underline{\boldsymbol{\beta}})$	[6]
$\boldsymbol{V} = \boldsymbol{Q}_{\beta\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$	$\overline{\boldsymbol{V}} = \left(\boldsymbol{X}'\boldsymbol{X} + \underline{\boldsymbol{V}}^{-1}\right)^{-1}$	[7]

The left column of Table 1 shows the classical regression approach, which is used in real estate valuation up to now. In contrast to this, the right column opposes the Bayesian parameter estimation based on the analytical solution. In these equations, available prior information is marked per underbar, resulting posterior information per overbar. As it can be seen, the main difference lies in the consideration of the uncertainty in the Bayesian approach, reflected by the cofactor matrix $\underline{\boldsymbol{V}}$ of the parameters as well as parameters themselves resulting from prior knowledge. By comparing both columns, it becomes obvious how prior information affects the estimation of posterior parameters as well as their uncertainty reflected by the VCM. In addition to the estimation of the parameters and their uncertainty, the posterior empirical variance factor can be calculated by

$$\bar{s}^2 = \left(n + 2\,\frac{\left(\underline{s}^2\right)^2}{V_{\sigma^2}} + 2\right)^{-1} \left(2\left(\frac{\left(\underline{s}^2\right)^2}{V_{\sigma^2}} + 1\right)\underline{s}^2 + \left(\underline{\boldsymbol{\beta}} - \overline{\boldsymbol{\beta}}\right)\underline{\boldsymbol{V}}^{-1}\left(\underline{\boldsymbol{\beta}} - \overline{\boldsymbol{\beta}}\right) + (\boldsymbol{y} - \boldsymbol{X}\overline{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\overline{\boldsymbol{\beta}})\right) \quad [8]$$

where $\underline{s}^2$ denotes the priori variance factor and V_{σ^2} its variance.For the whole deduction of the equations in Table 1 andEq.8the reader is referred to Koch (2007). The results can now be used to derive the confidence intervals (Highest Posterior Density Intervals, HPDI) of the parameters (assuming a Student's-t-distribution):

$$P\left(\bar{\beta}_i - t_{1-\alpha;f}\,/\bar{s}_{\bar{\beta}_i} < \beta_i < \bar{\beta}_i + t_{1-\alpha;f}/\bar{s}_{\bar{\beta}_i}\right) = 1 - \alpha. \quad [9]$$

In Eq. 9 the term$t_{1-\alpha;f}$ denotes the quantile of the Student's-t-distribution, α the significance level, f the degrees of freedom and $\bar{s}_{\bar{\beta}_i}$the standard deviation of the parameter resulting from Eq. 7. Once the parameters are estimated, the results allow to predict new, unobserved values $\boldsymbol{y}_p$ with value-influencing characteristics$\boldsymbol{X}_p$. For this purpose a so called *posterior predictive distribution* has to be formulated. Following Gelman et al. (2004, p. 358), this distribution $p(\boldsymbol{y}_p|\boldsymbol{y})$ contains two types of unvertainty: (1) the fundamental variability of the model, represented by the variance factor in $\boldsymbol{y}$ not accounted for by $\boldsymbol{X\beta}$, and (2) the posterior uncertainty of the parameters and the posterior variance factor due to the finite sample size of $\boldsymbol{y}$. The normal linear Bayesian regression model allows the analytical determination of the expectation value *E*(.) as well as the variance *var*(.) of the posterior predictive distribution:

$$\begin{aligned} E(\boldsymbol{y}_p|\bar{s}^2,\boldsymbol{y}) &= E\left(E(\boldsymbol{y_p}|\overline{\boldsymbol{\beta}},\bar{s}^2,\boldsymbol{y})|\bar{s}^2,\boldsymbol{y}\right) \\ &= E(\boldsymbol{X}_p\overline{\boldsymbol{\beta}}|\bar{s}^2,\boldsymbol{y}) \\ &= \boldsymbol{X}_p\overline{\boldsymbol{\beta}}, \end{aligned} \qquad [10]$$

$$\begin{aligned} var(\boldsymbol{y}_p|\bar{s}^2,\boldsymbol{y}) &= E\left(var(\boldsymbol{y_p}|\overline{\boldsymbol{\beta}},\bar{s}^2,\boldsymbol{y})|\bar{s}^2,\boldsymbol{y}\right) + var\left(E(\boldsymbol{y_p}|\overline{\boldsymbol{\beta}},\bar{s}^2,\boldsymbol{y})|\bar{s}^2,\boldsymbol{y}\right) \\ &= E(\bar{s}^2\boldsymbol{I}|\bar{s}^2,\boldsymbol{y}) + var(\boldsymbol{X}_p\overline{\boldsymbol{\beta}}|\bar{s}^2,\boldsymbol{y}) \\ &= \left(\boldsymbol{I}+\boldsymbol{X}_p\underline{\boldsymbol{V}}\boldsymbol{X}_{\boldsymbol{p}}'\right)\bar{s}^2. \end{aligned} \qquad [11]$$

