Anwendung der Kollokation als erweitertes Vergleichswertverfahren in der Immobilienwertermittlung

Sebastian Zaddach und Hamza Alkhatib

Zusammenfassung

Das in der Immobilienwertermittlung genutzte Vergleichswertverfahren beruht aus mathematisch-statistischer Sicht auf der multiplen linearen Regressionsanalyse und wird überwiegend in der Form des mittelbaren Preisvergleichs genutzt. Seit Jahrzehnten hat sich die Regressionsanalyse als Standardverfahren für die Verkehrswertermittlung etabliert, sich seitdem in der Praxis jedoch nicht signifikant weiterentwickelt. In diesem Beitrag wird die Erweiterung der klassischen Regressionsanalyse als Vergleichswertverfahren um einen Kollokationsansatz untersucht. Ausgehend von einem einfachen Trendmodell ermöglicht die Kollokation durch die Berücksichtigung eines Signalanteils die Ausschöpfung bislang nicht erfasster stochastischer Informationsanteile, die in der Regression nicht modelliert werden können. Die Ergebnisse der Regressionsanalyse werden damit verbessert und liefern im Idealfall einen höheren Erklärungsbeitrag zu den realen Kaufpreisen.

Summary

The use of the sales comparison approach for real estate valuation purposes is from a mathematical-statistical point of view based on the multiple linear regression analysis and is used primarily in the form of indirect price comparison. For decades, the regression analysis has become a standard method for determining the market value. Nevertheless, since its introduction the method has not been enhanced significantly. The aim of this contribution is to enhance the use of regression analysis as a sales comparison approach by introducing a collocation approach. Starting from a simple trend model the collocation approach allows to exhaust additional information, which cannot be modeled within the regression approach. In this way, it can be assumed, that the results of the regression analysis are improved and the extended approach is able to give a higher explanation to the real purchase prices.

Schlüsselwörter: Wertermittlung, Vergleichswertverfahren, multiple lineare Regressionsanalyse, Kollokation

1 Anwendung des Vergleichswertverfahrens

Ziel einer Immobilienwertermittlung ist die Ermittlung marktkonformer und allgemein anerkannter Verkehrswerte nach § 194 des Baugesetzbuches (BauGB). Im Fokus der hier vorgestellten Forschung steht das Vergleichswertverfahren als eines der nach der Immobilienwertermittlungsverordnung (ImmoWertV) gesetzlich normierten Standardmethoden. Das Vergleichswertverfahren kann grundsätzlich sowohl bei der Wertermittlung unbebauter wie auch bebauter Grundstücke zum Einsatz kommen. Es basiert auf der Überlegung, den Verkehrswert eines Wertermittlungsobjekts aus einer ausreichenden Anzahl von zum Wertermittlungsstichtag zeitnahen Kaufpreisen zu bestimmen, die mit dem Zielobjekt in hinreichender Art in den wesentlichen wertbeeinflussenden Merkmalen übereinstimmen. In diesem Fall kann es als die Methode mit der höchsten Marktnähe betrachtet werden, da die Vergleichspreise für im Wesentlichen gleiche Objekte den sichersten Anhalt für die Verkehrswertermittlung bieten und das aktuelle Marktgeschehen widerspiegeln (Kleiber 2010). Zu den Arten des Preisvergleichs zählt u. a. die mittelbar statistische Methode, bei der die durch Abweichungen in den wesentlichen wertbeeinflussenden Merkmalen hervorgerufenen Wertunterschiede bei den Objekten bestimmt werden. Im Entwurf zur Vergleichswertrichtlinie wird in diesem Zusammenhang allgemein die Nutzung einer »ausgleichenden mehrdimensionalen Schätzfunktion« vorgeschlagen. Diese Formulierung stellt die Wahl eines geeigneten mathematisch-statistischen Verfahrens grundsätzlich frei. Die Anwendung der Regressionsanalyse ist bei der Analyse von Grundstücksdaten durch die Forschungen von Ziegenbein (1977) seit ca. 35 Jahren Standard für Wertermittlungszwecke in vielen Gutachterausschüssen. Seit der Adaption in diesem Bereich hat sich das Verfahren für vielseitige Zwecke in der Praxis bewährt; so wird es neben der Ermittlung des Verkehrswertes nach § 194 BauGB für bebaute oder unbebaute Grundstücke auch für die Ableitung weiterer wertermittlungsrelevanter Daten wie Umrechnungskoeffizienten oder Indexreihen verwendet.

Durch die Modellierung im Regressionsansatz wird eine geeignete Zielgröße, z. B. der Kaufpreis pro Bezugseinheit, durch ein lineares (oder ggf. nichtlineares) Modell als Trendfunktion erklärt. Der hier vorgestellte Ansatz basiert auf der Tatsache, dass bei Grundstücksmarktanalysen zum einen Einflussgrößen bekannt sind, die nicht oder nur sehr schwer in eine funktionale Beziehung zum Kaufpreis gebracht werden können (z. B. Lage), zum anderen eine große Zahl von Einflussgrößen und deren Wirken auf den Kaufpreis nicht bekannt sind oder erhoben werden können (z. B. Käuferverhalten). Das deterministische Modell der Regressionsanalyse kann damit nicht umfassend und korrekt aufgestellt werden. Differenziertere Wertverhältnisse sind durch einfache Funktionen nicht zu erfassen. Durch die Einführung eines Kollokationsansatzes soll die Anwendung statistischer Methoden im Vergleichswertverfahren signifikant erweitert werden. Dabei sollen die Grundlagen für aufbauende Forschungsansätze hinsichtlich einer Verbesserung der Unsicherheitsmodellierung geschaffen werden sowie die Ausschöpfung von Informationen, welche bislang im Regressionsansatz nicht berücksichtigt werden, im Vordergrund stehen. Zu diesem Zweck wird die Systemgleichung der Regressionsgleichung als funktionaler Trend um einen Signalanteil als zusätzliche Komponente erweitert, mit dem Ziel, Teile der Fehlspezifikationen aufzufangen.

2 Die Kollokation als Prädiktionsmodell

Der Verkehrswert einer Immobilie wird durch den Vergleich mit den Kaufpreisen weiterer Grundstücke ermittelt, die in den wesentlichen wertbeeinflussenden Merkmalen mit dem Zielgrundstück übereinstimmen. Die Regressionsanalyse gibt die Möglichkeit, Variationen der Zielgröße durch die Variabilität der Einflussgrößen zu erklären, um so einen funktionalen Zusammenhang aufzuzeigen. Die Zusammenhänge zwischen der Zielgröße und den Einflussgrößen sind häufig nur vage bekannt; Kenntnisse über eine exakte Funktion sind nicht gegeben. Die Regressionsgleichung kann in allgemeiner Matrixnotation mit

