

Aggregation und Gewichtung von unterschiedlichen Wertermittlungsdaten in kaufpreisarmen Lagen mittels Varianzkomponentenschätzung

Aggregation and Weighting Technique of Different Real Estate Valuation Data in Regions with Few Transactions by Means of Variance Component Estimation

Alexander Dorndorf, Matthias Soot, Alexandra Weitkamp, Hamza Alkhatib

Für die Immobilienwertermittlung in Deutschland werden drei normierte Verfahren genutzt, wobei das Vergleichswertverfahren als marktnächstes Verfahren anzusehen ist. Dieses nutzt regelmäßig die Regressionsanalyse zur Lösung des funktionalen Zusammenhangs. Hierbei ist es notwendig, dass pro modellierter Einflussgröße ca. 15 Kaufpreise vorliegen. Dies führt insbesondere in kaufpreisarmen Lagen zu Problemen, in denen selbst bei einer weiten Fassung des Gebiets nur zehn bis 30 Kauffälle pro Auswertung für alle Einflussgrößen zur Verfügung stehen. Als Lösungsansatz wird in diesem Beitrag der aktuelle Entwicklungsstand eines mathematisch-statistischen Modells zur Berechnung von zuverlässigen Regressionskoeffizienten in kaufpreisarmen Lagen vorgestellt. Dieses Ziel erreicht der vorgestellte Ansatz durch die Verwendung von zusätzlichen Marktdaten, die mit den vorhandenen Kauffällen kombiniert werden. Herausforderungen dabei sind die Akquirierung der zusätzlichen Marktdaten und die optimale Kombination der Datensätze im Ausgleichsmodell. In diesem Beitrag werden als zusätzliche Marktdaten eine Expertenbefragung sowie Angebotsdaten verwendet. Die Kombination mit den Kauffällen erfolgt über die Varianzkomponentenschätzung (VKS). Die Qualitätsuntersuchung der verschiedenen Marktdaten zeigt eine gute Übereinstimmung zwischen den Kauffällen und den Befragungsdaten. Die Angebotsdaten unterscheiden sich von den anderen beiden Datensätzen. Die Untersuchung der VKS zur Gewichtung der verschiedenen Datensätze im Ausgleichsmodell führt zu zwei Aussagen: Mit der VKS werden die Regressionskoeffizienten präziser geschätzt, aber dies ermöglicht keine bessere Prädiktion der Verkehrswerte.

Schlüsselwörter: Kaufpreisarme Lage, Expertenwissen, Regressionsanalyse, Varianzkomponentenschätzung

In Germany, real estate valuation is realized by three standardized methods, among which the approach based on sales comparison is considered as being closest to the real estate market. With this approach, regression analysis is frequently used to solve the linear relationship. The regression model, however, normally requires approximately 15 purchases per independent variable for an accurate estimate in real estate valuation. This causes a problem for areas with few transactions in which only 10 to 30 purchases are available. To solve this problem, the current paper presents a mathematical-statistical model for an

accurate estimation of regression coefficients in areas with few transactions. This aim is realized by using additional market data, which are combined with the available purchase cases. Challenges with this approach are the acquisition of the additional market data and the optimal combination of the data in the adjustment model. In this paper, survey data from real estate experts and offer prices are used as additional market data. The combination of these data with the purchase cases is carried out by means of variance component estimation (VCE). The quality assessment of the various market data shows a good conformity between the purchase cases and the experts' survey data. The offer prices differ from the other two datasets. The investigation concerning the VCE for the weighting of the different datasets within the adjustment model has two results: VCE yields more precise estimates of the regression coefficients, but this does not allow for a better prediction of the market values.

Keywords: *Regions with few transactions, expert knowledge, regression analysis, variance component estimation*

1 EINLEITUNG

Die zuverlässige Bestimmung von Immobilienwerten ist für eine Vielzahl von Aufgaben, z. B. bei der Ermittlung von Beleihungswerten, unerlässlich. Eine Fehleinschätzung des Immobilienmarkts kann zu negativen Ereignissen führen. Das Resultat einer solchen Fehleinschätzung wird im Rückblick auf die Immobilienmarktkrise in den USA im Jahr 2008 deutlich. Eine Ursache für die unzuverlässige Einschätzung des Immobilienmarkts ist beispielsweise fehlende Markttransparenz. Abhilfe kann durch gute Marktdaten geschaffen werden. Allerdings ist für die zuverlässige Ermittlung von Immobilienwerten eine ausreichende Anzahl von Transaktionen im gewöhnlichen Geschäftsverkehr am Immobilienmarkt notwendig, was zu einer Herausforderung in kaufpreisarmeren Lagen führt. Kaufpreisarmeren Lagen existieren z. B. in der Innenstadtlage für den Teilmarkt der unbebauten Grundstücke oder in peripheren ländlichen Räumen für fast alle Teilmärkte.

Um dieser Herausforderung zu begegnen, wird in diesem Beitrag der aktuelle Entwicklungsstand eines Ansatzes für eine zuverlässige Immobilienbewertung in diesen Lagen mittels multipler linearer Regression vorgestellt. Für eine akkurate Schätzung der Regressionskoeffizienten benötigt das Regressionsmodell normalerweise ca. 15 Kauffälle pro unabhängiger Variable /Ziegenbein 2010/, /Kleiber et al. 2014/. Somit sind in kaufpreisarmeren Lagen, in denen ca. zehn bis 30 Transaktionen getätigt werden, nicht genügend Vergleichsfälle für eine zuverlässige Regressionsanalyse mit zumeist vier bis fünf wertbeeinflussenden Einflussgrößen vorhanden. Zusätzlich zu der geringen Anzahl von Transaktionen am Markt ergibt sich die Herausforderung der Detektion von Ausreißern in den vorhandenen Kauffällen. Dies ist aufgrund der wenigen Beobachtungen mit statistischen Ansätzen nicht zuverlässig zu realisieren, weil die Redundanz in dem Ausgleichsmodell klein ist. Selbst bei erfolgreicher Detektion würde dies eine Reduzierung der sowieso schon geringen Anzahl an Beobachtungen zur Folge haben. Dadurch wird die Redundanz im Modell weiter verringert, was zu einer nicht lösbaren Aufgabe für die zuverlässige Schätzung der Regressionskoeffizienten führen kann.

Zurzeit erfolgt die Immobilienbewertung in kaufpreisarmeren Lagen vornehmlich durch den gutachterlichen Sachverstand: Die wenigen vorhandenen Kauffälle fließen allerdings nur indirekt (oft subjektiv)

über den Gutachter mit in die Bewertung ein, wobei zumeist kein einheitliches methodisches Vorgehen erfolgt. Zur Reduzierung der individuellen Wahrnehmung der Experten adaptierte /Reuter 2006/ die Delphi-Methode, um einen intersubjektiven Preisvergleich mit den Expertisen der Gutachter durchzuführen. Erste statistische Ansätze, welche das Expertenwissen mit vorhandenen Kauffällen mittels bayesscher multipler Regression kombinieren, wurden in /Weitkamp & Alkhatib 2012a/ und /Alkhatib & Weitkamp 2012/ veröffentlicht. In /Alkhatib & Weitkamp 2013/ und /Weitkamp & Alkhatib 2014/ wird ein robuster Ansatz für das bayessche Regressionsmodell vorgestellt, welcher den Einfluss von mit Ausreißern behafteten Kauffällen auf das Schätzergebnis reduziert. Die Robustheit wird in diesem Modellansatz durch die Verwendung der Student-Verteilung anstatt der Normalverteilung realisiert. Als Vorwissen verwendeten /Weitkamp & Alkhatib 2014/ für das bayessche Regressionsmodell Gutachten und Ergebnisse einer durchgeführten Befragung des Gutachterausschusses Osnabrück. Allerdings erfolgte die Kombination der Kauffälle in diesem Ansatz mit nur einer zusätzlichen Datenquelle.

In diesem Beitrag liegt deswegen der Fokus auf der Verwendung verschiedener Marktdaten im selben Regressionsmodell zur besseren Stützung der wenigen Kauffälle. Als zusätzliche Datenquellen werden die Ergebnisse einer durchgeführten Befragung des Gutachterausschusses Nienburg sowie Angebotsdaten von der Immobilienplattform Immobilienscout24 der Firma Immobilien Scout GmbH verwendet. Die Herausforderung ist die untereinander existierende Heterogenität zwischen den verschiedenen Datensätzen, die im Regressionsmodell mitberücksichtigt werden muss. Ein für diesen Zweck entwickelter statistischer Ansatz ist die Varianzkomponentenschätzung (VKS), die zur optimalen Kombination von Datensätzen mit unterschiedlich aufgenommenen Messinstrumenten in anderen geodätischen Teildisziplinen (z. B. Ingenieurgeodäsie) schon seit Längerem verwendet wird. Zur Beurteilung, ob die VKS auch für die Kombination der verschiedenen Marktdaten in einer kaufpreisarmeren Lage geeignet ist, wird dieser Ansatz an echten Marktdaten untersucht.

2 MATHEMATISCHE GRUNDLAGEN

2.1 Einsatz der multiplen linearen Regression in der Wertermittlung

Seit Jahrzehnten wird die multiple lineare Regression in der Wertermittlung für das Vergleichswertverfahren eingesetzt /Ziegenbein 1977/, /Pelzer 1978/. In diesem Modell erklären die beschreibenden Merkmale einer Immobilie (z. B. die Wohnfläche oder der Bodenrichtwert) ihren Kaufpreis. Daraus ergibt sich das folgende funktionale Modell:

$$y_i = \beta_0 + x_{i,1}\beta_1 + \dots + x_{i,u}\beta_u + e_i; \\ i = 1, \dots, n; e_i \sim N(0, \sigma^2). \quad (1)$$

Die abhängige Variable y_i (hier der Kaufpreis pro m^2 Wohnfläche) wird für n beobachtete Kauffälle durch eine Linearkombination von unabhängigen Variablen $x_{i,1}, \dots, x_{i,u}$ und den $u + 1$ unbekannten Regressionskoeffizienten β beschrieben. In der Regel werden die unbekannten Regressionskoeffizienten mit der Methode der kleinsten Quadrate geschätzt /Fahrmeir et al. 2009/, /Koch 1997/. Die Residuen e_i entsprechen den Differenzen zwischen dem Modell und den Beobachtungen, die aufgrund der Unvollkommenheit zwischen Modell und Realität entstehen. Für eine korrekte Schätzung der Regressionskoeffizienten wird in der Regel angenommen, dass die Residuen der Normalverteilung mit dem Erwartungswert 0 und der Varianz σ^2 folgen. Die Autoren verweisen für eine ausführlichere Diskussion der Regressionsanalyse auf die Fachliteratur, z. B. /Fahrmeir et al. 2009/ oder /Urban & Mayerl 2011/.