Under the assumption of Student's-t-distribution the HPD-Intervals of predicted values can now easily calculated by:

$$P\left(\hat{y}_{p,i} - t_{1-\alpha;f}/\sqrt{var(\hat{y}_{p,i})} < y_{p,i} < \bar{\beta}_i + t_{1-\alpha;f}/\sqrt{var(\hat{y}_{p,i})}\right) = 1-\alpha. \qquad [12]$$

2.4 Recursive Bayesian Estimation

Referring to issues of real estate valuation, former classical regression results, which reflect the market behavior in one year,should influence the analysis of the next year. This point of view is justified, because the market, of course, does not depend on closed periods. Until now, most expert committees in Germany performa singleanalysisfor one year,without usinginformation fromrecent years, as far as the estimation of comparative values is concerned. Only few experts use overlapping periods, althoughthe behavior of the real estate marketis continuously changing and data of past periods may help to improve the regression function.The advantage of the informative Bayesian regression approach compared to the classical, non-informative one is the propagation of uncertainty of the parameters concerning data sets collected in different periods. In this way, the posterior information is used as a prior data set for subsequent periods. The way of gaining prior knowledge is not strictly specified. For this purpose, experts' opinion can be used as well as the results of former analysis. The first mentioned approach has been applied successfully to real estate valuation in Alkhatib and Weitkamp (2012), Weitkamp and Alkhatib (2012a) and Weitkamp and Alkhatib (2012b). The second approach is presented in the following.Figure 1 depicts the approach, which is suggested for recursive Bayesian filtering based on gaining prior knowledge from past analysis, i. e. data driven prior information (known as *empirical Bayes*).

The first step of the analysis starts with the definition of the spatial and objective submarket: The response variable as well as the predictor variables have to be defined. It is initially irrelevant, whether each predictor variable has an effect on the response variable or not, because the significance of the parameters is tested within the process. After defining the variables, prior information has to be generated by applying a non-informative (classical) regression approach to a data set of a past period. In this context, it is important to choose the same response and predictor variables as in the actual data set.

Once the prior information (cf. Eq. 6 and 7) is obtained, a first informative Bayesian regression approach can be applied. According to the last section, a Student's-t-distribution is assumed for the parameters.

The significance of the parameters of the model can then be tested by formulating the hypothesis

$$H_0:\ E(\bar{\beta}_i) = 0,\ \ H_A:\ E(\bar{\beta}_i) \neq 0, \qquad [13]$$

with the null hypothesis H_0 (parameter is not significant) and the alternative hypothesis H_A (parameter is significant) following the Student's-t-distribution with *n-m* degrees of freedom. The test size is given by

$$t_\beta = \frac{|\bar{\beta}_i|}{\bar{s}_{\bar{\beta}_i}} \qquad [14]$$

and the null hypothesis is rejected, if $t_\beta > t_{1-\alpha/2,f}$. In case of a non-significant parameter, the parameter has to be eliminated and the Bayesian regression model has to be calculated again. In this context, it is important that a re-calculation of the prior regression is also necessary, as new prior information has to be generated without consideration of the non-significant parameter. The whole process runs iteratively, until an optimal regression model is reached. Based on the results, the HPD-Intervals as well as the prediction of new values can be performed. Additionally, the posterior Bayesian regression model serves as a prior information for subsequent analysis.