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1}$$

formuliert werden. Der Vektor $\boldsymbol{y}$ beinhaltet die Beobachtungen (hier: Kaufpreise pro Bezugseinheit), $\boldsymbol{X}$ bezeichnet die Designmatrix der wertrelevanten Einflussgrößen und $\boldsymbol{\beta}$ den Vektor der unbekannten und zu schätzenden Regressionskoeffizienten. Der Term $\boldsymbol{X\beta}$ stellt die systematische Komponente (Trendanteil) dar, der aus gegebenen Daten abgeleitet werden kann. Der Vektor $\boldsymbol{\varepsilon}$ enthält als stochastische Komponente die durch den Trend nicht erklärbaren, unsystematischen Abweichungen von der Zielgröße (Pelzer 1976). Das funktionale Modell nach Gl. 1 wird ergänzt durch das stochastische Modell, welches durch die Varianz-Kovarianzmatrix (VKM)

$$\boldsymbol{\Sigma}_{yy} = \sigma^2 \boldsymbol{P}^{-1} \tag{2}$$

gegeben ist. Dabei bezeichnet σ^2 die unbekannte Varianz der Gewichtseinheit und $\boldsymbol{P}$ die bekannte Gewichtsmatrix der Beobachtungen. In Wertermittlungsanwendungen werden i.d.R. die einzelnen Beobachtungen als voneinander unabhängig mit gleichen Varianzen angenommen. Ist dies nicht der Fall, wird dieses durch Transformation der Beobachtungen in eine symmetrische Verteilung erreicht. Die Gewichtsmatrix lässt sich dann als Einheitsmatrix formulieren mit $\boldsymbol{P} = \boldsymbol{I}$ (Koch 1995). Die Schätzung der Regressionskoeffizienten erfolgt nach der Methode der kleinsten Quadrate.

Für die Herleitung zur Lösung der Regressionsaufgabe sowie Ansätze zur Beurteilung der Modellgüte, insbesondere hinsichtlich der Anforderungen aus Wertermittlungssicht, wird auf weiterführende Literatur verwiesen (vgl. u.a. Pelzer 1976, Ziegenbein 1977, Fahrmeir, Kneib und Lang 2009).

Mittels der Regressionsanalyse kann der Vergleichswert y_i eines Wertermittlungsobjektes geschätzt werden. Der geschätzte Wert kann grundsätzlich nur ein angenäherter Wert sein; zu komplex sind die wertrelevanten Einflussgrößen und deren Zusammenwirken in der marktorientierten und akteursabhängigen Wertermittlung für eine mathematisch zutreffende Formulierung. Zwischen den Funktionswerten der Regressionsfunktion und den tatsächlichen Kaufpreisen verbleibt i.d.R. eine Differenz, welche die unvollkommene Modellierung widerspiegelt.

Für die Nutzung eines zusätzlichen Informationsgehaltes in den Daten, welcher jedoch bislang nicht berücksichtigt wird, bietet sich das Verfahren der Kollokation nach kleinsten Quadraten an. Die Methode wurde bereits um 1970 entwickelt (vgl. Moritz 1980). Gegen Ende der 1970er Jahre wurde die Theorie erstmals auf Fragen der Wertermittlung übertragen (vgl. Pelzer 1978, Ziegenbein und Hawerk 1978, Uhde 1982). Die Grundidee der Kollokation beruht auf der Überlegung, dass sich jede Beobachtung (hier: Kaufpreis) zerlegen lässt in einen regelmäßig systematischen Anteil *(Trend)*, einen unregelmäßig systematischen Anteil *(Signal)* sowie einen unregelmäßig zufälligen Anteil *(Rauschen)*.

Der *Trend* ermöglicht die Erklärung eines Kaufpreises durch meist einige wenige dominierende Wertmerkmale. Das *Signal* enthält in Form von unregelmäßigen, aber systematischen Anteilen einen Informationsgehalt, den es für die Verbesserung der Trendfunktion und damit der Modellierung der Zielgröße auszuschöpfen gilt. Dabei handelt es sich um eine nach Trendabzug verbleibende Restsystematik, die nicht in weitere einzelne Anteile bestimmter wertbeeinflussender Merkmale zerlegt werden kann. Die Interpretation kann lediglich als Gesamteffekt erfolgen, der sich aus dem rein stochastischen Verhalten der Daten ergibt, aber einen Beitrag zur Erklärung der Zielgröße in sich trägt. Im *Rauschen* verbleibt die zufällige, durch das Modell nicht erklärbare Streuung.

2.1 Funktionales Modell

Das Ausgleichungsmodell nach Gl. 1 wird erweitert um die Trennung der Residuen $\boldsymbol{\varepsilon}$ in einen Signalanteil $\boldsymbol{s}$ und einen Rauschanteil $\boldsymbol{v}$, sodass sich als funktionales Modell die Gleichung

$$\boldsymbol{y} = \underbrace{\boldsymbol{X\beta}}_{Trend} + \underbrace{\boldsymbol{Rs}}_{Signal} + \underbrace{\boldsymbol{v}}_{Rauschen} \tag{3}$$

ergibt (Moritz 1980). Die Matrix $\boldsymbol{R}$ des Signalanteils enthält den Zusammenhang zwischen den Beobachtungen und dem Signal. Aufgrund fehlender Kenntnis über den funktionalen Zusammenhang wird $\boldsymbol{R} = \boldsymbol{I}$ gesetzt (einfache Kollokation). Das unregelmäßig-zufällige Rauschen tritt nur in den Kaufpreisen als Beobachtungen auf. Basierend auf der Gl. 3 ist die Aufgabe der Kollokation die Trennung der beschriebenen Anteile:

- *Ausgleichung:* Die Trendparameter werden so bestimmt, dass sich die Trendfunktion $\boldsymbol{X\beta}$ den Messpunkten möglichst gut anpasst.
- *Filterung:* Für die Kaufpreise sollen die Trendwerte um die Signalwerte $\boldsymbol{s}$ verbessert und das Rauschen herausgefiltert werden, um den unregelmäßigen, aber systematischen Informationsgehalt voll auszuschöpfen.
- *Prädiktion:* Es sollen an Stellen, an denen keine Kaufpreise vorliegen, die Werte des prädizierten Trendanteils $\boldsymbol{X}_p\hat{\boldsymbol{\beta}}_{Koll}$ (mit $\boldsymbol{X}_p$ als Designmatrix der Einflussgrößen der zu prädizierenden Objekte) und das prädizierte Signal $\hat{\boldsymbol{s}}_p$ genutzt werden, um Schätzungen neuer Werte zu ermöglichen.

2.2 Stochastisches Modell

Mathematisch handelt es sich bei dem Schritt der Prädiktion um eine Interpolationsaufgabe, bei der ein Wert an einer beliebigen Stelle aus den vorhandenen umliegenden Größen bestimmt wird. Zu diesem Zweck werden in der Kollokation im Gegensatz zur Regression stochastische Informationen der Beobachtungen genutzt. Die Prädiktion erfolgt bezüglich des Signalanteils über die Korrelationen der Beobachtungen untereinander, welche durch deren Kovarianzen beschrieben werden: Die Beziehung zwischen den Beobachtungen und dem Signal nach Gl. 3 wird allein durch den stochastischen Zusammenhang beschrieben.