2.2 Bayesscher Ansatz für die Regressionsanalyse

In kaufpreisarmen Lagen ist die Anzahl verfügbarer Kauffälle für eine zuverlässige Regressionsanalyse zu gering. Aus diesem Grund werden weitere Beobachtungen benötigt. Diese können aus anderen Datenquellen, z. B. aus der Marktexpertise von Experten, akquiriert werden. Ein mathematischer Ansatz, der die Kombination dieser verschiedenen Daten ermöglicht, ist das Bayes-Theorem:

$$p(\bar{\beta} | y) \propto p(\bar{\beta}) \cdot p(y | \bar{\beta}). \quad (2)$$

In dieser Gleichung wird $p(\bar{\beta} | y)$ als Posteriori-Dichte bezeichnet. Sie enthält die unbekannten Regressionskoeffizienten in Abhängigkeit von den Beobachtungen y . Die Posteriori-Dichte ist proportional zu dem Produkt der Priori-Dichte $p(\bar{\beta})$ und der Likelihood-Funktion $p(y | \bar{\beta})$. Hierbei repräsentiert die Likelihood-Funktion die Informationen über die Kauffälle am Markt. Alle zusätzlichen Marktinformationen werden in der Priori-Dichte formuliert.

Die bayessche Inferenz unterscheidet zwischen konjugierten und nichtkonjugierten Priori-Dichten /Koch 2007/. Bei einer konjugierten Priori-Dichte folgt die Posteriori-Dichte der Verteilungsfamilie der Priori-Dichte. Dies hat den Vorteil, dass die Posteriori-Dichte analytisch gelöst werden kann. Für Gl. (1) ergibt sich bei einer konjugierten Priori-Dichte der folgende Schätzer:

$$\bar{\beta} = (X^T X + V^{-1})^{-1} (X^T y + V^{-1} \underline{\beta}). \quad (3)$$

Gl. (3) enthält zusätzlich zu Gl. (1) den Vektor $\underline{\beta}$ mit dem Vorwissen über die Regressionskoeffizienten und die zugehörige Kofaktormatrix V der Koeffizienten. Die Posteriori-Regressionskoeffizienten $\bar{\beta}$ können somit analytisch geschätzt werden.

Im Gegensatz hierzu muss bei nichtkonjugierten Priori-Dichten die Posteriori-Dichte mittels Monte-Carlo-Methoden numerisch berechnet werden. Dies ist z. B. in dem von /Alkhatib & Weitkamp 2013/ und /Weitkamp & Alkhatib 2014/ vorgestellten robusten bayesschen Regressionsansatz der Fall. Durch die Einführung der Student-Verteilung für die robuste Schätzung der Regressionskoeffizienten aus den Kauffällen entsteht eine nichtkonjugierte Priori-Dichte. Die numerische Lösung der Posteriori-Dichte erfolgt durch einen Gibbs-Sampler, der die unbekannte Verteilung der Regressionskoeffizienten approximiert. Somit ist das Lösen einer nichtkonjugierten Priori-Dichte komplexer und aufwendiger als das Lösen einer konjugierten Priori-Dichte. Detailliertere mathematische Grundlagen über die bayessche Inferenz sind z. B. in /Koch 2007/ oder /Kacker & Jones 2003/ zu finden. Konkrete Anwendungen der bayesschen Regressionsanalyse für Wertermittlungsaufgaben werden z. B. in /Weitkamp & Alkhatib 2012a/ und /Zaddach 2016/ behandelt.

2.3 Varianzkomponentenschätzung

Wie im Abschnitt zuvor erläutert, ist es mit dem Bayes-Theorem möglich, die vorhandenen Kauffälle einer kaufpreisarmen Lage mit weiteren Datenquellen zu kombinieren. Zugunsten der in der Wertermittlung gängigen Notationen für die Regressionsanalyse wird in diesem Abschnitt auf die bayesschen Notationen für die VKS verzichtet, um die Praxisnähe zur Immobilienbewertung zu erleichtern. Die bayesschen Notationen für die nachfolgenden Gleichungen werden z. B. in /Koch 2007/ vorgestellt.

Bei der Kombination der verschiedenen Datensätze ergibt sich die Herausforderung, dass sie heterogen sind. Dies kann sich beispielsweise an den Posteriori-Varianzfaktoren der Datensätze zeigen, wenn diese sich signifikant unterscheiden. Der Varianzfaktor wird aus der Residuenquadratsumme und der Redundanz berechnet (äquivalent zu Gl. (9)). Er beschreibt somit, wie präzise die geschätzten Regressionskoeffizienten die Beobachtungen des Datensatzes im Durchschnitt erklären. Eine ausführlichere Diskussion der Unterschiede zwischen den verwendeten Datensätzen erfolgt in Abschnitt 4.4.

Um die Heterogenität zwischen den Datensätzen zu berücksichtigen, können diese im Ausgleichsmodell unterschiedlich gewichtet werden. Für die Bestimmung der Gewichte existieren verschiedene Ansätze. Sie können z. B. aus Erfahrungswerten von früheren Experimenten abgeleitet oder mit statistischen Ansätzen berechnet werden. Um eine unterschiedliche Gewichtung der Datensätze im mathematischen Ansatz von Gl. (3) mit zu berücksichtigen, wird die Gleichung wie folgt erweitert:

$$\bar{\beta} = \left(\frac{1}{\sigma_L^2} X^T X + \frac{1}{\sigma_P^2} V^{-1} \right)^{-1} \left(\frac{1}{\sigma_L^2} X^T y + \frac{1}{\sigma_P^2} V^{-1} \underline{\beta} \right). \quad (4)$$

In die Gleichung werden die Variablen σ_L^2 und σ_P^2 als unbekannte Varianzkomponenten¹ für die Likelihood-Funktion und die Priori-

Dichte eingeführt. Die Gewichtung funktioniert nun wie folgt: Ist $\sigma_i^2 < \sigma_p^2$, hat die Likelihood-Funktion aufgrund der Division durch eins einen größeren Vorfaktor als die Priori-Dichte. Dadurch haben die Likelihood-Daten einen größeren Einfluss auf die Schätzung der Posteriori-Dichte als die Daten der Priori-Dichte. Den Likelihood-Daten wird somit in der Schätzung mehr vertraut bzw. die Priori-Daten werden herabgewichtet. Bei der Verwendung von mehreren Datensätzen als Priori-Information benötigt jeder Datensatz eine separate Varianzkomponente. Werden konjugierte Priori-Dichten verwendet, so kann dies durch die Erweiterung von Gl. (4) zu Gl. (5) erreicht werden:

$$\bar{\beta} = \left(\frac{1}{\sigma_1^2} \mathbf{X}_1^T \mathbf{X}_1 + \dots + \frac{1}{\sigma_i^2} \mathbf{X}_i^T \mathbf{X}_i \right)^{-1} \left(\frac{1}{\sigma_1^2} \mathbf{X}_1^T \mathbf{y}_1 + \dots + \frac{1}{\sigma_i^2} \mathbf{X}_i^T \mathbf{y}_i \right),$$

$$i = 1, \dots, k. \quad (5)$$

In Gl. (5) werden die Priori-Reggressionskoeffizienten und die zugehörige Kofaktormatrix durch die originären Beobachtungen $\mathbf{y}_i$ von den Priori-Daten und den zugehörigen unabhängigen Einflussgrößen $\mathbf{X}_i$ ersetzt, wodurch eine Kombination von k verschiedenen Datensätzen mit unterschiedlichen Gewichten im Modell ermöglicht wird. Die Bestimmung der unbekannten Varianzkomponenten σ_i^2 kann mit dem statistischen Ansatz der VKS durchgeführt werden /Koch 2007, S. 139 ff./ Diese wird in anderen geodätischen Disziplinen seit Jahren erfolgreich eingesetzt, z.B. in der Satellitengeodäsie zur Bestimmung des Geopotenzials oder für die Ausgleichung von geodätischen Netzen /Koch & Kusche 2002/, /Yang et al. 2005/. In der Literatur werden für die VKS verschiedene Schätzprinzipien vorgeschlagen, wie beispielsweise die Helmert-Methode oder die Maximum-Likelihood-Schätzung. Weiterführende Informationen hierzu sind z.B. in /Amiri-Simkooei 2007/ zu finden.

Der in diesem Beitrag verwendete Ansatz zur Schätzung der Varianzkomponenten erfolgt

nach dem in /Koch & Kusche 2002/ vorgestellten Algorithmus (Abb. 1). Als Input werden die Designmatrizen und die Beobachtungen der verschiedenen Datensätze benötigt. Da die geschätzten Regressionskoeffizienten von den Varianzkomponenten der Datensätze abhängig sind, müssen die Koeffizienten und die Varianzkomponenten in zwei separaten Schritten iterativ bis zu ihrer Konvergenz geschätzt werden. Der in Abb. 1 hellrot markierte Pseudocode stellt die Schätzung der Regressionskoeffizienten dar. Dieser Teil des Algorithmus ist nahezu äquivalent mit der multiplen linearen Regression: Lediglich Gl. (5) wird anstelle von Gl. (1) zur Schätzung der Koeffizienten verwendet.

Die Schätzung der Varianzkomponenten in Abhängigkeit von den Regressionskoeffizienten wird im hellblau markierten Abschnitt von Abb. 1 dargestellt. Als Startwerte für die VKS werden die Residuen $\hat{e}_i$ für die Datensätze 1 – k aus Gl. (6) benötigt und die

Daten: Designmatrix $\mathbf{X}_i$
Beobachtungsvektor $\mathbf{y}_i$

Ergebnisse: Varianzkomponenten $\hat{\sigma}_i^2$
Regressionskoeffizienten $\bar{\beta}$

Startwerte: $j = 0$
 $\hat{\sigma}_i^{2(j=0)} = 1$
 $\hat{\sigma}_i^{2(j=1)} = 0$

mit $i = 0, \dots, k$

while $\max(|\hat{\sigma}_i^{2(j+1)} - \hat{\sigma}_i^{2(j)}|) > 10^{-5}$ **do**

Startwerte: $\bar{\beta}$ mit Gl. (5) und $\hat{\sigma}_i^{2(j)}$

$\hat{e}_i = \mathbf{X}_i \bar{\beta} - \mathbf{y}_i$ Gl. (6)

$t = 0$

$\hat{\sigma}_i^{2(t=0)} = 1$

$\hat{\sigma}_i^{2(t=1)} = 0$

while $\max(|\hat{\sigma}_i^{2(t+1)} - \hat{\sigma}_i^{2(t)}|) > 10^{-5}$ **do**

$\bar{\mathbf{V}}^{(t+1)} = \left(\frac{1}{\hat{\sigma}_1^{2(t)}} \mathbf{X}_1^T \mathbf{X}_1 + \frac{1}{\hat{\sigma}_2^{2(t)}} \mathbf{X}_2^T \mathbf{X}_2 + \dots + \frac{1}{\hat{\sigma}_k^{2(t)}} \mathbf{X}_k^T \mathbf{X}_k \right)^{-1}$ Gl. (7)

$r_i^{(t+1)} = n_i - \text{spur} \left(\frac{1}{\hat{\sigma}_i^{2(t)}} \mathbf{X}_i^T \mathbf{X}_i \bar{\mathbf{V}}^{(t+1)} \right)$ Gl. (8)

$\hat{\sigma}_i^{2(t+1)} = \frac{\hat{e}_i^T \hat{e}_i}{r_i^{(t+1)}}$ Gl. (9)

$t = t + 1$

end

$\hat{\sigma}_i^{2(j+1)} = \hat{\sigma}_i^{2(t)}$

$j = j + 1$

end

Abb. 1 | Pseudocode für die iterative Berechnung der Varianzkomponenten und der Regressionskoeffizienten. Der hellrot markierte Code berechnet die Koeffizienten und der hellblau markierte Code die Varianzkomponenten

¹ Zur Unterscheidung der Berechnungsmethoden wird der Begriff Varianzfaktor verwendet, wenn $\hat{\sigma}^2$ bei der Regressionsanalyse wie üblich geschätzt wird. Der Begriff Varianzkomponente wird für die geschätzten $\hat{\sigma}_i^2$ der unterschiedlichen Datensätze (Kaufpreise, Angebotsdaten, Expertenwissen) aus der VKS verwendet.