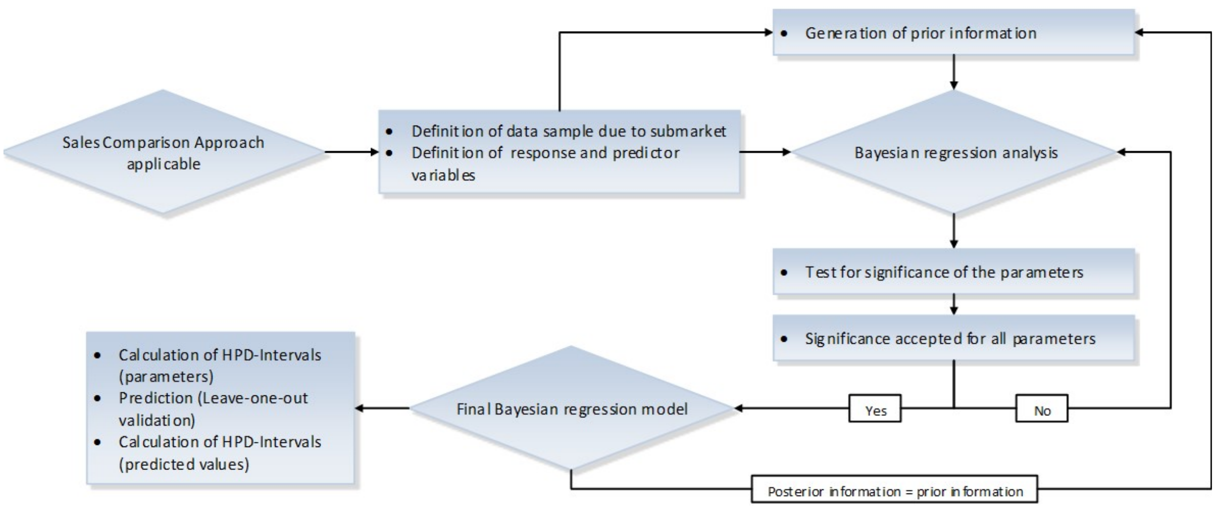

Figure 1: Process of recursive Bayesian regression approach.

3. PRACITCAL APPROACH

The presented research approach in Figure 1 is applied by means of real data set. Overall, three periods are used. In this case, one period accords to one year where the local expert committee collected purchase prices and their main value-influencing characteristics. The evaluation of the first year is performed by the use of a non-informative Bayesian regression approach, as no prior information is available for this data set. The results of this first year serve as a prior data set for the second year according to Figure 1. Thesustainingresults are thenusedfor the next calculation of the third year. In order to be able to predict new values, a *leave-one-out cross-validation* approach is applied. The basic idea of this method is to separate one purchase case of the original sample and to perform the estimation with the remaining data. The whole process is repeated, until every purchase case has been used for prediction (Cressie 1993).

3.1 Description of Submarket

The analysis are based on data sets of theAutomated Purchase Records (APR),a digital database containing all property transactions within the province of Lower Saxony, in the north of Germany.Due to the available data from the APR, the city of Hanover, capital of Lower Saxony, is selected as a spatial submarket. The objective submarket is set to condominiums. Based on former analysis concerning the significant value affecting variables, the comparative value (purchase price per m² (€/m²), respectively) can be described by six variables: standard ground value (€/m²), age (years), area of living space (m²), distance to the next park/recreation area (km) and the quality of the location (based on assessment of experts, categorial variable: good location, average location, bad location).

3.2 Recursice Bayesian Estimation

The data set is available for the years 2008 to 2010. In the first year of evaluation, 2008, 489 purchases are available, in 2009 132 purchsases and in 2010 184 purchases. The non-informative solution of the Bayesian approach is based on the evaluation of 2009. According to Table 1, the results are equal to the classical multiple regression analysis. The regression coefficients and their variance a s result of the analysis of the data set 2008 form the prior knowledge due to Eq. 6 and 7 ($\underline{\boldsymbol{\beta}}$ and$\underline{\boldsymbol{V}}$, respectively). The prior knowledge is estimated by the non-informative approach (Eq. 3 and 4), as the data set of 2008 can be regarded as zero point on a scale within this approach and it is assumed, that no further knowledge is available. Instead of the random choice of 2008, any other data set may initialize the procedure. The response variable is appointed to purchase price per area of living space (€/m²). The determination of an optimal regression function is based on the requirements that are applied in the context of the valuation standardsin Germany. These are not discussed in this paper (see e. g. Ziegenbein 1977).