Für einen beliebigen zufälligen Beobachtungsvektor $\boldsymbol{y}$ ist die VKM gegeben durch

$$\boldsymbol{\Sigma}_{yy} = \begin{bmatrix} \sigma_1^2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \cdots & \sigma_n^2 \end{bmatrix}. \tag{4}$$

Hierbei bezeichnet σ_i die Standardabweichung der i-ten Beobachtung und ρ_{ik} den Korrelationskoeffizienten zwischen der i-ten und der k-ten Beobachtung. Für die Anwendung im Rahmen der Wertermittlung wird zur Vereinfachung

$$\boldsymbol{\Sigma}_{yy} = \sigma_0^2\boldsymbol{I} \tag{5}$$

(mit dem a priori Varianzfaktor $\sigma_0^2 = 1$) angenommen. Analog kann für das Signal ebenfalls eine VKM $\boldsymbol{\Sigma}_{ss}$ aufgestellt werden, wobei die Verteilungsaussage

$$\boldsymbol{s} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{ss}) \tag{6}$$

gilt. Für die Kreuzkorrelationen des Signals und der VKM der Beobachtungen wird darüber hinaus Unkorreliertheit angenommen ($\boldsymbol{\Sigma}_{ys} = \boldsymbol{\Sigma}_{sy} = \boldsymbol{0}$). Unter diesen Bedingungen und unter Einbeziehung des prädizierten Signals $\boldsymbol{s}_p$ ergibt sich das stochastische Modell der Kollokation zu

$$\bar{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{ss} & \boldsymbol{\Sigma}_{ss_p} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{s_ps} & \boldsymbol{\Sigma}_{s_ps_p} \end{bmatrix} = \bar{\sigma}_0^2\bar{\boldsymbol{Q}}\ , \tag{7}$$

wobei $\bar{\boldsymbol{\Sigma}}$ die VKM, $\bar{\sigma}_0^2$ den a priori Varianzfaktor und $\bar{\boldsymbol{Q}}$ die Kofaktormatrix bezeichnen (Moritz 1980).

Eine wichtige Aufgabe in der Kollokation ist die Besetzung von $\boldsymbol{\Sigma}_{ss}$, der VKM des Signals. Das Signal ist stochastisch abhängig, da es systematische Effekte beinhaltet. Dies kann in der Annahme umgesetzt werden, dass benachbarte Kaufpreise ein gleichwertiges oder ähnliches Verhalten aufweisen, während bei Kaufpreisen, die weiter entfernt voneinander liegen, entsprechend von einem unabhängigen Verhalten ausgegangen werden kann. Wie die Nachbarschaft bzw. der Abstand der Kaufpreise als Beobachtungen definiert werden, erfolgt anhand der Problemstellung. Denkbar sind rein geografische Nachbarschaften, aber auch die Definition von problemorientierten Maßzahlen. Letzterer Weg wird im vorliegenden Ansatz gewählt und im Folgenden erläutert. Die Matrix $\boldsymbol{Q}_{ss}\left(=\frac{1}{\bar{\sigma}_0^2}\boldsymbol{\Sigma}_{ss}\right)$ ergibt sich damit zu

$$\boldsymbol{Q}_{ss} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix}, \tag{8}$$

in der die Korrelationskoeffizienten ρ_{ik} sich auf den Zusammenhang der Signalkomponenten s_i und s_k und damit im vorliegenden Fall und im Vorgriff auf die Erläuterungen im folgenden Kapitel auf die wertmäßige Übereinstimmung zweier Immobilien i und k beziehen. Von besonderer Bedeutung ist in diesem Zusammenhang der Ansatz der Funktion $\rho_{ik} = f(s_i, s_k)$, die diese stochastische Abhängigkeit zutreffend darstellt. Die Lösung des Problems erfolgt über Autokovarianzfunktionen, die das Verhalten des Signals mathematisch approximieren und aus empirischen Daten gewonnen werden können. Die Bestimmung einer geeigneten Autokovarianzfunktion für Anwendungen in der Wertermittlung erfordert aufgrund spezieller Anforderungen eine Modifikation der geodätischen Vorgehensweise nach Möser et al. (2000).

2.3 Ausgleichung und Filterung

Mittels der stochastischen Verknüpfung der Signalanteile lassen sich die Schritte der Ausgleichung und Filterung durchführen. Die Berechnung der gesuchten Koeffizienten erfolgt im Gauß-Helmert-Modell. Für eine tiefer gehende Betrachtung dieses Ausgleichungsansatzes wird an dieser Stelle auf die weiterführende Literatur verwiesen

Tab. 1: Vergleich zwischen Regressions- und Kollokationsansatz in Wertermittlungsanwendungen

	Regressionsansatz	Kollokationsansatz
Funktionales Modell	$y = X\beta + \varepsilon$	$y = X\beta + s + v$
Stochastisches Modell	$\Sigma_{yy} = I = P$	$H = (Q_{yy} + Q_{ss})$
Schätzung	$\hat{\beta}_{Reg} = (X'PX)^{-1} X'Py$	$\hat{\beta}_{Koll} = (X'HX)^{-1} X'Hy$ $\hat{s} = Q_{ss}\hat{k}$ $\hat{s}_p = Q_{s_p s}\hat{k}$
Prädiktion	$\hat{y}_{Reg} = X_p \hat{\beta}_{Reg}$	$\hat{y}_{Koll} = X_p \hat{\beta}_{Koll} + \hat{s}_p$

(vgl. u.a. Niemeier 2008). Die ausgeglichenen Koeffizienten ergeben sich zu:

$$\hat{\beta}_{Koll} = (X^T H^{-1} X)^{-1} X^T H^{-1} y \tag{9}$$

mit $H = Q_{yy} + Q_{ss}$, (10)

wobei $\overline{\sigma}_0^2 = 1$ angenommen wird. Nach der Schätzung des Korrelatenvektors

$$\hat{k} = H^{-1}(y - X\hat{\beta}) \tag{11}$$

ergeben sich die Residuen zu

$$\hat{v} = Q_{yy}\hat{k}\,. \tag{12}$$

Das geschätzte Signal kann ermittelt werden über

$$\hat{s} = Q_{ss}\hat{k}\,. \tag{13}$$

Aufbauend auf diesem Ergebnis kann der Schritt der Prädiktion erfolgen. Hierbei sollen auf Grundlage der Schätzung und Filterung die Werte prädiziert werden, an denen keine Kaufpreise, sondern lediglich die Einflussgrößen gegeben sind. Durch die Beziehung

$$\hat{s}_p = Q_{s_p s}\hat{k} \tag{14}$$

lassen sich die Signalanteile in den Prädiktionspunkten ermitteln. Tab. 1 zeigt zusammenfassend die wesentlichen Unterschiede zwischen dem Modellansatz der klassischen linearen Regressionsanalyse und dem Kollokationsansatz.

2.4 Voraussetzungen für die Anwendung der Kollokation in der Wertermittlung

Die Anwendung der Kollokation stellt grundsätzliche Anforderungen an die Residuen. Dies sind im einzelnen Homogenität, Isotropie und Normalverteilung. Dabei wird der Residuenvektor $\hat{\varepsilon} = y - X\hat{\beta}$ als eine Realisierung eines stochastischen Prozesses aufgefasst, die eine Zufallsstichprobe aus der Grundgesamtheit des gewählten Teilmarktes darstellt (Uhde 1982).