Varianzkomponenten im Iterationsschritt $t = 0$ mit eins initialisiert. Als Abbruchkriterium dient die maximale Differenz zwischen den neu berechneten Varianzkomponenten der aktuellen Iteration und den Varianzkomponenten aus der vorherigen Iteration. Wenn diese Differenz den Schwellenwert von 10^{-5} unterschreitet, wird die While-Schleife beendet. Der verwendete Schwellenwert ist für die Schätzung der Varianzkomponenten im Rahmen der Wertermittlung ausreichend, da der Preis pro Quadratmeter höchstens im einstelligen Cent-Bereich nicht mehr ausreichend präzise geschätzt wird.

In jedem Schleifendurchlauf wird zuerst die Kofaktormatrix $\bar{V}$ der Regressionskoeffizienten nach Gl. (7) berechnet. Mit der Kofaktormatrix wird als nächstes der Redundanzanteil r_i eines jeden Datensatzes bestimmt (vgl. Gl. (8)). Die Variable n_i ist die Anzahl der Beobachtungen im entsprechenden Datensatz. Der Redundanzanteil eines Datensatzes beschreibt, wie groß der Beitrag seiner unabhängigen Variablen X_i zum Gesamtmodell (bestehend aus den unabhängigen Variablen aller Datensätze) ist.

Anschließend werden die Varianzkomponenten von den Datensätzen mit ihren Residuenquadratsummen und aktualisierten Redundanzanteilen nach Gl. (9) neu berechnet. Ein Datensatz mit kleinen Residuen und einem großen Redundanzanteil hat somit eine kleinere Varianzkomponente als ein Datensatz mit großen Residuen und einem kleinen Redundanzanteil. Als letzter Schritt eines Schleifendurchlaufs wird der Laufindex t_i um eins erhöht. Die in der hellblau markierten While-Schleife geschätzten Varianzkomponenten werden im Anschluss für die Berechnung der neuen Regressionskoeffizienten $\bar{\beta}$ in der hellrot markierten While-Schleife verwendet. Für die hellrot markierte While-Schleife kann dasselbe Abbruchkriterium wie für die hellblau markierte Schleife verwendet werden. Detailliertere mathematische Herleitungen zu dem verwendeten VKS-Ansatz sind in /Koch & Kusche 2002/ zu finden.

3 Entwicklung einer Auswertestrategie zur Gewichtung unterschiedlicher Marktdaten

In diesem Beitrag liegt der Fokus auf der Gewichtung von verschiedenen Marktdaten mittels VKS. Die Methodik zur Ableitung

von Gewichten wird bei der zukünftigen Entwicklung eines robusten bayesschen Regressionsansatzes für kaufpreisarme Lagen berücksichtigt. Zur Vereinfachung des mathematischen Modells wird die Annahme getroffen, dass in den Kauffällen keine Ausreißer vorhanden sind. Aufgrund der Verwendung eines Teilmarkts mit ausreichend vielen Markttransaktionen können die Ausreißer aus den Kauffällen vorab detektiert und entfernt werden. Somit ist die Annahme ausreißerfreier Daten gerechtfertigt. Zur Berechnung der Posteriori-Regressionskoeffizienten kann Gl. (5) direkt verwendet werden. Für die Untersuchung des VKS-Ansatzes werden Kauffälle, Befragungsdaten und Angebotsdaten, also $k = 3$ Datensätze, verwendet (vgl. Abschnitt 4).

Die VKS wird, wie bereits erwähnt, in anderen geodätischen Disziplinen seit Längerem eingesetzt. Im Bereich der Wertermittlung hat /Uhde 1982/ eine erste Anwendungsmöglichkeit für die VKS präsentiert. Hier wurden für einen Datensatz mehrere Varianzkomponenten in Abhängigkeit von der Streuung der Kaufpreise geschätzt. Eine erste Anwendung zur Kombination unterschiedlich erfasster Marktdaten wird in /Dorndorf et al. 2016/ vorgestellt: Anhand einer Closed-Loop-Simulation wurde die Anwendungsmöglichkeit der VKS für die Wertermittlung in einer kaufpreisarmen Lage mittels simulierter Datensätze analysiert. Die simulierten Datensätze hatten die markttypischen Streuungseigenschaften von Kauffällen, Expertenbefragung und Angebotsdaten. Die Closed-Loop-Simulation zeigte für die VKS zur Kombination der drei verschiedenen Datensätze gute Resultate. Bei der anschließenden Untersuchung des Ansatzes an echten Marktdaten konnte der Mehrwert der VKS allerdings nicht so eindeutig wie in der Simulation festgestellt werden. Aus diesem Grund wird die Untersuchung des VKS-Ansatzes an echten Marktdaten in diesem Beitrag weiterführend analysiert. Die Grundidee bei der hier durchgeführten Untersuchung besteht darin, dass ein Teilmarkt mit ausreichend vorhandenen Kauffällen für ein Vergleichswertverfahren mittels Regressionsanalyse verwendet wird. Aus diesen Kauffällen wird eine kaufpreisarme Lage simuliert, indem aus allen Kauffällen n_{kl} zufällige Kauffälle ausgewählt werden. Das Ergebnis der simulierten kaufpreisarmen Lage kann so mit dem Regressionsergebnis aus allen Kauffällen verglichen werden. Zusätzlich zu dem Vergleich der

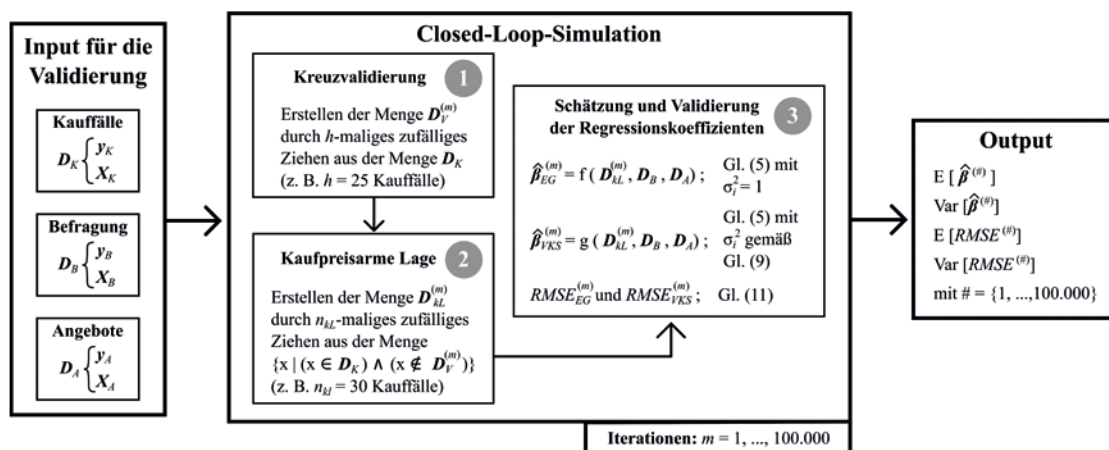


Abb. 2 | Schematische Darstellung des Ablaufs für die Simulation einer kaufpreisarmen Lage

Regressionskoeffizienten wird eine Kreuzvalidierung zur Bewertung der Ergebnisse durchgeführt. Diese Auswertestrategie wird auch in /Alkhatib & Weitkamp 2013/ präsentiert.

Aufgrund der simulierten kaufpreisarmen Lage ist das Ergebnis der Regressionskoeffizienten von der Auswahl der zufälligen Kauffälle abhängig. Für zuverlässigere Ergebnisse wird die in Abb. 2 dargestellte Closed-Loop-Simulation durchgeführt. Die Simulation verwendet als Inputparameter die Datensätze der Kauffälle, der Befragung und der Angebote. In jeder durchgeführten Iteration m werden zuerst aus dem Kauffalldatensatz D_K zufällig $h = 25$ Kauffälle für die Kreuzvalidierungsstichprobe $D_V^{(m)}$ zurückgelegt (Schritt 1). Für die anschließende Erstellung der simulierten kaufpreisarmen Lage in Schritt 2 werden nur noch die Kauffälle von D_K verwendet, die nicht Teil der Kreuzvalidierungsstichprobe sind. Aus dieser Datenmenge werden $n_{KL} = 30$ zufällige Kauffälle gezogen, welche die kaufpreisarme Lage $D_{KL}^{(m)}$ repräsentieren. Im Schritt 3 werden aus den Datensätzen der kaufpreisarmen Lage, der Befragung und den Angeboten die gemeinsamen Regressionskoeffizienten mit zwei unterschiedlichen Schätzstrategien berechnet. In der Schätzung mit $f(D_{KL}^{(m)}, D_B, D_A)$ werden die drei Datensätze mit einheitlichen Gewichten ($\hat{\sigma}_i^2 = 1$) kombiniert. Die Funktion $g(D_{KL}^{(m)}, D_B, D_A)$ verwendet hingegen die VKS nach Abb. 1 zur Bestimmung von unterschiedlichen Gewichten für die drei Datensätze. Somit werden in jeder Iteration die Regressionskoeffizienten $\hat{\beta}_{EG}$ und $\hat{\beta}_{VKS}$ geschätzt. Die beiden Ergebnisse werden anschließend mit der zuvor zurückgelegten Kreuzvalidierungsstichprobe $D_V^{(m)}$ analysiert. Zuerst werden hierfür mit Gl. (10) die Verbesserungen $\hat{e}_V$ im Datensatz der Kreuzvalidierungsstichprobe berechnet und anschließend mit Gl. (11) der Root-Mean-Square-Error (RMSE) bestimmt:

$$\hat{e}_{V,i}^{(m)} = X_V^{(m)} \hat{\beta}_i^{(m)} - y_V^{(m)}, i = EG \text{ oder VKS.} \quad (10)$$

$$RMSE_i^{(m)} = \sqrt{\frac{1}{h} \sum (\hat{e}_{V,i}^{(m)})^2}, i = EG \text{ oder VKS.} \quad (11)$$

Insgesamt werden in der Simulation 100 000 Iterationen durchgeführt, womit sich für jede Schätzfunktion 100 000 verschiedene Regressionskoeffizienten und Quadratwurzeln des durchschnittlichen Prognosefehlers (bzw. RMSE) ergeben. Das endgültige Ergebnis für die Simulation sind der Erwartungswert und die Varianz (bzw. Standardabweichung nach Ziehen der Quadratwurzel) der zuvor berechneten m Einzelergebnisse.