3.3 Results of the Practical Approach

The results of the recursive Bayesian estimation due to Figure 1 are depicted in Table 3 concerning the evaluation of 2008 and 2009 and in Table 4 the propagation of the uncertainty from 2009 to 2010 is shown. As the focus of this contribution lies on the examination of the uncertainty and due to the lack of space, the presentation of the key figures of the regression analysis is omitted. In the first column of each table, the significant coefficients are listed. The next column, "*Prior informative*", shows the results of the prior regression analysis, accomplished with the data set of 2008 and 2009, respectively. The estimated coefficients for the "*non-informative prior*" solutioncan be found in the subsequent column. The last column, "*informative prior*", shows the results under consideration of the particular prior analysis. It can be seen, that the mean values of the regression coefficients of the informative solutions lie between the mean values of the prior information and those of the posterior based on non-informative prior. The different approaches lead to numerical different coefficients, which can be expected because of the modifications due to Eq. 6 and 7. Nevertheless, the tendency of the coefficents in regard to their prefix as well as their absolute numeric value indicates, that the chosen submarket of condominiums reflects a very homogeneous behaviour concerning the real estate market from 2008 to 2010. Furthermore, the improvement of the regression function is not the primary aim of the Bayesian approach, since the main focus of this method lies in decreasing the uncertainty of the single coefficients and its propagation to the predicted values.

Table 3: Prior and posterior estimates for the regression coefficients (2009).

		Prior	*Posterior based on*	
		Informative (2008)	*non-informative prior (2009)*	*informative prior (2009)*
Intercept	β_0	776.97	926.18	820.56
Standard ground value	β_1	9.65	2.37	8.67
Age	β_2	-3.05	-2.40	-3.36
Area of living space	β_3	5.43	6.00	5.49
Dist. Park	β_4	-13.15	-16.99	-14.69
Quality of location	β_5	-70.91	-51.33	-69.86

Table 4: Prior and posterior estimates for the regression coefficients (2010).

		Prior	*Posterior based on*	
		informative (2009)	*non-informative prior (2010)*	*informative prior (2010)*
Intercept	β_0	820.56	740.13	836.59
Standard ground value	β_1	8.67	15.12	9.13
Age	β_2	-3.36	-3.29	-3.33
Area of living space	β_3	5.49	5.40	5.50
Dist. Park	β_4	-14.69	-11.59	-13.17
Quality of location	β_5	-69.86	-74.49	-80.71

The marginal distributions of the coefficients because of the joint posterior density are depicted in Figure 2a and Figure 2b. Under the assumption of Student's-t-distribution each figure shows the distribution of one coefficient at a time as a result of the non-informative solution (dark grey color) and of the informative solution (bright grey color). Additionally, the mean values are plotted as well as the the 95%-HPD-Intervals (dashed-dotted line in case of non-informative solution, solid line in case of informative solution). The left columns in Figure 2a and 2b show the coefficients resulting from the analysis of 2009, the right columns those of the subsequent evaluation of 2010. In every case, it is obvious, that the informative solution is limited to a more narrow range in comparison to the non-informative solution. This corresponds to the fact that the use of prior information leads to a more concentrated result due to the associated mean value. The uncertainty, which is in statistics usually expressed by the variance, is reduced significantly. This conclusion is confirmed by the computation of the lower and higher boundaries of the HPD-Intervals according to Eq. 9. As expected, the boundaries have a more narrow range, if prior knowledge is integrated. Comparing the informative results from 2009 and 2010 it can be noticed, that numerical values are very similiar, whereas those of the non-informative solution are more different.This leads to the fact, that the use of prior informaion not only reduces the uncertainty but also provides a better stability of the functional model. The informative Bayesian regression approach is able to consider slight fluctuations of the market.

Figure 3 depicts the HPD-Intervals resulting for the cross validation approach. As explained in the beginning of chapter 3, 132 comparative values are predicted for the year 2009 and 184 for the year 2010. The first row shows the margins of the HPD-Intervals of the non-informative and the informative solution for the years of 2009 (left) and those for 2010 (right). The second row displays the change in % of the margins resulting from the first row, caused by the consideration of prior information. In both year, the consideration of prior information allows the reduction of the HPD-Intervals for the predicted values for about 100 €/m² and 150 €/m², respectivley. This corresponds to a mean improvment of about 12 % in 2009 and 11.6 % in 2010.