Homogenität bezeichnet in diesem Fall, dass die Verteilungsfunktion der Residuen invariant gegenüber Translationen ist. Dies ist gleichbedeutend mit der Annahme, dass die Varianzen bei beliebig gewählten Stichproben aus der Grundgesamtheit des Teilmarktes im gesamten betrachteten Gebiet gleich bleiben, womit die anzusetzende Autokovarianzfunktion unabhängig vom Ort konstant ist. Der Begriff der Isotropie bezeichnet die Invarianz der Verteilungsfunktion der Residuen gegenüber Rotation, was bedeutet, dass die Funktion richtungsunabhängig gültig ist (Ziegenbein und Hawerk 1978).

Im Gegensatz zur Prüfung der Normalverteilung der Residuen, die im Rahmen der optimierten Regressionsfunktion erfolgt, können die Annahmen über Homogenität und Isotropie in Wertermittlungsanwendungen nur schwer empirisch geprüft werden. Ziegenbein und Hawerk (1978) verweisen auf die generelle Annahme, dass die Bedingungen hinreichend erfüllt sind, sofern die Kaufpreise als Zielgröße normalverteilt sind bzw. in eine Normalverteilung transformiert werden und zusätzlich von der Trendfunktion $\hat{y} = X\hat{\beta}$ angenommen werden kann, dass sie den deterministischen Anteil in den Daten richtig und nahezu vollständig erfasst. Gleichwohl bleibt der empirische Nachweis der Homogenität und Isotropie für Wertermittlungsanwendungen eine offene Frage, die Gegenstand künftiger Forschungen sein wird. Im vorliegenden Fall wird die Erfüllung der Eigenschaften unterstellt.

3 Praktischer Untersuchungsansatz

Im Folgenden soll die Kollokation in der Wertermittlung umgesetzt und anhand eines realen Datensatzes geprüft werden. Als Datengrundlage dient dabei eine Stichprobe der Automatisierten Kaufpreissammlung (AKS) des Gutachterausschusses für Grundstückswerte (GAG) Hannover. Die Untersuchung gliedert sich nach Abb. 1 in die Schritte der Datenvorbereitung, der Schätzung und Filterung sowie der Prädiktion mit anschließender Validierung der Ergebnisse.

3.1 Stichprobenbeschreibung

Als räumlicher Teilmarkt wird die niedersächsische Landeshauptstadt Hannover gewählt. Aufgrund der Datenlage wird der sachliche Teilmarkt der Eigentumswohnungen untersucht. Als Ausnahme finden keine Kauffälle

Berücksichtigung, die im Stadtzentrum gelegen sind, da bei Objekten in zentraler Lage erfahrungsgemäß von erhöhten Kaufpreisen ausgegangen werden kann. Zusätzlich zu den wertrelevanten Merkmalen, die für jeden Kauffall in der Datenbank der AKS enthalten sind, werden ergänzende Daten, insbesondere hinsichtlich der Lagecharakterisierung, erhoben. Insgesamt ergibt sich ein Datensatz von 1.855 vollständig belegten Kauffällen aus den Jahren 2008 bis 2011 mit insgesamt 37 erhobenen, potenziell wertrelevanten Merkmalen.

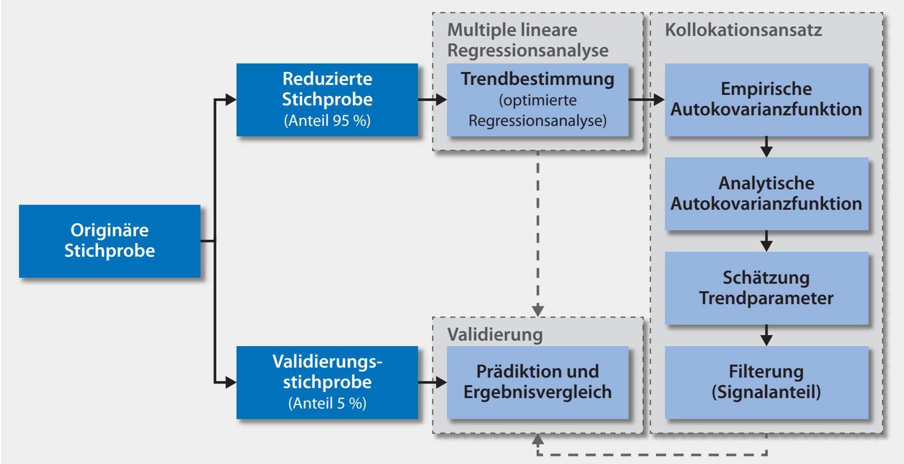

Abb. 1: Praktischer Untersuchungsansatz

Für die vergleichende Beurteilung der Güte der Ergebnisse des Regressions- und des Kollokationsansatzes ist eine zweite Stichprobe erforderlich, die die Prädiktion von Schätzwerten für die Vergleichszwecke ermöglicht. Daher wird das Verfahren der Kreuzvalidierung angewandt: Grundidee dieses Verfahrens ist es, einen Teil der ursprünglichen Stichprobe zu separieren und mit den übrig gebliebenen Daten die Schätzung durchzuführen. Für die anschließende Validierung werden die Vergleichswerte aus der zuvor separierten Stichprobe einmal mit den aus der Regression, einmal mit den aus der Kollokation erhaltenen Ergebnissen geschätzt und den originären Beobachtungen der Validierungsstichprobe gegenübergestellt. Aus der Gesamtstichprobe wird daher per Zufallsprinzip ein für Validierungsstichproben üblicher Anteil von 5 % (93 Kauffälle) separiert (vgl. Abb. 1) (Cressie 1993). Als Kennzahl für die Güte der Ansätze wird basierend auf den Prädiktionsergebnissen die Wurzel des mittleren quadratischen Fehlers (Root Mean Square Error, RMSE) bestimmt (Willmott und Matsuura 2005). Der RMSE ergibt sich zu

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}\ , \qquad (15)$$

wobei n die Größe der Stichprobe (hier: $n = 93$) bezeichnet und für $\hat{y}_i$ jeweils die prädizierten Vergleichswerte der Regressions- oder Kollokationsfunktion eingesetzt werden.

3.2 Bestimmung der Trendfunktion und Prädiktion im Regressionsmodell

Ein weiterer Schritt der Datenvorbereitung nach Abb. 1 ist die Ermittlung einer optimierten Regressionsanalyse mit dem Ziel der Schätzung einer geeigneten Trendfunktion. Durch die Bestimmung der signifikanten Regressionskoeffizienten werden die erhobenen wertbeeinflussenden Merkmale auf einige wenige dominierende Faktoren reduziert. Nach der Reduzierung der Gesamtstichprobe um die Datensätze für die Validierungsstichprobe stehen für die Analyse noch insgesamt 1.762 Kauffälle zur Verfügung. Als Zielgröße wird der Kaufpreis pro Wohn- und Nutzfläche (Euro/m²) verwendet.