3.1 Einfluss der Ziehungsstrategie auf das Closed-Loop-Ergebnis

Das in Abb. 2 vorgestellte Prinzip zur Erstellung der simulierten kaufpreisarmen Lage beruht auf Ziehen mit Zurücklegen. In der Closed-Loop-Simulation wäre es möglich, eine bereits gezogene Kombination von Kauffällen ein weiteres Mal zu ziehen, wodurch die Ergebnisse aus zwei Iterationen identisch wären. Das Simulationsergebnis der Closed-Loop-Simulation wäre in diesem Fall stärker durch die identische Ziehungskombination beeinflusst als durch die anderen nicht identischen Kombinationen.

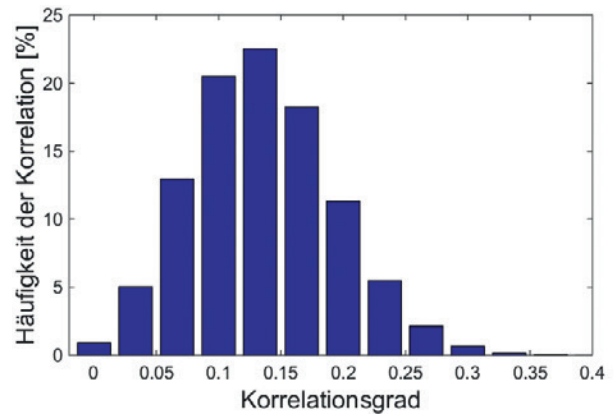


Abb. 3 | Häufigkeitsverteilung der Korrelationen zwischen zwei zufällig gezogenen kaufpreisarmen Lagen in den 100 000 durchgeführten Closed-Loop-Simulationen

Zwischen den einzelnen Iterationsergebnissen der Simulation existieren Korrelationen. Eine Korrelation von 1,0 zwischen zwei gezogenen kaufpreisarmen Lagen bedeutet, dass die gezogenen Stichproben identisch sind, wohingegen bei einer Korrelation von 0,5 nur die Hälfte der zufällig gezogenen Kauffälle in beiden (simulierten) kaufpreisarmen Lagen identisch ist.

Die zu erwartenden Korrelationsgrößen zwischen den Ziehungen in der Simulation hängen von der Anzahl der Iterationen und der Anzahl an möglichen Ziehungskombinationen für die kaufpreisarme Lage ab. Bei z. B. 197 Kauffällen, aus denen 30 Kauffälle gezogen werden, existieren $2,5 \cdot 10^{35}$ Kombinationsmöglichkeiten. Somit beträgt die Wahrscheinlichkeit bei 100 000 Iterationen, dass eine kaufpreisarme Lage öfter als einmal in der Simulation gezogen wird, nach der Binomialverteilung $p \approx 7,9 \cdot 10^{-62}$. Die Wahrscheinlichkeit in der Simulation, identische kaufpreisarme Lagen zu ziehen, liegt also nahe null.

In Abb. 3 ist die Verteilung der Korrelationen zwischen den in der Simulation gezogenen kaufpreisarmen Lagen dargestellt. Die am häufigsten auftretende Korrelation beträgt 0,13. Das entspricht vier identischen Kauffällen in zwei verschiedenen Ziehungen. Korrelationen größer als 0,3 existieren bei ca. 0,9 % der berechneten Korrelationen. Die größte Korrelation in der Simulation beträgt 0,57 (18 identische Kauffälle in zwei Ziehungen), wobei diese nur einmal in den 100 000 Iterationen auftritt.

Insgesamt zeigt das Ergebnis der Korrelationsverteilung in Abb. 3, dass die zufällig gezogenen kaufpreisarmen Lagen am häufigsten nur schwach miteinander korreliert sind. Eine systematische Verzerrung des Simulationsergebnisses, die aufträte, wenn z. B. dieselbe Ziehungskombination in der Simulation wiederholt verwendet würde, wird daher ausgeschlossen.

Eine weitere Möglichkeit, die Verzerrung des Simulationsergebnisses durch die Korrelationen zwischen den Iterationsergebnissen zu reduzieren, besteht in der Eliminierung von zufällig gezogenen kaufpreisarmen Lagen mit großen Korrelationen. Aufgrund der in dieser Simulation überwiegend existierenden kleinen Korrelationen wird hierauf verzichtet.

4 UNTERSUCHUNGSGEBIET

Als Untersuchungsgebiet für die in Abschnitt 3 beschriebene Auswertestrategie wurde ein Teilmarkt mit ausgeglichenem Angebots- und Nachfrageverhalten ausgewählt. Für den räumlichen Teilmarkt fiel die Wahl auf die Stadt Nienburg (Weser) in Niedersachsen mit ihrem sachlichen Teilmarkt der Ein- und Zweifamilienhäuser sowie Reihenhäuser. Die Kreisstadt Nienburg (Weser) liegt auf halber Strecke zwischen Hannover und Bremen und ist verkehrstechnisch gut angebunden (IC-Bahnhof und Bundesstraße). Der Landkreis Nienburg hat ca. 120 000 Einwohner. Die dominierende Bebauungsart ist mit 90 % dem Segment der Ein- und Zweifamilienhäuser (54 000 Haushalte) zuzuordnen /LK Nienburg 2015/. In der Stadt Nienburg leben ca. 30 000 Einwohner /Zensusdatenbank 2015/.

Für die Untersuchung wurden drei unterschiedliche Datenquellen verwendet. Neben den Kaufpreisen aus der Kaufpreissammlung des Gutachterausschusses der Stadt Nienburg standen Angebotsdaten von der Firma Immobilien Scout GmbH zur Verfügung. Zudem wurde durch die Befragung von Experten ein weiterer Datensatz erzeugt. In den nachfolgenden Abschnitten werden die verwendeten Datensätze detaillierter vorgestellt und analysiert. Die hier vorgestellten Analysen der Angebotsdaten beziehen sich auf das gesamte Kreisgebiet von Nienburg. Die für die Auswertestrategie genutzten Daten (Angebotsdaten, Kaufpreise und Expertenbefragung) beschränken sich auf den räumlichen Teilmarkt der Stadt Nienburg. Weiterführende Informationen zu den verwendeten Daten sind in /Soot et al. 2016/ zu finden.

4.1 Kaufpreise

In dem ausgewählten Untersuchungsgebiet stehen im Zeitraum 2011 bis 2015 ca. 250 Kauffälle zur Verfügung. Für alle diese Kauffälle existieren die beschreibenden Objektmerkmale Wohnfläche (WF in m²), Grundstücksgröße (F in m²), Baujahr (Jahr – 1946 $\triangleq$ 70 Jahre Nutzungsdauer), Bodenrichtwert (BRW in €/m²) und Ausstattungsstandard. Die abhängige Variable y ist der Kaufpreis pro Quadratmeter Wohnfläche (€/m²).

Dieselben Kauffälle wurden auch in der Untersuchung von /Dorndorf et al. 2016/ verwendet. Hier wurde festgestellt, dass ca. 70 % der Kauffälle eine Ausstattungsstufe von 3,0 besitzen und die übrigen Kategorien der Ausstattungsstufe nur schwach besetzt sind. Aufgrund dieser Verteilung unterscheidet sich der abgeleitete Regressionskoeffizient für den Ausstattungsstandard aus Kauffällen signifikant von dem abgeleiteten aus Befragungsdaten (hier variieren die Ausstattungsstandards mehr). Die Ausstattungsstufe

wird in der Automatisierten Kaufpreissammlung Niedersachsen aus den verschiedenen Eigenschaften des Objekts (Heizung, Sanitär, Fassade und Fenster) berechnet und als Ganzzahlwert in der Kaufpreissammlung gespeichert. Eine Ursache für die häufige Vergabe eines mittleren Ausstattungsstandards kann an fehlenden Informationen („Rückzug auf den Mittelwert“) sowie an dem einfachen Modell liegen. Neben dem Ausstattungsstandard ist zusätzlich bei 228 Kauffällen die Standardstufe bekannt, die eine größere Variation über die Qualitätskategorien als die Ausstattung aufweist. Die Standardstufe der Objekte ist in Anlehnung an die Anlage 2 der Sachwertrichtlinie /BMVBS 2012/ abgeleitet und in die Datenbank eingetragen. Das Modell der Standardstufe ist deutlich komplexer (neun Kategorien) und berechnet Dezimalwerte. Hier besitzen nur ca. 15 % der Kauffälle die Kategorie 3,0. Den größten prozentualen Anteil hat die Kategorie 2,0 mit ca. 25 % der Kauffälle. Die Standardstufen der restlichen Kauffälle sind etwa gleichmäßig auf die anderen Qualitätskategorien verteilt. Aufgrund des prozentual großen Anteils von Kauffällen mit mittlerem Ausstattungsstandard und der deutlichen Abweichung zur Verteilung der Standardstufe kann mit hoher Wahrscheinlichkeit davon ausgegangen werden, dass der Ausstattungsstandard die Qualität der Immobilien nur unzureichend beschreibt. Um eine bessere Bewertung des Einflusses der Qualität auf die Kaufpreise der Immobilien zu ermöglichen, wird im Folgenden die Standardstufe verwendet. Somit ergibt sich die in Tab. 1 zusammengefasste deskriptive Statistik für die Kauffälle. Die Auswirkungen der Verwendung der Standardstufe auf die geschätzten Regressionskoeffizienten werden in Abschnitt 4.4 diskutiert. Sowohl für die Ausstattungsstufe als auch für die Standardstufe werden in der Praxis lineare, stetige Modelle verwendet. Für die Vergleichbarkeit mit vorhandenen Modellen werden auch hier die Einflussgrößen linear modelliert; eine Verbesserung der Schätzung durch eine kategoriale Modellierung muss zukünftig geprüft werden.

4.2 Angebotsdaten

Insgesamt stehen auf Immobilienscout24 ca. 1 400 Angebote von Ein- und Zweifamilienhäuser im Kreisgebiet Nienburg für den Zeitraum 2011 bis 2015 zur Verfügung. Davon besitzen 760 Angebote eine Information über die genaue Lage der Immobilie. Über diese ist es möglich, 440 Angebote einem Kauffall aus der Kaufpreissammlung zuzuordnen. /Soot et al. 2016/ hat über diesen Zusammenhang die durchschnittliche Vermarktungsdauer der Immobilien und einen Offset zwischen Kaufpreis und Angebotspreis bestimmt. Für ca. 80 % der Angebote beträgt die Vermarktungsdauer weniger

als 1,5 Jahre und der Kaufpreis liegt im Durchschnitt 13 % unter dem Angebotspreis.