4. CONCLUSIONS

The presented method shows the implementation of a recursive Bayesian approach to real estate valuation. The researches have to be interpreted as a preparatory step for upcoming approaches concerning the Bayesian approach for real estate purposes. The outstanding finding of the presented research is the successful recursive integration of prior knowledge which has been derived empirically from former data sets. The presented results show, that the empirical Bayes method is applicable and provides regression coefficients with a significantly reduced uncertainty concerning the classical approach and the propagation of this uncertainty to new predicted comparative values. In this way, the impact of the uncertainty of the single value affecting variables on the market value can be reduced, as far as the mathematical model is concerned. Nevertheless, the results only show the general application of the Bayesian method, but the main benefits of the method are reached in cases, where only a few data sets are available. In these cases, the integration of prior knowledge about the parameters leads to estimates, which are reliable, whereas the use of the classical approach may fail, due to a lack of data. Furthermore, besides the improvement of the propagation of uncertainty, the functional model can be improved by applying a so-called Collocation approach. This approach can be regarded as an enhancement to the classical regression analysis and leads to predictions closer to the purchase price than those of the regression analysis can. This approach has already been tested for purposes of real estate valuation successfully (cf. Zaddach and Alkhatib 2012, Zaddach and Alkhatib 2013b, Zaddach and Alkhatib 2014). The combination of the Collocation approach and Bayesian Theoremallows the improvement of the functional model and its uncertainty and is the next step of applying Bayesian inference to the field of real estate valuation.