Die Bestimmung einer optimalen Regressionsfunktion erfolgt nach den Anforderungen der Wertermittlung, auf die hier nicht weiter eingegangen wird (vgl. u. a. Ziegenbein 1977, Uhde 1982). Durch die angewendete Rückwärtsstrategie werden neun signifikante Einflussgrößen ermittelt. Tab. 2 gibt in der Spalte *Koeffizienten* $\hat{\beta}_{Reg}$ einen Überblick über das Trendmodell als Ergebnis der Regressionsanalyse. Die Stichprobe reduziert sich dabei ein weiteres Mal durch eine Ausreißersuche nach der 2,5σ-Regel im iterativen Regressionsablauf auf eine Anzahl von 1.706 Kauffällen.

Bei der Trendfunktion ist zu beachten, dass es sich um das Ergebnis einer zufälligen Stichprobe handelt, aus der ohne Systematik die Validierungsstichprobe entfernt wurde. Für jede andere Kombination ergeben sich leicht abweichende Werte.

Als signifikante Einflussgrößen für die Beschreibung des gewählten Teilmarktes der Eigentumswohnungen erweisen sich das normierte Datum des Kauffalls *(KD)*, der Bodenrichtwert *(BRW)*, das Alter der Immobilie (*ALTER*, unter Berücksichtigung signifikanter baulicher Veränderungen), die Anzahl der Vollgeschosse *(VGES)*, die kombinierte Wohn- und Nutzfläche *(WOF/NUF)*, die Distanz zum Stadtzentrum *(DISTZENTR)*, die Distanz zur nächsten Grün- und Erholungsfläche *(DISTGRUEN)* sowie als Dummy-Variablen die Verfügbarkeit eines Stellplatzes *(STPLATZ)* und vorhandene Balkone oder Terassen *(BALKON)*. Das Bestimmtheitsmaß der ermittelten Regressionsgleichung liegt mit 0,43 in einer für Wertermittlungsanwendungen üblichen Größenordnung.

Die optimierte Regressionsfunktion kann nun genutzt werden, um die Vergleichswerte der Validierungsstichprobe zu prädizieren, was nach Tab. 1 (Zeile *Prädiktion*) durch das Aufstellen der Designmatrix anhand der signifikanten wertrelevanten Merkmale und der anschließenden Multiplikation mit den geschätzten Regressionskoeffizienten $\hat{\beta}_{Reg}$ erfolgt. Dabei ist zu beachten, dass die Regressionsfunktion nach Tab. 2 nur in den durch die

Wertebereiche der Einflussgrößen gegebenen Grenzen gültig ist. Jeder Prädiktionsdatensatz muss innerhalb dieser Grenzen liegen. Ist dies für einzelne Einflussgrößen nicht der Fall, muss der komplette Datensatz aus der Validierungsstichprobe entfernt werden. Nach dieser Prüfung verbleiben alle Datensätze in der Stichprobe, sämtliche Datensätze liegen innerhalb der Stichprobengrenzen. Für die Validierungsstichprobe wird im Anschluss die jeweilige Zielgröße $\hat{y}_{Reg}$ prädiziert.

Die ermittelte optimierte Trendfunktion wird im Weiteren in den Kollokationsansatz eingeführt, um die bislang nicht ausgeschöpften Informationsanteile zu bestimmen.

Tab. 2: Ergebnisse der Koeffizientenschätzungen in beiden Modellansätzen

Einflussgröße	Koeffizienten $\hat{\beta}_{Reg}$ (Regression)	Koeffizienten $\hat{\beta}_{Koll}$ (Kollokation)	Wertebereich	
Konstante	214,61	308,57	–/–	[–]
KD	53,53	58,67	2008–2012	[–]
BRW	0,85	0,60	74–820	[Euro/m²]
ALTER	–2,01	–1,93	0–119	[Jahre]
VGES	–30,09	–30,03	1–19	[–]
WOF/NUF	4,78	3,97	17–253	[m²]
DISTZENTR	–30,00	–37,17	0,76–9,02	[km]
DISTGRUEN	–99,12	–88,31	0,01–2,17	[km]
STPLATZ	82,24	51,28	$0 \vee 1$	[–]
BALKON	174,16	192,51	$0 \vee 1$	[–]

Stichprobenumfang n = 1.703
Bestimmtheitsmaß B der optimierten Regression = 0,43

3.3 Ableitung der empirischen Autokovarianzfunktion

Nachdem die Daten aufbereitet sind und eine optimierte Trendfunktion mittels des klassischen Regressionsansatzes bestimmt wurde, folgt im nächsten Schritt die Ausgleichung und Filterung im Kollokationsansatz. Wie in Abschnitt 2.2 und im Ablaufschema nach Abb. 1 beschrieben, ist es hierfür notwendig, eine geeignete Autokovarianzfunktion zu ermitteln, die das stochastische Verhalten der Daten zutreffend widerspiegelt. Für Anwendungen in der Wertermittlung muss die Definition dieser Funktion im Gegensatz zur klassischen Zeitreihenanalyse (vgl. Möser et al. 2000) über einen modifizierten Weg erfolgen, da zu jeder Beobachtung eine komplexe Zahl von Einflussgrößen mit unterschiedlichen Einheiten zu berücksichtigen ist. Pelzer (1976) und Ziegenbein (1977) haben das Problem über die Einführung der sogenannten Mahalanobis-Distanz gelöst, die die Verschiedenartigkeit von Grundstücken als einen m-dimensionalen Wertabstand ausdrückt. Die Anzahl m der Dimensionen bezeichnet hierbei die Anzahl der Merkmale, die für eine vorliegende Stichprobe als wertrelevant angesehen werden (im vorliegenden Fall also $m = 9$). Werden die Wertmerkmale zweier Vergleichsobjekte i und k jeweils in einem Vektor zusammengefasst, ergeben sich diese zu $\boldsymbol{x}_i = [x_{i1}\ x_{i2}\ \dots\ x_{im}]$ sowie $\boldsymbol{x}_k = [x_{k1}\ x_{k2}\ \dots\ x_{km}]$. Aus den Vektoren wird eine gewichtete Differenz gebildet, die die Maßzahl δ_{ik}^2 bezeichnet und welche die Verschiedenartigkeit der beiden Objekte ausdrückt, soweit sich diese aus den gewählten Wertmerkmalen ergibt:

$$\begin{aligned}\delta_{ik}^2 &= \left(\boldsymbol{x}_i - \boldsymbol{x}_k\right)\boldsymbol{D}\left(\boldsymbol{x}_i - \boldsymbol{x}_k\right)' \\ &= \Delta\boldsymbol{x}_{ik}\boldsymbol{D}\Delta\boldsymbol{x}_{ik}'.\end{aligned} \tag{16}$$