Damit die Angebotsdaten für die Untersuchung des VKS-Ansatzes verwendet werden können, benötigen sie dieselben beschreibenden Objekteigenschaften wie die Kauffälle. Auf Immobilienscout24 werden meistens die Wohnfläche, die Grundstücksgröße und

	Kaufpreis in €/m ²	WF in m ²	F in m ²	Baujahr	BRW in €/m ²	Standard- stufe
Mittelwert	1 065,56	133,59	648,18	26,64	74,49	2,45
Median	1 004,63	130,00	628,50	23,75	70,00	2,40
Standardabweichung	315,98	36,82	291,83	14,82	24,57	0,48
Minimum	475,63	67,00	118,00	4,00	33,00	1,40
Maximum	2 214,31	248,00	1 516,00	60,00	180,00	4,00

Tab. 1 | Deskriptive Statistik der Kauffälle

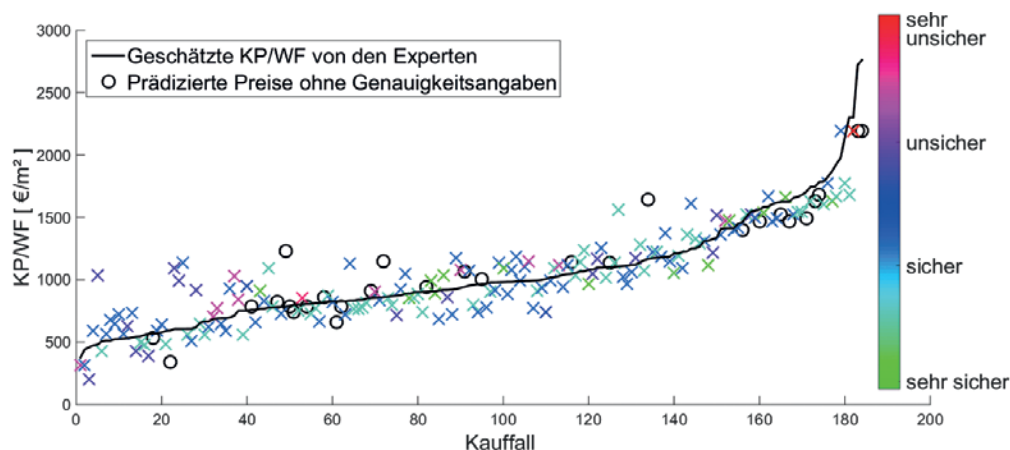


Abb. 4 | Gegenüberstellung der geschätzten Verkehrswerte der Experten (schwarze Linie) zu den prädizierten Kaufpreisen aus der Regressionsanalyse (Kreise/Kreuze). Die Farbe der Kreuze stellt die vom Experten eingeschätzte Genauigkeit seines geschätzten Verkehrswerts dar

das Baujahr für die Angebote mit angegeben. Bei Untersuchungen zu diesen Angaben wurde festgestellt, dass teilweise signifikant unterschiedliche Ausprägungen zwischen den Angaben von Immobilienscout24 und der Kaufpreissammlung existieren.

Die auf Immobilienscout24 fehlenden Angaben zum Bodenrichtwert können über die Adresse der Angebote ermittelt werden. Der Bodenrichtwert dient hier als Lageindikator. Die Bestimmung des Ausstattungsstandards ist allerdings eine größere Herausforderung, da in dem von der Firma Immobilien Scout GmbH zur Verfügung gestellten Datensatz keine ausreichenden Informationen über die Ausstattung vorliegen. In /Dorndorf et al. 2016/ wurde die Ausstattung aus den Angeboten zugeordneten Kauffällen bestimmt, wodurch für 109 Angebote im räumlichen Teilmarkt der Stadt Nienburg (Weser) ein Ausstattungsstandard vorlag. Durch die in Abschnitt 4.1 erläuterte Verwendung der Standardstufe als Qualitätsbeschreibendes Merkmal für die Immobilien reduziert sich die Anzahl der zur Verfügung stehenden Angebote auf 78. Somit ist von dem umfangreichen Angebotsdatensatz nur ein kleiner Teil der Angebote für die Untersuchung des VKS-Ansatzes geeignet. In zukünftigen Arbeiten soll deswegen untersucht werden, wie für eine größere Anzahl der vorhandenen Angebote die unabhängigen Variablen zuverlässiger bestimmt werden können.

4.3 Expertenwissen

Die Expertenbefragung wurde im September 2015 in Nienburg (nach einem ähnlichen wie in /Weitkamp & Alkhatib 2012a/ beschriebenen Verfahren) durchgeführt. An der Befragung haben zehn Gutachter teilgenommen. Diese Gutachter sind Mitglieder und Mitarbeiter des Gutachterausschusses Nienburg (Weser) und verfügen über lokale Marktkennntnisse. Die Gutachter wurden aufgefordert, auf Grundlage von Objektbeschreibungen Werteinschätzungen abzugeben. Dazu wurden den Gutachtern die Qualitätsbeschreibungen aus Gutachten sowie ca. zehn bis 15 Fotos des Objekts zur Verfügung gestellt. Des Weiteren hatten die Gutachter Zugriff auf den örtlichen Grundstücksmarktbericht und ein Berechnungsformular zur Durchführung der Sach- und Vergleichswertverfahren. Auf diesem Formular sollten die Gutachter auch einschätzen, wie sicher sie sich mit ihrem Ergebnis sind. Insgesamt wurden 50 verschiedene Objekte beschrieben, die jeweils von drei bis vier

Gutachtern bewertet wurden. Dadurch konnte eine Anzahl von 184 Verkehrswerten abgeleitet werden, wovon 24 bewertete Objekte ein Baujahr vor 1946 besitzen. Für die Untersuchung der VKS werden deswegen nur 150 Objekte als Pseudo-Kaufpreise weiterverwendet. Bei 85 % der geschätzten Verkehrswerte wurde eine Genauigkeitseinschätzung von den Experten mit angegeben.

4.3.1 Einschätzung der Genauigkeit

Die Genauigkeits- bzw. Unsicherheitseinschätzungen der Gutachter wurden in der Befragung mit erfasst, um zusätzliche Erkenntnisse über die geschätzten Verkehrswerte zu erhalten. Für die Analyse der Einschätzungen wurde mit den Befragungsdaten eine Regressionsanalyse durchgeführt. Die geschätzten Regressionskoeffizienten wurden anschließend für die Prädiktion von Preisen im Befragungsdatensatz verwendet. Dieses Ergebnis ist in Abb. 4 dargestellt. Die schwarze Linie sind die Verkehrswerte der Gutachter und die farbigen Kreuze sind die zugehörigen prädizierten Preise mit den geschätzten Genauigkeiten aus der Befragung. Beim Vergleich der prädizierten Preise mit den Verkehrswerten ist festzustellen, dass die Gutachter, die sich mit Ihrem Ergebnis sehr unsicher bzw. sehr sicher sind, also die extremen Werte zur Genauigkeit angekreuzt haben, mit am besten zum Regressionsmodell passen. Die schlechtesten Prädiktionen sind unter anderem jene, bei denen die Experten keine Angaben zur Genauigkeit gemacht haben (schwarze Kreise). Ebenso sind die prädizierten Preise, die eine durchschnittliche Angabe für die Unsicherheit erhalten haben, weiter von den geschätzten Verkehrswerten der Gutachter entfernt. Dieses Phänomen soll in zukünftigen Arbeiten genauer untersucht werden. Sollte sich eine derartige Systematik bestätigen, könnte sich hieraus evtl. eine Gewichtung für die Befragungsdaten ableiten lassen.

Neben der Experteneinschätzung zur Ermittlung der Genauigkeit lagen auch realisierte Kaufpreise zu einigen Befragungsobjekten vor. Dadurch konnten acht Objekte, für die der Kaufpreis vorlag, mit 31 Expertenbewertungen verglichen werden. Der Vergleich der durch die Gutachter geschätzten Verkehrswerte mit den realisierten Kaufpreisen zeigt eine gute Übereinstimmung (Streuung rd. 10 %). Somit kann davon ausgegangen werden, dass die von den Experten geschätzten Verkehrswerte mit dem Markt gut übereinstimmen.

4.4 Qualitätsunterschiede zwischen den Datensätzen

Alle drei Datensätze beschreiben denselben räumlichen wie sachlichen Teilmarkt. Deswegen ist davon auszugehen, dass zwischen den drei genutzten Datenquellen Korrelationen existieren. Angebotspreise sind die Verhandlungsgrundlage für Kaufpreise. Die zwischen Angebotspreisen und Kaufpreisen bestehenden Zusammenhänge wurden beispielsweise in /Soot et al. 2016/ untersucht. Auch das Expertenwissen stützt sich durch direkte Kaufpreisanalysen der Experten und Kaufpreisanalysen der Gutachterausschüsse (z. B. sonstige zur Wertermittlung erforderliche Daten in Marktberichten) an den Kaufpreisen.

Es ist wichtig, die Existenz dieser Korrelationen für den in Abschnitt 2.3 vorgestellten VKS-Ansatz zu beachten, da dieser nur für unabhängige bzw. unkorrelierte Datengruppen gültig ist. Allerdings ist eine genaue Abschätzung der existierenden Korrelationseffekte und ihrer Größen nicht trivial zu geben und stellt eine eigene Arbeit dar. Deswegen werden hier die Datengruppen derzeit als unkorreliert betrachtet. In zukünftigen Arbeiten sollen die Korrelationen zwischen den Datensätzen untersucht werden, um diese zum Beispiel durch eine Kovarianzkomponentenschätzung im mathematischen Modell mit zu berücksichtigen.

Für die Kombination der drei Datensätze in einem gemeinsamen Ausgleichungsansatz müssen sie sich auf denselben Wertermittlungstichtag beziehen. Als Stichtag wird das Datum der durchgeführten Expertenbefragung verwendet. Die zwei anderen Datensätze werden mithilfe des Bodenpreisindexes von Nienburg (extrapoliert) auf September 2015 angepasst. Damit die Datensätze frei von Ausreißern sind, wird wie in /Dorndorf et al. 2016/ für jeden Datensatz eine Ausreißersuche mit dem Data-Snooping

nach Baarda durchgeführt. Somit stehen für die Auswertestrategie 222 Kauffälle, 130 Befragungsdaten und 76 Angebotsdaten zur Verfügung. Die Regressionsergebnisse für die ausreißerbereinigten Datensätze sind in Tab. 2 und Tab. 4 zusammengefasst.