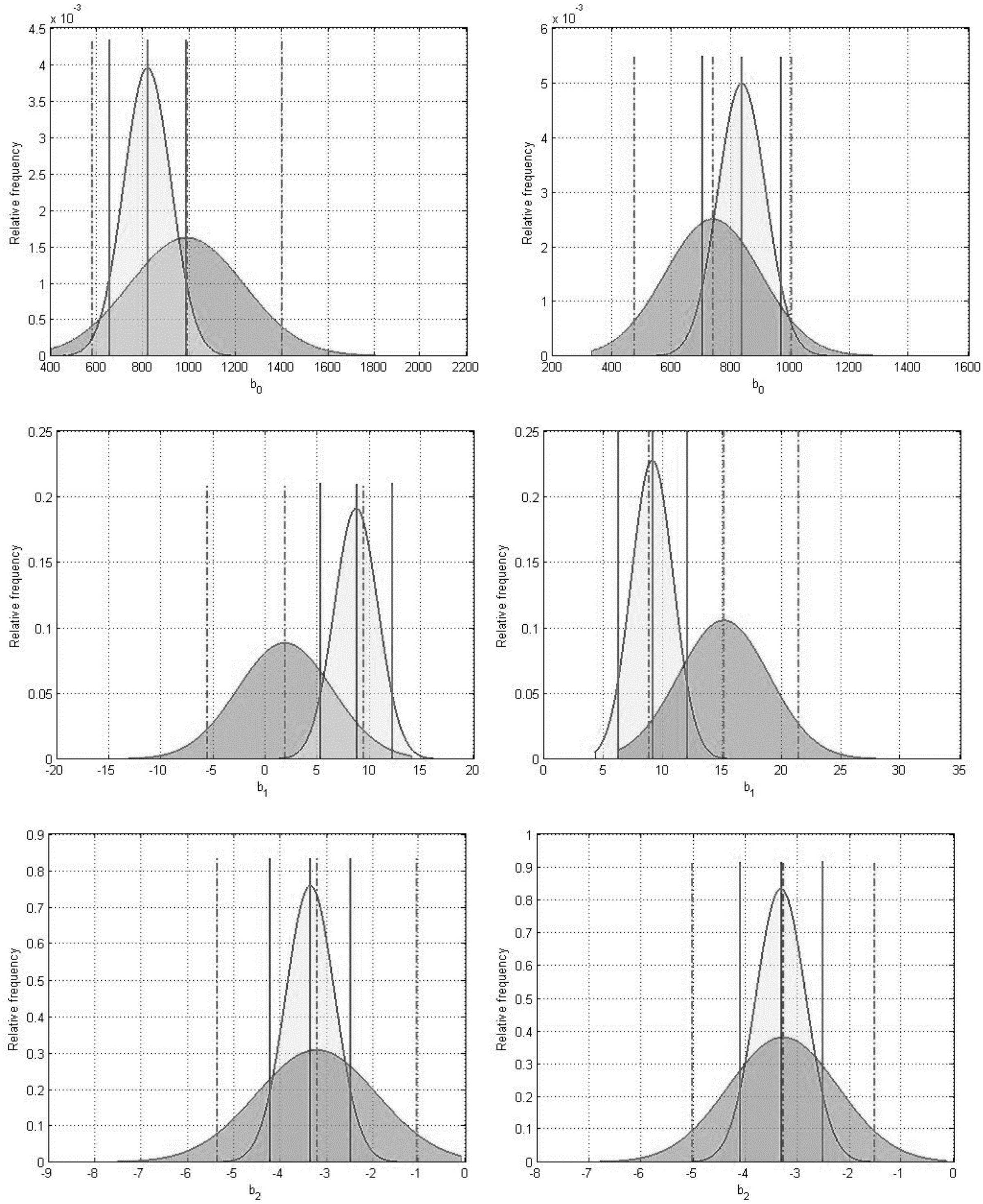

Figure 2a: Estimated Student's-t-distribution of the parameters β_0 toβ_2. The left column show the results of the informative Bayesian regression analysis based on the data of 2009, the right column those of 2010. Dashed-dot lines represent the 95%-confidence intervals and the mean value of the non-informative, solid lines those of the informative solution.

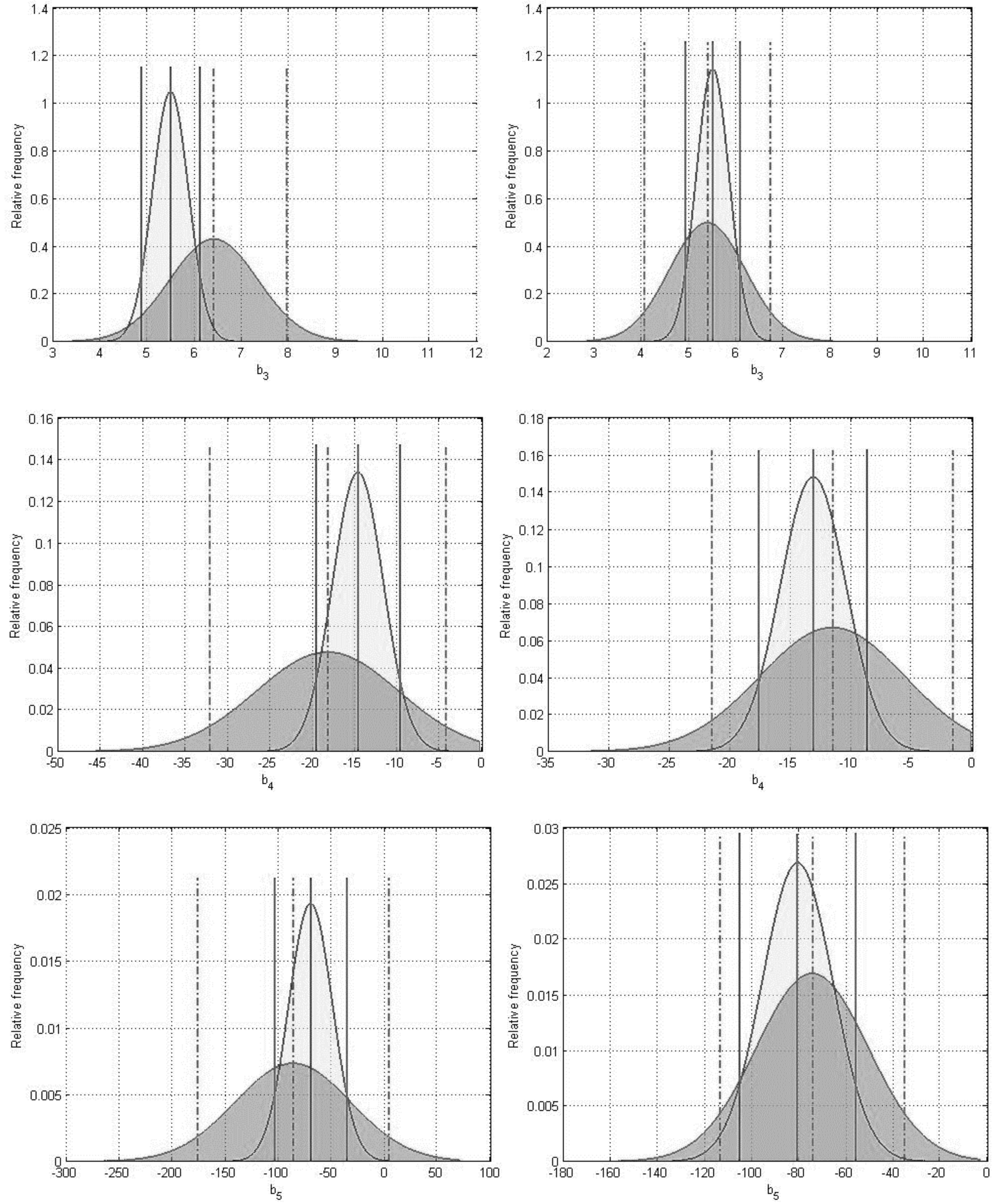

Figure 2b: Estimated Student's-t-distribution of the parameters β_3 toβ_5. The left column show the results of the informative Bayesian regression analysis based on the data of 2009, the right column those of 2010. Dashed-dot lines represent the 95%-confidence intervals and the mean value of the non-informative, solid lines those of the informative solution.