Die Matrix $\boldsymbol{D}$ wird als Gewichtsmatrix der Dimension $[m \times m]$ eingeführt, die auf der Hauptdiagonalen mit individuellen Gewichten für jede Einflussgröße besetzt ist und damit die Bedeutung der wertrelevanten Merkmale bezeichnet. Die Besetzung der Matrix kann dabei auf verschiedene Arten erfolgen. Im vorliegenden Fall wird analog zu den Untersuchungen von Ziegenbein und Hawerk (1978) der Ansatz gewählt, die partiellen Korrelationskoeffizienten ρ_{yx_j} zwischen der Zielgröße und der j-ten Einflussgrößen als Gewichte einzuführen. Als Ergänzung zu den erhobenen wertrelevanten Merkmalen wird zusätzlich als ein weiteres wertbeeinflussendes Merkmal die geografische Distanz zwischen dem Wertermittlungsobjekt und den Vergleichsobjekten eingeführt. In der Bewertungspraxis ist es üblich, dass geografisch näher liegende Vergleichsobjekte einen höheren Einfluss auf das Wertermittlungsobjekt haben, als weiter entfernt liegende. Durch die Einführung der Distanzen wird diesem Sachverhalt Rechnung getragen. Zudem erhält das Merkmal *Distanz* das Gewicht 1,0, um die besondere Bedeutung der gegenseitigen Lage zu betonen. Die Dimension der Gewichtsmatrix ändert sich damit auf $[m + 1 \times m + 1]$:

$$\boldsymbol{D} = \begin{bmatrix} \rho_{yx_1} & 0 & \dots & 0 \\ 0 & \rho_{yx_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_{yx_{10}} \end{bmatrix}. \tag{17}$$

Für die Berechnung der Maßzahl δ_{ik}^2 nach Gl. 16 werden zudem alle Realisierungen der Einflussgrößen auf den Wertebereich zwischen 0 und 1 normiert:

$$\tilde{x}_{ij} = \frac{x_{ij} - x_j^{min}}{x_j^{max} - x_j^{min}}$$
$$\text{mit}\quad i = 1,\dots,n \ \text{ und } \ j = 1,\dots,m. \tag{18}$$

Die sich dann aus Gl. 16 ergebene Größe $\sqrt{\delta_{ik}^2} = \delta_{ik}$ wird als Wertabstand zwischen zwei Vergleichsobjekten bezeichnet. Mathematisch beschreibt Gl. 16 einen m-dimensionalen Raum, in dem δ_{ik} den Abstand zwischen den Ortsvektoren $\boldsymbol{x}_i$ und $\boldsymbol{x}_k$ ausdrückt. Je geringer dieser Abstand ist, desto geringer fallen auch die Wertunterschiede zwischen den Wertermittlungsobjekten aus. Pelzer (1978) hat gezeigt, dass sich der Korrelationskoeffizient ρ_{ik} (vgl. Gl. 8) zwischen den für die Vergleichsobjekte geltenden Signalen $\boldsymbol{s}_i$ und $\boldsymbol{s}_k$ als Funktion des Wertabstands auffassen lässt:

$$\rho_{ik} = f\left(\delta_{ik}\right), \tag{19}$$

wobei offensichtlich gilt, dass $f(0) = 1$ und $f(\infty) = 0$.

Für die Bestimmung der Werte der empirischen Autokovarianzfunktion werden für jedes Vergleichsobjekt in der Stichprobe die Wertabstände zu allen übrigen Vergleichsobjekten nach Gl. 16 berechnet, woraus sich eine symmetrische Matrix $[n \times n]$ mit n als Anzahl der Datensätze der Stichprobe ergibt (im vorliegenden Fall demnach $[1.706 \times 1.706]$). Die so gebildeten Wertabstände werden in Klassen zusammengefasst, für die nach Sachverstand eine geeignete Klassenbreite d vorzugeben ist. Für die vorliegende Stichprobe werden 16 Wertabstandsklassen als optimale Anzahl ermittelt.

Im Weiteren müssen die linearen Abhängigkeiten der aufeinanderfolgenden Abstandsklassen abgeschätzt werden. Dieser Effekt wird als Erhaltensneigung bezeichnet (Niemeier 2008). Die Tendenz, dass die aufeinanderfolgenden Werte, hier die Klassen, k_i und k_{i+1} stochastisch abhängig voneinander sind, wird durch die Autokovarianzfunktion $C(k)$ ausgedrückt. Für die Berechnung gilt

$$C(k) = \frac{1}{n-k} \sum_{i=1}^{n-k} \varepsilon_i \, \varepsilon_{i+k} \tag{20}$$

mit k = Klasse,
n = Anzahl der Klassenwerte,
ε = Residuen aus Trendbestimmung.

Für die Klasse $k = 0$ entspricht die Berechnung nach Gl. 20 der empirischen Varianz s^2 der Beobachtungen, für $k > 0$ werden die Kovarianzen der Werte, die im klassifizierten Wertabstand d voneinander liegen, berechnet. Durch die Normierung der empirischen Autokovarianzfunktion mit der Varianz $C(0)$ (bzw. s^2) erfolgt der Übergang zur empirischen Autokorrelationsfunktion, welche mit $K(k)$ bezeichnet wird (Niemeier 2008). Der Verlauf der empirischen Autokorrelationsfunktion für die vorliegende Stichprobe kann Abb. 2 entnommen werden. Empirische Autokorrelationsfunktionen bilden die Grundlage für die Bestimmung von analytischen Autokorrelationsfunktionen und somit für die Besetzung der gesuchten Matrix $\boldsymbol{\Sigma}_{ss}$. Anhand des Werteverlaufs in Abb. 2 liegt die Vermutung eines exponentiellen Funktionsverlaufs der Form

$$\rho_{ik} = e^{a\left(\delta_{ik}\right)} \tag{21}$$

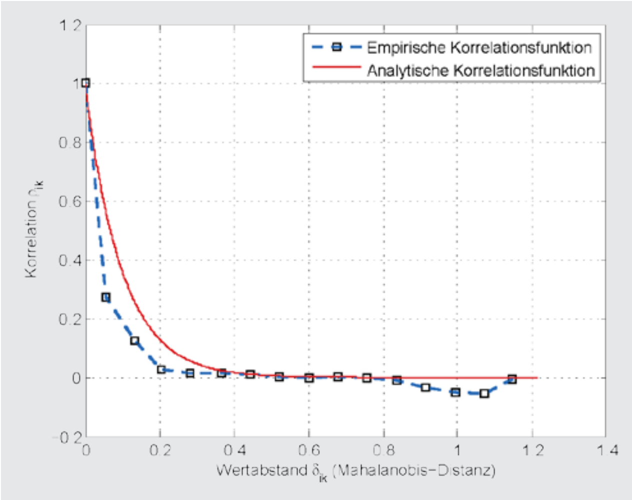

Abb. 2: Verlauf der empirischen und der analytischen Autokorrelationsfunktion

nahe. Dieses Ergebnis entspricht den Erwartungen, da Vergleichsobjekte, die einen geringen Wertabstand aufzeigen, höher miteinander korreliert sind als Objekte, die wertmäßig größere Differenzen aufweisen. Die Autokorrelation muss daher mit steigendem Wertabstand abnehmen. Für den Übergang auf eine analytische Lösung ist der Parameter a in Gl. 21 als funktionsbestimmender Koeffizient mittels einer einfachen nichtlinearen Regressionsanalyse zu schätzen. Für den vorliegenden Datensatz ergibt sich $a = -6{,}88$, wobei für eine bessere Anpassung der analytischen an die empirische Funktion der Wertabstand zusätzlich gewichtet wird:

$$\rho_{ik} = e^{-6{,}88\left(\left(\delta_{ik}\right)/max\left(\delta_{ik}\right)^2\right)}. \tag{22}$$