Die Kauffälle, Befragungsdaten und Angebotsdaten zeichnen sich untereinander durch ihre unterschiedlichen Erhebungsarten aus. Die hierdurch entstehende Heterogenität zwischen den Datensätzen wird zum Beispiel an den unterschiedlichen Posteriori-Varianzfaktoren und Bestimmtheitsmaßen in Tab. 2 deutlich. Bei den Befragungsdaten ist die Quadratwurzel aus dem Varianzfaktor mit 93,85 nur halb so groß wie bei den anderen beiden Datensätzen. Zusammen mit dem für die Wertermittlung untypisch hohen Bestimmtheitsmaß von 92 % bei den Befragungsdaten kann darauf geschlossen werden, dass die befragten Experten unabhängig voneinander die Verkehrswerte auf dem untersuchten Teilmarkt identisch einschätzen. Das Bestimmtheitsmaß spiegelt unter anderem auch die Transparenz des Grundstücksmarkts wider. Dies bestätigt das höhere Bestimmtheitsmaß insofern, dass Sachverständige den Immobilienmarkt besser kennen, als dies Verkäufer bzw. Käufer vermögen. Sachverständige geben einen objektivierten Wert² an. Im Gegensatz zu den Befragungsdaten weist die Quadratwurzel aus dem Varianzfaktor der Kauffälle mit 182,86 eine signifikant höhere Streuung auf, die unter anderem durch die individuellen Preisverhandlungen der Vertragspartner am Markt zu erklären ist. Es besteht allerdings auch die Gefahr, dass die Experten präziser sind, jedoch weiter als die Kauffälle vom „wahren Wert“ entfernt liegen. In aktuellen Forschungsansätzen sollen das Expertenwissen und die Informationsquellen der Experten näher untersucht werden. Die Angebotsdaten besitzen mit 201,96 die größte Quadratwurzel aus dem Varianzfaktor und mit 57 % das kleinste Bestimmtheitsmaß von den drei Datensätzen. Hier ist zu vermuten, dass die höhere

	Anzahl Beobachtungen	Posteriori $\sqrt{\sigma^2}$	Bestimmtheitsmaß
Kauffälle	222	182,86	67 %
Expertenwissen	130	93,85	92 %
Angebotsdaten	76	201,96	57 %

Tab. 2 | Qualitative Unterschiede zwischen den Datensätzen

	β_0 Absolutglied	β_1 WF	β_2 F	β_3 Baujahr	β_4 BRW	β_5 Ausstattungsstandard
Kauffälle	506,65	-4,89	0,31	15,40	3,41	110,46

Tab. 3 | Ergebnis der multiplen linearen Regression für den Datensatz der Kauffälle unter Verwendung des Ausstattungsstandards

	β_0 Absolutglied	β_1 WF	β_2 F	β_3 Baujahr	β_4 BRW	β_5 Standardstufe
Kauffälle	243,85	-4,82	0,37	10,80	4,07	258,05
Expertenwissen	507,28	-4,71	0,17	9,89	3,09	221,45
Angebotsdaten	349,29	-3,25	0,56	10,76	3,53	172,29

Tab. 4 | Ergebnis der multiplen linearen Regression für die verschiedenen Datensätze unter Verwendung der Standardstufe

² Die Objektivierung entsteht durch vorliegende Modelle und verfügbare Marktinformationen.

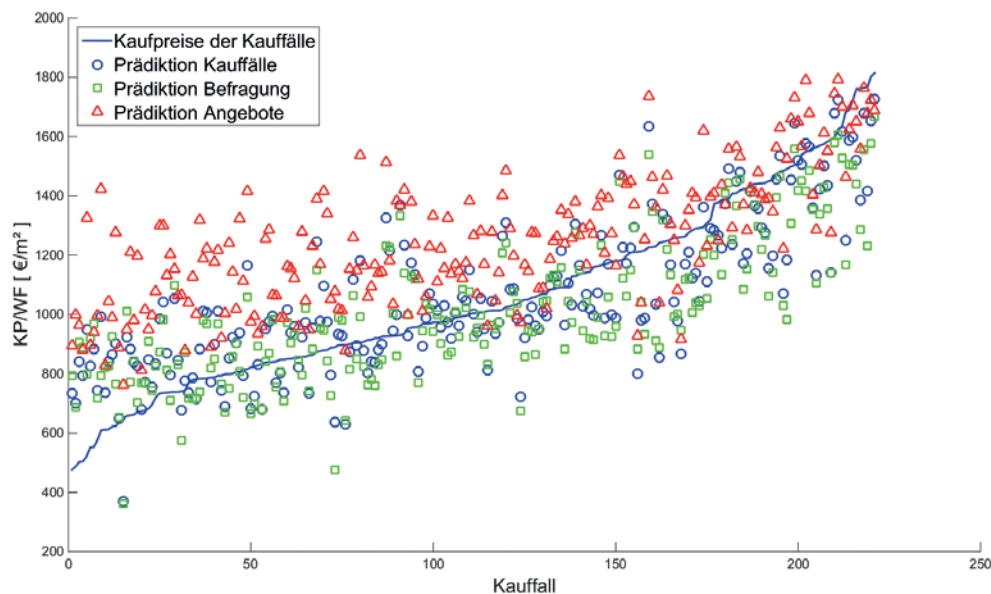


Abb. 5 | Prädiktion der Verkehrswerte aus verschiedenen Datensätzen jeweils unter Verwendung der Standardstufe

Variation neben der fehlenden Kenntnis des Grundstücksmarkts auch auf unterschiedliche Angebotsstrategien (einstellen höherer Preise zu Verhandlungszwecken) zurückgeführt werden kann. Allerdings ist der Unterschied der Angebotsdaten zu den Kaufpreisen nicht so signifikant, wie dies im Vergleich zu den Befragungsdaten der Fall ist.

Die Auswirkungen der Verwendung der Standardstufe auf die geschätzten Regressionskoeffizienten zeigen sich in Tab. 4. Das Ergebnis der Tab. 3 entspricht /Dorndorf et al. 2016/, wo für den Koeffizienten β_5 der Ausstattungsstandard verwendet wurde. In Tab. 4 wird dieser, wie im Abschnitt 4.1 beschrieben, durch die Standardstufe ersetzt. Bei der Verwendung des Ausstattungsstandards für β_5 wird ein Koeffizient von 110,46 geschätzt, wohingegen bei dem neuen Modell der Koeffizient für die Qualität (Standardstufe) 258,05 beträgt. Dieses Ergebnis liegt näher an dem von den Experten eingeschätzten Einfluss der Immobilienqualität auf den Kaufpreis. Derselbe Effekt trifft auch auf den Regressionskoeffizienten des Baujahrs zu, der mit 10,80 für die Kauffälle in Tab. 4 deutlich näher an der Experteneinschätzung liegt als das Ergebnis der Kauffälle in Tab. 3 mit 15,40. Allerdings hat die Verwendung der Standardstufe auch einen Einfluss auf die restlichen Regressionskoeffizienten. Dadurch passen das Absolutglied, die Grundstücksgröße und der Bodenrichtwert unter Verwendung der Standardstufe schlechter als unter Verwendung des Ausstattungsstandards zu den Befragungsdaten. Es ist hierbei zu berücksichtigen, dass die Koeffizienten der Grundstücksgröße und des Bodenrichtwerts bei der Verwendung des Ausstattungsstandards auch nicht optimal zu dem Expertenwissen passen.

Die Regressionskoeffizienten der Angebotsdaten unterscheiden sich teilweise signifikant von denen der anderen beiden Datensätze, wie z. B. bei der Wohnfläche. Die Ursache hierfür ist nicht zwingend die Verwendung der Standardstufe in der Regressionsanalyse, da die geschätzten Koeffizienten mit dem Ausstattungsstandard in /Dorndorf et al. 2016/ ebenso große Abweichungen zu den Kauffällen und den Befragungsdaten haben. Vielmehr sind die in Abschnitt 4.2 bereits erwähnten unterschiedlichen Merkmalsaus-

prägungen zwischen den Angebotsdaten und Kauffällen sowie der kleine Stichprobenumfang von 76 Angeboten unter anderem als Ursache in Betracht zu ziehen. Allerdings ist es genau wegen des kleinen Stichprobenumfangs zurzeit nicht möglich, eine zuverlässige Aussage über den Grund der signifikanten Unterschiede zwischen den Angebotsdaten und den anderen beiden Datensätzen zu geben.

Bei der zuvor durchgeführten Gegenüberstellung der Regressionskoeffizienten von den verschiedenen Datensätzen sind die existierenden Korrelationen zwischen den verschiedenen Koeffizienten zu beachten. Das Absolutglied und die Standardstufe im Kaufalldatensatz haben mit $-0,65$ die stärkste Korrelation. Durch die Korrelationen beeinflussen sich die geschätzten Regressionskoeffizienten gegenseitig, wodurch eine Änderung eines Koeffizienten auch zu einer Änderung eines anderen Koeffizienten führt. Um die Auswirkungen der drei unterschiedlichen Regressionsergebnisse mit ihren zwischen den Koeffizienten existierenden Korrelationen anschaulicher vergleichen zu können sind in Abb. 5 für jeden Datensatz prädizierte Preise dargestellt. Die Prädiktion ist mit der Designmatrix X_K der Kauffälle und den Regressionskoeffizienten $\hat{\beta}$ der drei Datensätze erfolgt. Auf der x-Achse sind die 222 Kauffälle nach ihrem Kaufpreis pro Quadratmeter Wohnfläche aufsteigend sortiert angeordnet. Die blaue Linie entspricht dem Quadratmeterpreis, bezogen auf die Wohnfläche als Beobachtung y des entsprechenden Kauffalls. Beim Vergleich der prädizierten Kaufpreise der Kauffälle (blaue Kreise) mit denen der Befragung (grüne Quadrate) ist zu erkennen, dass die beiden Prädiktionen nur wenig differieren. Allerdings sind die prädizierten Preise der Befragungsdaten meistens geringfügig niedriger als die der Kaufpreise. Das Prädiktionsergebnis der Angebotsdaten (rote Dreiecke) unterscheidet sich hingegen signifikant zu den Ergebnissen der anderen beiden Datensätze. Im niedrigen Preissegment sind die prädizierten Preise der Angebote überwiegend höher als die der Kauffälle oder der Befragung. Mit einem steigenden Preisniveau der Immobilien wird der Unterschied zwischen den Prädiktionsergebnissen der Angebotsdaten und den anderen zwei Datensätzen kleiner. Der

	Regression mit einheitlichen Gewichten	Regressionsanalyse mit VKS		
		Kauffälle	Befragung	Angebotsdaten
$\sqrt{\hat{\sigma}^2}$ bzw. $\sqrt{\hat{\sigma}_i^2}$	181,92	190,11	95,07	290,01

Tab. 5 | Varianzfaktor bzw. Varianzkomponenten für die zwei unterschiedlichen Gewichtungsansätze in der Regressionsanalyse

in /Soot et al. 2016/ festgestellte Preisoffset von durchschnittlich 13 % zwischen den Angeboten und Kauffällen kann in der hier durchgeführten Prädiktion somit nicht bestätigt werden. Die geschätzten Regressionskoeffizienten der Angebotsdaten repräsentieren stattdessen ein vom Quadratmeterpreis abhängigen Offset zu den Kauffällen. Die Stichprobe von /Soot et al. 2016/ umfasste allerdings 440 Angebotsdaten, wohingegen für die Regressionsanalyse nur 76 Angebote zur Verfügung standen. Wie bereits zuvor diskutiert, ist der Stichprobenumfang der Angebotsdaten in dieser Untersuchung sehr klein. Deswegen kann keine zuverlässige Aussage über das Verhalten des Preisoffsets gewährleistet werden. Aus diesem Grund erfolgt im Moment keine Kalibrierung der Angebotsdaten mittels Offset für die im Abschnitt 5 berechneten Untersuchungsergebnisse.

5 ANALYSE DES VKS-ANSATZES FÜR DIE DATENKOMBINATION

5.1 Ergebnis der Regressionsanalyse

Die in Abschnitt 3 beschriebene Auswertestrategie wird mit den in Abschnitt 4 vorgestellten Datensätzen durchgeführt. Aufgrund der zufälligen Erstellung der kaufpreisarmen Lage in der Closed-Loop-Simulation ist die Genauigkeit der Regressionsergebnisse

von der Iterationsanzahl abhängig. Deswegen wurde die Closed-Loop-Simulation mit 100 000 Iterationen insgesamt zehnmal wiederholt. Der Vergleich der Ergebnisse der zehn Experimente zeigt, dass sich die gemittelten Regressionskoeffizienten maximal um $\pm 0,1$ % voneinander unterscheiden. Somit reichen 100 000 Iterationen aus, um die Ergebnisse der Regressions-

analyse mit einheitlichen Gewichten den Ergebnissen der Regressionsanalyse mit VKS zuverlässig gegenüberzustellen.