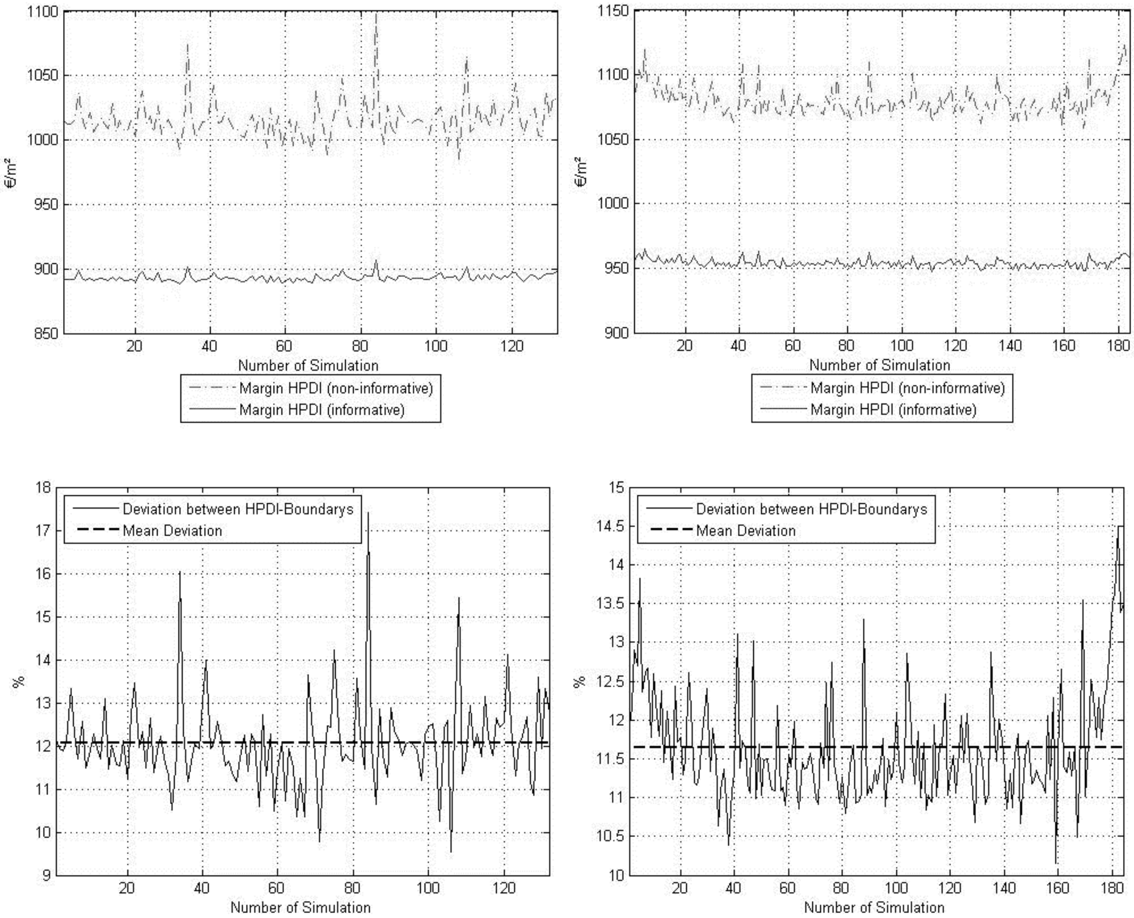

Figure 3: Comparison of the HPD-Intervals. The first row depicts the margins of the HPD-Intervals of the non-informative and the informative solution for the years of 2009 (left) and 2010 (right) for the predicted values. Second row depicts the change in % of the margins resulting from the first row, caused by the consideration of prior information.

REFERENCES

Alkhatib, H., Weitkamp, A. (2012*). Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung (Teil 1).zfv,* 137. Jg., 2/2012, pp. 93-102.