Die resultierende analytische Autokorrelationsfunktion kann ebenfalls Abb. 2 entnommen werden. Der ermittelte funktionale Zusammenhang ermöglicht nun die Besetzung der Kovarianzmatrix $\boldsymbol{\Sigma}_{ss}$ durch Einsetzen von Gl. 22 in Gl. 8. Die Matrix $\boldsymbol{\Sigma}_{s_p s_p}$ nach Gl. 7 entspricht vom Aufbau her $\boldsymbol{\Sigma}_{ss}$ nach Gl. 8, jedoch werden hier für die Größe δ_{ik} die berechneten Wertabstände zwischen allen zu prädizierenden Objekten eingeführt. Entsprechend werden für die Besetzung von $\boldsymbol{\Sigma}_{ss_p}$ die Wertabstände zwischen den Vergleichsobjekten und den zu prädizierenden Objekten genutzt.

3.4 Prädiktion von Verkehrswerten im Kollokationsmodell

Analog zu dem in Abschnitt 3.2 beschriebenen Vorgehen erfolgt die Prädiktion der Zielgröße im Kollokationsmodell, wobei die signifikanten Einflussgrößen aus der optimierten Regression übernommen werden. Die Berechnung des Trendanteils erfolgt durch die Lösung der Gleichung

$$\hat{\boldsymbol{y}}_{Koll}^{+} = \boldsymbol{X}_p \hat{\boldsymbol{\beta}}_{Koll}, \tag{23}$$

wobei $\hat{\boldsymbol{y}}^{+}_{Koll}$ die prädizierten Beobachtungen, $\boldsymbol{X}_p$ die Einflussgrößen der Validierungsstichprobe und $\hat{\boldsymbol{\beta}}_{Koll}$ die in der Kollokation ermittelten Koeffizienten des Trendanteils bezeichnen (vgl. Tab. 1). Die gegenüber der optimierten Regression neu geschätzten Koeffizienten können Tab. 2 in der Spalte *Koeffizienten* $\hat{\boldsymbol{\beta}}_{Koll}$ *(Kollokation)* entnommen werden. Durch die erneute Schätzung unter Berücksichtigung des stochastischen Anteils in Form des Signals sind zwischen dem Trendmodell der Regression und dem der Kollokation leichte Änderungen zu erkennen. Sind die gesuchten Werte $\hat{\boldsymbol{y}}^{+}_{Koll}$ auf diese Weise geschätzt, werden sie um den prädizierten Signalanteil nach Gl. 14 verbessert:

$$\hat{\boldsymbol{y}}_{Koll} = \hat{\boldsymbol{y}}^{+}_{Koll} + \hat{\boldsymbol{s}}_p \, . \tag{24}$$

3.5 Vergleich der Ergebnisse

Als Ergebnis der verschiedenen Auswertungen können die 93 prädizierten Vergleichswerte der Regression $(\hat{\boldsymbol{y}}_{Reg})$ und der Kollokation $(\hat{\boldsymbol{y}}_{Koll})$ sowie die Originalbeobachtungen $(\boldsymbol{y})$ gegenübergestellt werden. Eine direkte Gegenüberstellung der Schätzwerte und der Originalkaufpreise wird in Abb. 3 gezeigt. Die Kaufpreise der Validierungsstichprobe sind hier in aufsteigender Reihenfolge sortiert und als schwarze Linie dargestellt. Die Ergebnisse der Schätzung mittels Regressionsansatz sind als rote Linie gekennzeichnet, die der Kollokation als blaue Linie.

Für beide Modellansätze ist zu erkennen, dass sie in vielen Fällen stark um die Originalkaufpreise streuen. Dies hängt unmittelbar mit der eingangs erläuterten unzureichenden Möglichkeit der Modellierung der komplexen Wertermittlungsdaten durch das mathematische Modell zusammen. An dieser Stelle wird auch deutlich, dass Regression wie Kollokation lediglich Approximationsmodelle darstellen, beide mathematischen Formulierungen erlauben keine vollständige Erklärung der Kaufpreise. Die maximale Abweichung beträgt für die Regressionsergebnisse ca. 823 Euro/m², die minimale 2 Euro/m², für die Kollokationsergebnisse im Maximum 739 Euro/m² und im Minimum 3 Euro/m². Der direkte Vergleich der beiden Modellansätze zeigt jedoch auch, dass in den meisten Fällen die Schätzungen der Kollokation unterhalb der Regression liegen und somit wie erwartet einen zusätzlichen Informationsanteil ausschöpfen, der für eine bessere Anpassung sorgt. Es ist allerdings zu berücksichtigen, dass die Funktion über einen relativ großen räumlichen Teilmarkt erstellt wurde und demnach eine Auswahl kleinerer Gebiete eine bessere Trendbestimmung und verfeinerte Modellierung der stochastischen Zusammenhänge der Objekte ermöglichen würde. Es ist zu erwarten, dass sich die Reststreuung durch Reduzierung der Stichprobe auf räumliche Teilgebiete verringert. Die Untersuchung der Ergebnisse bei einer weiteren räumlichen Untergliederung wird in künftigen Forschungsansätzen aufgegriffen.

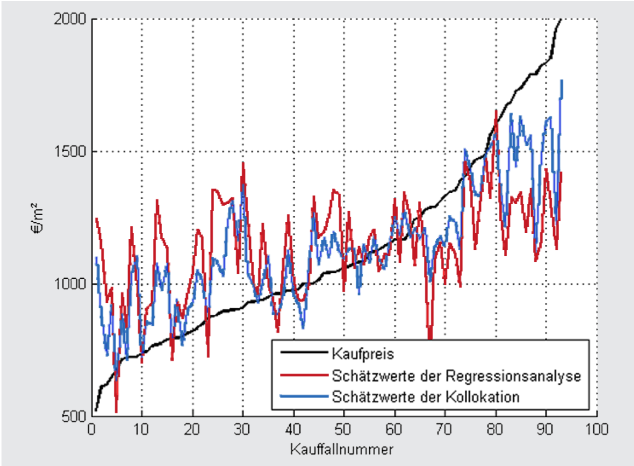

Abb. 3: Vergleich der Ergebnisse der jeweiligen Modellansätze

In Abb. 4 werden die Residuen aus beiden Modellansätzen in Histogrammen dargestellt. Während die Residuen des Regressionsansatzes $(\hat{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{y}}_{Reg} - \boldsymbol{y})$ relativ breit über den Bereich von ca. –650 Euro/m² bis ca. 800 Euro/m² streuen, ist in den Residuen des Kollokationsansatzes $(\hat{\boldsymbol{v}} = \hat{\boldsymbol{y}}_{Koll} - \boldsymbol{y})$ eine Verbesserung zu erkennen: Die Abweichungen liegen zwar ähnlich wie die vorherigen Ergebnisse in einer Spanne von ca. –600 Euro/m² bis ca. 750 Euro/m², konzentrieren sich durch die Ausschöpfung zusätzlicher Information jedoch deutlicher um den Bereich von 0. Die Streuung wird somit auf einen engeren Bereich begrenzt.