In Tab. 5 sind die mittels VKS durchschnittlich geschätzten Varianzkomponenten für die drei Datensätze aus den 100 000 simulierten kaufpreisarmen Lagen angegeben. Bei der Berechnung der Regressionskoeffizienten mit dem VKS-Ansatz werden die Befragungsdaten mit einem Gewicht von $1/95,07^2$ fast doppelt so stark berücksichtigt wie die Kauffälle mit einem Gewicht von $1/190,11^2$. Den geringsten Einfluss auf das Ergebnis der Koeffizienten haben die Angebotsdaten mit einer Varianzkomponente von $290,01^2$. Die Varianzkomponenten der Befragungsdaten und der Kauffälle sind ungefähr identisch mit den in Tab. 2 geschätzten Varianzfaktoren für die beiden Datensätze. Bei den Angebotsdaten unterscheidet sich allerdings die Quadratwurzel aus der Varianzkomponente mit 290,01 signifikant zu der Quadratwurzel des Varianzfaktors mit 201,96. Der Unterschied der Angebotsdaten zu den anderen beiden Datensätzen ist somit größer als es allein der Varianzfaktor der Angebote vermuten lässt. Dies entspricht dem Ergebnis des Vergleichs zwischen den Datensätzen in Abb. 5, welches verdeutlicht, dass sich die Angebotsdaten im niedrigen Preissegment signifikant von den prädierten Preisen der Kauffälle und Befragung unterscheiden. Für die VKS kann somit ein hohes Potenzial für die Schätzung von unterschiedlichen Gewichten zur Kombination der Datensätze anhand ihrer Datenqualität nachgewiesen werden. Im Gegensatz dazu wird bei der Verwendung von einheitlichen

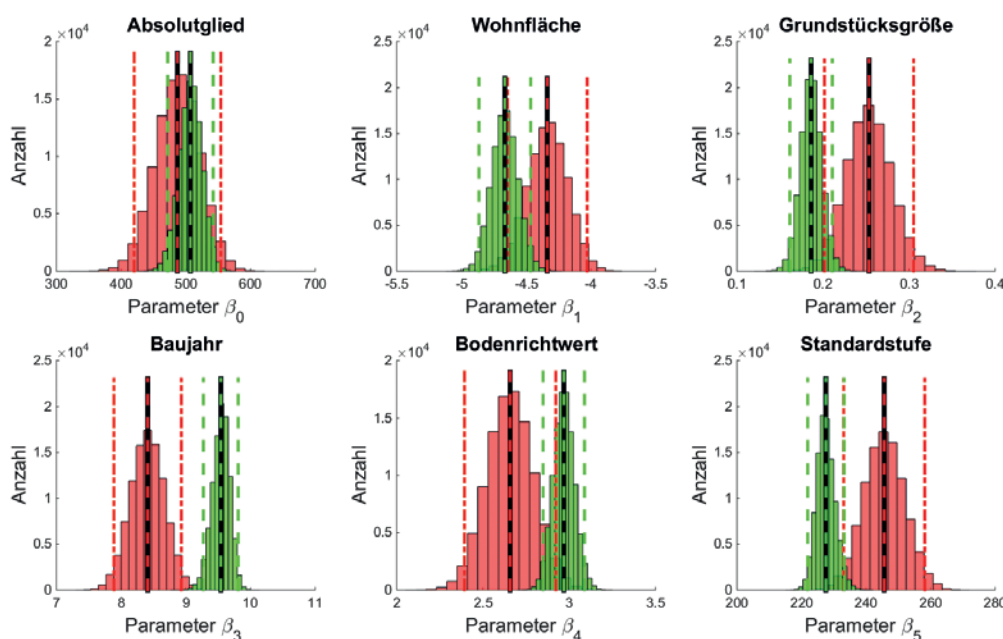


Abb. 6 | Geschätzte Regressionskoeffizienten für die 100 000 simulierten kaufpreisarmen Lagen. Rot: Regression mit einheitlichen Gewichten. Grün: Regression mit VKS

Parameter	Regressionsanalyse mit allen Kauf-fällen Koeffizienten	Regressionsanalyse für die kaufpreisarme Lage			
		Einheitliche Gewichte für jeden Datensatz		Gewichtung der Datensätze mittels VKS	
		Koeffizienten	σ	Koeffizienten	σ
β_0 [Absolutglied]	243,85	487,14	34,00	507,25	17,81
β_1 [WF]	-4,82	-4,34	0,16	-4,66	0,10
β_2 [F]	0,37	0,25	0,03	0,19	0,01
β_3 [Baujahr]	10,80	8,41	0,26	9,54	0,14
β_4 [BRW]	4,07	2,66	0,13	2,97	0,06
β_5 [Standardstufe]	258,05	245,65	6,37	227,56	2,86

Tab. 6 | Gemittelte Regressionskoeffizienten aus den 100 000 Simulationsergebnissen

Gewichten nur ein Varianzfaktor in Höhe von 181,92² geschätzt. Dies entspricht etwa dem Mittel der drei separaten Varianzkomponenten.

In Abb. 6 sind die Ergebnisse der Regressionskoeffizienten für 100 000 simulierte kaufpreisarme Lagen dargestellt. Die in Grün dargestellten Histogramme sind das Ergebnis für den Regressionsansatz mit VKS. Die roten Histogramme stellen das Resultat für die Kombination der drei Datensätze mit einheitlichen Gewichten in der Regression dar. Die farbig gestrichelten Linien repräsentieren das 95-%-Konfidenzintervall des entsprechenden Histogramms und die schwarz gestrichelten Linien sind die berechneten Mittelwerte aus den 100 000 Regressionskoeffizienten. An den Konfidenzintervallgrenzen ist zu erkennen, dass die geschätzten Regressionskoeffizienten mit der VKS eine kleinere Streuung aufweisen als die Koeffizienten aus der Regression mit gleichgewichteten Datensätzen. Beim Vergleich der Mittelwerte der Koeffizienten zeigt sich ein differenzierteres Bild zwischen den beiden Gewichtungsansätzen für die Datenkombination. Im Fall des Absolutglieds liegen die Mittelwerte innerhalb des Konfidenzintervalls der jeweils anderen Verteilung, wohingegen beim Baujahr ein signifikanter Unterschied zwischen den beiden Mittelwerten zu erkennen ist.

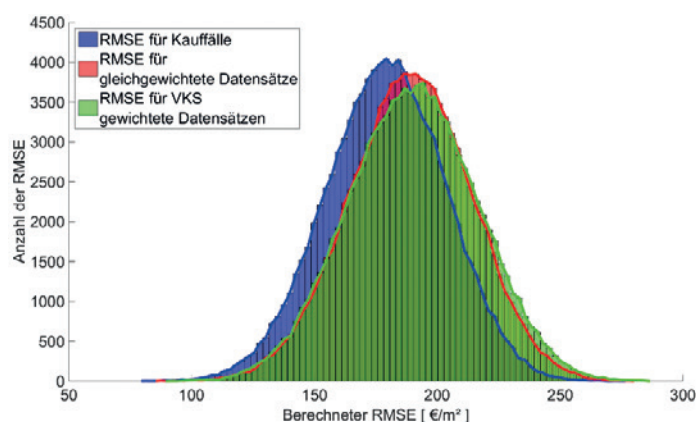


Abb. 7 | Histogramme der 100 000 berechneten RMSE aus den Kreuzvalidierungsstichproben. Blau: RMSE mit den Regressionskoeffizienten der Kauffälle berechnet. Rot: RMSE für die Regression mit einheitlichen Gewichten. Grün: RMSE für die Regression mit VKS (unterschiedliche Gewichte)

In Tab. 6 sind die Mittelwerte aus Abb. 6 den geschätzten Regressionskoeffizienten vom Kauf-falldatensatz aus Abschnitt 4.4 gegenübergestellt. Der Vergleich der Koeffizienten zeigt, dass das Absolutglied, die Grundstücksgröße und die Standardstufe bei der Schätzung mit einheitlichen Gewichten näher an dem Ergebnis der Kauffälle liegen. Im Gegensatz hierzu passen beim VKS-Ansatz die Wohnfläche, das Baujahr und der Bodenrichtwert besser zu den Regressionskoeffizienten der Kauffälle als es bei einer Gleichgewichtung der Datensätze der Fall ist. Der Vergleich der Koeffizienten der zwei unterschiedlichen Kombinerungsstrategien mit denen der Kauffälle ermöglicht somit keine eindeutige Aussage darüber, ob die Verwendung der VKS zur

Kombination der hier verwendeten Datensätze zu einem marktüblicheren Verkehrswert führt. Um dies zu beurteilen, wird im nachfolgenden Abschnitt das Ergebnis der Kreuzvalidierung betrachtet.

5.2 Ergebnis der Kreuzvalidierung

Die berechneten RMSE aus den Verbesserungen der 100 000 Kreuzvalidierungsstichproben sind in Abb. 7 dargestellt. Das rote Histogramm repräsentiert die RMSE für die Regressionsanalyse mit einheitlichen Gewichten. Die in Grün dargestellten RMSE sind das Ergebnis für die Regression mit VKS. Als Referenz zur Beurteilung dieser beiden Ergebnisse wurden zusätzlich mit den Regressionskoeffizienten der Kauffälle (Tab. 4) für dieselben Kreuzvalidierungsstichproben die RMSE berechnet. Das blaue Histogramm weist daher die kleinsten RMSE auf, weil die Regressionskoeffizienten der Kauffälle durch die Schätzung mit der Methode der kleinsten Quadrate die kleinste Verbesserungsquadratsumme für den Kaufalldatensatz besitzen. Die RMSE vom grünen und roten Histogramm sind aus diesem Grund im Durchschnitt ca. 10 €/m² größer als die RMSE der blauen Verteilung. Dieses Ergebnis zeigt, dass

in simulierten kaufpreisarmen Lagen nahezu identische Verkehrswerte ermittelt werden wie in der nicht reduzierten, insofern kaufpreisreichen Lage, wenn für die kaufpreisarme Lage alternative Daten wie Befragungen oder Angebotsdaten zur Verfügung stehen. Die roten und grünen RMSE weisen allerdings keinen signifikanten Unterschied auf. Dies ist auch an den fast identischen Mittelwerten und Standardabweichungen der beiden Verteilungen in Tab. 7 zu erkennen.

	Kauffälle	Einheitliche Gewichtung	VKS
Mittelwert in €/m ²	178,85	188,84	189,99
σ_{RMSE} in €/m ²	24,30	24,91	26,15

Tab. 7 | Statistische Momente der RMSE der Kreuzvalidierungsstichprobe

6 FAZIT UND AUSBLICK

6.1 Verwendung weiterer Marktdaten in kaufpreisarmen Lagen

Das Ergebnis der Kreuzvalidierung zeigt, dass durch die Kombination der 30 Kauffälle mit weiteren Marktdaten eine Wertermittlung für die simulierten kaufpreisarmen Lagen mittels Regressionsanalyse möglich ist. Die prädizierten Verkehrswerte der kaufpreisarmen Lagen streuen im Durchschnitt 10 €/m² mehr, als wenn mit den gesamten 222 Kauffällen prädiziert wird. Dies entspricht bei durchschnittlich 1 065 €/m² Wohnfläche einer Streuung von 0,9 %.