Alkhatib, H.; Weitkamp, A. (2013).*Robust Bayesian Regression Approach For Areas With Small Numbers Of Purchases.*In: *Proceedings of the RICS Cobra Research Conference*, September 10-12. 2013, New Delhi, India.

Chatterjee, S. (2013). *Handbook of regression analysis*. Wiley, A John Wiley & Sons, Inc., Publication: Hoboken, New Jersey.

Cressie, N. A. C. (1993).*Statistics for Spatial Data*. Revised Edition Auflage. New York, Chichester, Toronto, Brisbane, Singapore: John Wiley & Sons, Inc.

Fahrmeir, L., Kneib, T., Lang, S. (2009).*Regression. Modelle, Methoden und Anwendungen*. 2. Auflage. Heidelberg, Dordrecht, London, New York: Springer Verlag.

FBC (2013).Federal Building Code, in the version published on 23rd September 2004, as last amended by Article 1 of the Law of 11th June 2013 (Federal Law Gazette I, p. 1548).

Op't Veld, D. (2008). *Automated Valuation in the Dutch Housing Market: The Web-Application 'MarktPositie' Used by NVM-Realtors*. In: Kauko, T. and d'Amato, M.: *Mass Appraisal Methods – An international perspective for property valuers*. Wiley-Blackwell, John Wiley & Sons Ltd: The Atrium, Southern Gate, Chichester, West Sussex.

Weitkamp, A., Alkhatib, H. (2012a). *The Bayesian approach in the valuation – a strategy to handle markets with low purchasing prices?.* In: *FIG Publications, Proceedings of the FIG Working Week 2012*, May 6-10 2012, FIG, Rome, Italy.

Weitkamp, A., Alkhatib, H. (2012b*). Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung (Teil 2).zfv,* 137. Jg., 2/2012, 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*. In: AVN (Allgemeine Vermessungs-Nachrichten), Nr. 1/2014, p. 3-12.

Zaddach, S.; Alkhatib, H. (2012):*Use of Collocation Method as an Improved Sales Comparison Approach*. In: *FIG Publications, Proceedings of the FIG Working Week 2012*, May 6-10, 2012, FIG, Rome, Italy.

Zaddach, S., Alkhatib, H. (2013a). *Quantifying the impact of uncertainty to the market value by introducing a Bayesian sales comparison approach.* In: *Proceedings of the RICS Cobra Research Conference*, September 10-12. 2013, New Delhi, India.

Zaddach, S.; Alkhatib, H. (2013b).*Anwendung der Kollokation als erweitertes Vergleichswertverfahren in der Immobilienwertermittlung.* In: zfv (Zeitschrift für Geodäsie, Geoinformation und Landmanagement), Jg. 138, Nr. 2/2013, p. 144-153.

Zaddach, S.; Alkhatib, H. (2014). *Least Squares Collocation as an Enhancement to Multiple Regression Analysis in Mass Appraisal Applications.* In: Journal of Property Tax Assessment & Administration (Joint publication of the International Association of Assessing Officers and the International Property Tax Institute) (accepted).

Ziegenbein, W. (1977).*Zur Anwendung multivariater Verfahren der mathematischen Statistik in der Grundstückswertermittlung*. Dissertation, TechnischeUniversität Hannover, Hannover.

BIOGRAPHICAL NOTES

Dipl.-Ing Sebastian Zaddachgraduated in Geodesy and Geoinformatics at the Leibniz Universitaet Hannover in 2007. He passed the second state exam as "Graduate Civil Servant for Surveying and Real Estates" in Lower Saxony in 2010. Since 2010 he has been working as a scientific assistant at the Geodetic Instituteat the Leibniz Universitaet Hannover. His main research interests are:improvement of real estate valuation based on new statistical approaches, evaluation of uncertainty in real estate valuation.

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 Universitaet 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

Dipl.-Ing. Sebastian Zaddach
Leibniz Universitaet Hannover
Geodetic Institute Hanover
Nienburger Straße 1
30167 Hanover
GERMANY
Tel. +49 511 762 17201
Fax +49 511 762 2468
Email: zaddach@gih.uni-hannover.de
Web site: www.gih.uni-hannover.de

Dr.-Ing. Hamza Alkhatib
Leibniz Universitaet Hannover
Geodetic Institute Hanover
Nienburger Straße 1
30167 Hanover
GERMANY
Tel. +49 511 762 2464
Fax +49 511 762 2468
Email: alkhatib@gih.uni-hannover.de
Web site: www.gih.uni-hannover.de