6.2 VKS-Ansatz

Die Ergebnisse für die simulierten kaufpreisarmen Lagen zeigen, dass die geschätzten Regressionskoeffizienten mit dem VKS-Ansatz präziser sind (kleinere Streuung) als mit einer einheitlichen Gewichtung (vgl. Abb. 6). Der VKS-Ansatz ermöglicht in dieser Untersuchung aber keine bessere Schätzung der Verkehrswerte im Kauffalldatensatz. Für die gleichgewichteten Datensätze ist der gemittelte RMSE aus der Kreuzvalidierungsstichprobe mit 189 €/m² sogar geringfügig kleiner als der RMSE des VKS-Ansatzes mit 190 €/m². Der Grund hierfür sind die Befragungsdaten, deren Regressionskoeffizienten sich teilweise signifikant im Vergleich zu denen des Kauffalldatensatzes unterscheiden. In der Regression mit VKS werden die Befragungsdaten aufgrund ihrer geringen Streuung am stärksten in der Schätzung berücksichtigt. Hier ist jedoch Vorsicht geboten: Aus Präzision eines Datensatzes kann nicht zwingend auf die Richtigkeit der geschätzten Regressionskoeffizienten geschlossen werden. Die Richtigkeit ist unter anderem von der Korrektur aller systematischen Effekte eines Datensatzes zum wahren Wert vor der Schätzung abhängig.

6.3 Expertenwissen

Die Genauigkeitsuntersuchung der Expertenbefragung in Abschnitt 4.3.1 zeigt, dass für Befragungsobjekte, zu denen ein Kauf fall existiert, kein systematischer Unterschied zwischen dem Kaufpreis und dem geschätzten Verkehrswert der Experten festzustellen ist. Dieses Ergebnis wird auch bei der Analyse der Qualitätsunterschiede in Abschnitt 4.4 bestätigt: Die Angebotsdaten weisen nur einen signifikanten Offset im niedrigen Preissegment auf. Es wurde bei der Untersuchung der Befragungsdaten jedoch noch nicht betrachtet, ob die den Experten zur Bewertung vorgelegten Immobilien die Grundgesamtheit der im Teilmarkt existierenden Objekte repräsentieren.

Bei der Auswahl der Objekte für die Befragung wurde darauf geachtet, dass die Verteilungen der unabhängigen Variablen x_i der 50 Befragungsobjekte bestmöglich die Verteilungen der unabhängigen Variablen der Kauffälle repräsentieren. Zusätzlich wurde geprüft, ob die Einflussgrößen der Befragungsobjekte zueinander realistisch sind und die Immobilie somit theoretisch im Teilmarkt existieren könnte. Beispielsweise ist im Kauffalldatensatz die Wohn-

fläche immer kleiner als die zugehörige Grundstücksgröße der gehandelten Einfamilienhäuser.

Allerdings wurde bei der Erstellung des Befragungsdatensatzes nicht untersucht, ob eine Stichprobe von 50 unterschiedlichen Befragungsobjekten ausreichend groß ist, um die Grundgesamtheit des Teilmarkts adäquat zu repräsentieren. Bei der Untersuchung wurden zwar 130 Pseudo-Kaufpreise verwendet, womit die Anforderung von 15 Kauffällen pro Einflussgröße erfüllt ist. Die 50 unterschiedlichen Befragungsobjekte sind aber zu wenig, um der Forderung nach 75 Beobachtungen für die fünf Einflussgrößen zu entsprechen.

6.5 Ausblick

In zukünftigen Arbeiten sollen die Unterschiede zwischen dem Befragungsdatensatz und dem Kauffalldatensatz weiter analysiert werden, um die Erstellung und Durchführung zukünftiger Expertenbefragungen zu verbessern. Von besonderem Interesse ist hierbei die Frage, welche Objekte ein Befragungsdatensatz enthalten muss, um den zu untersuchenden Teilmarkt angemessen zu repräsentieren. Auch die systematischen Unterschiede zwischen den verschiedenen Marktdaten sind in diesen Zusammenhang genauer zu untersuchen, um eine Korrektur dieser Effekte (beispielsweise Offset) zu ermöglichen. Dies trifft insbesondere auf die Angebotsdaten zu, die einen signifikanten Offset in den bisher durchgeführten Untersuchungen aufwiesen. Die ersten Erkenntnisse über den Offset sollen in zukünftige Arbeiten weiter analysiert werden, um eine zuverlässige Modellierung des Offsets zu ermöglichen.

Neben den systematischen Unterschieden sind die existierenden Korrelationen zwischen den verschiedenen Datensätzen detaillierter zu untersuchen. Erst diese Ergebnisse werden zeigen, ob die Kombination der Datensätze mit dem VKS-Ansatz zielführend ist oder eine andere Vorgehensweise für die Gewichtung der verschiedenen Marktdaten in der Regressionsanalyse zu wählen ist.

DANKSAGUNG

Die Forschung in diesem Beitrag ist im Rahmen des DFG-Projekts 60451047 entstanden. Die Autoren danken der DFG herzlichst für die Finanzierung. Weiterhin danken wir der Immobilien Scout GmbH, der Katasterverwaltung in Niedersachsen und speziell dem Gutachterausschuss Nienburg für die Bereitstellung der Daten und die Teilnahme an der Befragung.

LITERATUR

Alkhatib, H.; Weitkamp, A. (2012): Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung. Teil 1. In: Zeitschrift für Geodäsie, Geoinformation und Landmanagement (zfv), 137(2012)2, 93–102.

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

Amiri-Simkooei, A. R. (2007): Least-squares variance component estimation: theory and GPS applications. PhD thesis, Delft University of Technology. Publication on Geodesy, 64. Netherlands Geodetic Commission, Delft, Niederlande.

BMVBS (2012): Richtlinie zur Ermittlung des Sachwerts (Sachwertrichtlinie – SW-RL) vom 5. September 2012 (BAN AT 18.10.2012 B1).

Dorndorf, A.; Soot, M.; Weitkamp, A.; Alkhatib, H. (2016): Development of a Robust Bayesian Approach for Real Estate Valuation in Areas with Few Transactions. FIG Working Week 2016, Christchurch, Neuseeland.

Fahrmeir, L.; Kneib, T.; Lang, S. (2009): Regression: Modelle, Methoden und Anwendungen. 2. Auflage. Springer, Berlin/Heidelberg/New York.

Hartung, J.; Elpelt, B.; Klöser, K.-H. (2009): Statistik: Lehr- und Handbuch der angewandten Statistik. 15. Auflage. Oldenbourg, München.

Kacker, R.; Jones, A. (2003): On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent. In: Metrologia, 40(2003), 235–248.

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. Auflage. Bundesanzeiger, Köln.

Koch, K.-R. (1997): Parameterschätzung und Hypothesentests in linearen Modellen. 3. Auflage. Dümmler, Bonn.

Koch, K.-R. (2007): Introduction to Bayesian Statistics. 2. Auflage. Springer, Berlin/Heidelberg/New York.

Koch, K.-R.; Kusche, J. (2002): Regularization of geopotential determination from satellite data by variance components. In: Journal of Geodesy, 76(2002), 259–268.

Kroese, D. P.; Taimre, T.; Botev, Z. (2011): Handbook of Monte Carlo Methods. Wiley, Hoboken, NJ.

LK Nienburg – Landkreis Nienburg (Weser) (2015): Leben im Landkreis – Wohnen. <http://www.lk-nienburg.de/leben-im-landkreis/zahlen-daten-fakten/wohnen/> (28.09.2015).

Pelzer, H. (1978): Ein indirektes Vergleichswertverfahren unter Anwendung statistischer Methoden. In: Zeitschrift für Vermessung (zfv), 103(1978)6, 245–254.

Reuter, F. (2006): Zur Ermittlung von Bodenwerten in kaufpreisarmen Lagen. In: Flächenmanagement und Bodenordnung (fub), 68(2006)3, 97–107.

Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. (2008): Global Sensitivity Analysis. The Primer. Wiley, Chichester, UK.

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, Neuseeland.

Uhde, C. (1982): Mathematische Modelle zur Analyse von Grundstücksmärkten. Dissertation, Technische Universität Hannover.

Urban, D.; Mayerl, J. (2011): Regressionsanalyse: Theorie, Technik und Anwendung. 4. Auflage. VS Verlag für Sozialwissenschaften, Wiesbaden.

Weitkamp, A.; Alkhatib, H. (2012a): Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung. Teil 2. In: Zeitschrift für Geodäsie, Geoinformation und Landmanagement (zfv), 137(2012)2, 103–114.

Weitkamp, A.; Alkhatib, H. (2012b): The Bayesian approach in the valuation – a strategy to handle markets with low purchasing prices? FIG Working Week 2012, Rom, Italien.

Weitkamp, A.; Alkhatib, H. (2014): Die Bewertung kaufpreisarmer Lagen mit multivariaten statistischen Verfahren: Möglichkeiten und Grenzen robuster Methoden bei der Auswertung weniger Kauffälle. In: allgemeine vermessungsnachrichten (avn), 121(2014)1, 3–12.

Yang, Y.; Xu, T.; Song, L. (2005): Robust Estimation of Variance Components with Application in Global Positioning System Network Adjustment. In: Journal of Surveying Engineering, 131(2005)4, 107–112.

Zaddach, S. (2016): Zum Beitrag Bayesscher Schätzverfahren in der Vergleichswertermittlung. Dissertation. Deutsche Geodätische Kommission, Reihe C, 772. München.

Zensusdatenbank (2015): Zensusergebnisse Gebäude und Wohnungen sowie Wohnverhältnisse der Haushalte (2011). <https://ergebnisse.zensus2011.de/#StaticContent:032560022022> (28.09.2015).

Ziegenbein, W. (1977): Zur Anwendung multivarianter Verfahren der mathematischen Statistik in der Grundstückswertermittlung. Dissertation, Technische Universität Hannover.

Ziegenbein, W. (2010): Immobilienwertermittlung. In: Kummer, K.; Frankenberg, J. (Hrsg.): Das deutsche Vermessungs- und Geoinformationswesen 2010. Wichmann, Heidelberg, 421–468.

M. Sc. Alexander Dorndorf

LEIBNIZ-UNIVERSITÄT HANNOVER
GEODÄTISCHES INSTITUT

Nienburger Straße 1 | 30167 Hannover
dorndorf@gih.uni-hannover.de



M. Sc. Matthias Soot

TU DRESDEN
GEODÄTISCHES INSTITUT

Helmholtzstr. 10 | 01067 Dresden
matthias.soot@tu-dresden.de



Prof. Dr.-Ing. Alexandra Weitkamp

TU DRESDEN
GEODÄTISCHES INSTITUT

Helmholtzstr. 10 | 01067 Dresden
alexandra.weitkamp@tu-dresden.de



Dr.-Ing. Hamza Alkhatib

LEIBNIZ-UNIVERSITÄT HANNOVER
GEODÄTISCHES INSTITUT

Nienburger Straße 1 | 30167 Hannover
alkhatib@gih.uni-hannover.de



Manuskript eingereicht: 24.10.2016 | Im Peer-Review-Verfahren begutachtet