In Abb. 5 werden die Residuen der beiden Ansätze als Balkendiagramme in Bezug zur ihrer geografischen Lage im Stadtgebiet Hannover dargestellt: In der linken

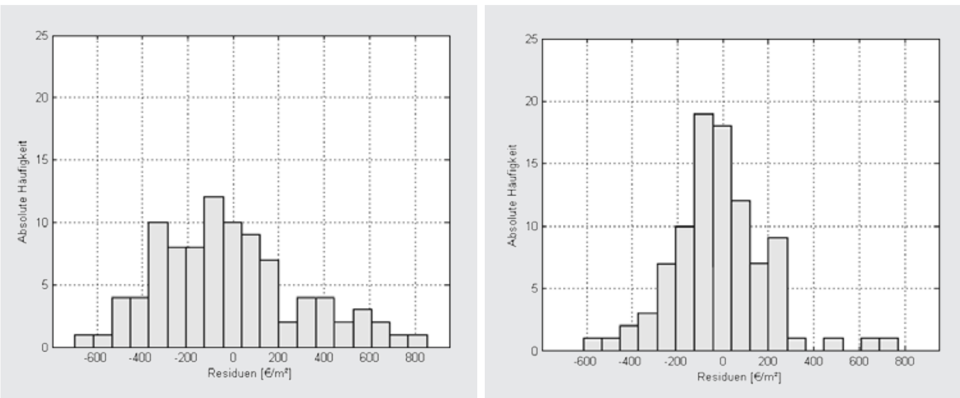

Abb. 4: Histogramm der Residuen im Regressionsansatz (links) und im Kollokationsansatz (rechts)

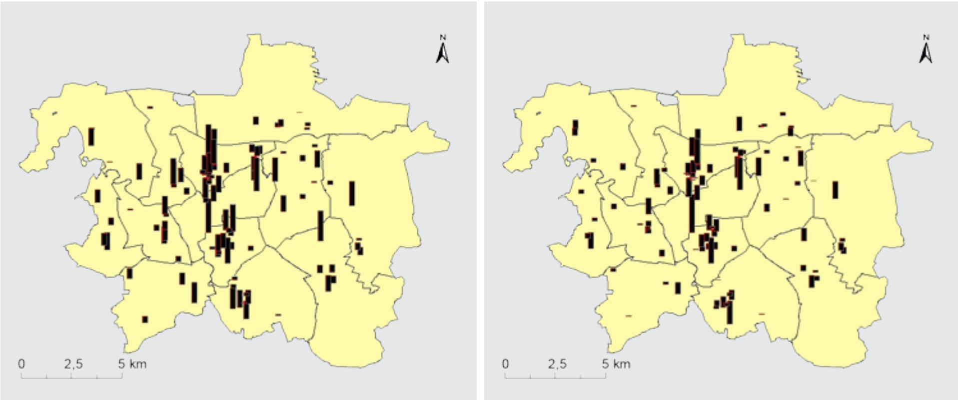

Abb. 5: Vergleich der Residuen des Regressionsansatzes (links) und des Kollokationsansatzes (rechts) für die Daten der Validierungsstichprobe

Darstellung sind die Residuen des Regressionsansatzes zu sehen, in der rechten die des Kollokationsansatzes. Insbesondere in den Randbereichen des Stadtgebietes ist eine Verringerung des nicht informativen Rauschens durch den Kollokationsansatz zu erkennen. Die Fälle, in denen sich nur geringe Änderungen ergeben haben, liegen vornehmlich um den Stadtteil Mitte, der bewusst ausgespart wurde. Ein möglicher Erklärungsansatz ist, dass die Zentrumslage nicht ausreichend repräsentiert wird und der räumliche Teilmarkt an dieser Stelle weiter beschränkt werden müsste. Dies unterstützt die Erkenntnisse aus Abb. 3, wonach wie beschrieben die Wertverhältnisse im Stadtgebiet Hannover nicht optimal durch einen einzigen Trend erfasst werden können, sondern sich in weitere Teilmärkte untergliedern. Jedoch ist festzustellen, dass der Kollokationsansatz bei einem entsprechend gewählten Teilmarkt die Regressionsergebnisse durch die Verringerung des Rauschanteils verbessern kann. Der *RMSE* der Schätzungen mittels Regressionsanalyse liegt bei ca. 318 Euro/m², der des Kollokationsansatzes bei ca. 211 Euro/m². Im Durchschnitt verbessert die Berücksichtigung des Signalanteils den klassischen Ansatz damit um einen weiteren Informationsanteil von etwa 33 %.

Da die Validierungsstichprobe im Rahmen der Kreuzvalidierung per Zufall aus der Gesamtstichprobe entfernt wird, sind die dargestellten Ergebnisse lediglich für diese Stichprobenkombination zutreffend. Um die generelle Verbesserung des statistischen Ansatzes durch die Kollokation zu belegen, wird der vorgestellte methodische Ablauf nach Abb. 1 insgesamt 500-mal wiederholt. Auf diese Weise kann die Aussagekraft einer einzelnen Auswertung auf eine generelle Beurteilung der Kollokation in Wertermittlungsanwendungen ausgeweitet werden. Als Ergebnis der 500 Simulationen werden die RMSE für jeden einzelnen Durchlauf dargestellt (vgl. Abb. 6). Zusätzlich werden die Mittelwerte über die 500 Simulationen dargestellt. Die Ergebnisse bestätigen die bisherigen Aussagen, nach denen von einer generellen Modellverbesserung ausgegangen werden kann. Zwar stellt sich das Ergebnis der erhöhten Informationsausschöpfung von 33 % der beispielhaft vorgestellten Validierungsstichprobe als eines der besseren Ergebnisse dar, der mittlere RMSE verbessert sich durch die Kollokation jedoch von ca. 330 Euro/m² auf ca. 240 Euro/m², die Informationsausschöpfung steigert sich um 27 %. Insgesamt bestätigen die Ergebnisse damit eine generelle Modellverbesserung durch die Berücksichtigung der zusätzlichen stochastischen Komponente im Kollokationsansatz.

4 Fazit und Ausblick

Der Fokus des hier vorgestellten Ansatzes liegt in der Prädiktion von zuverlässigen Vergleichswerten, welche auf der Datengrundlage von einigen wenigen dominierenden Wertmerkmalen beruhen und im Ergebnis bei einem modellierten funktionalen Zusammenhang eine gute Approximation der realen Kaufpreise darstellen. Insgesamt kann für den Teilmarkt der Eigentumswohnungen in Hannover eine Verbesserung in der mathematisch-statistischen Formulierung des Vergleichswertverfahrens erreicht werden. Signifikante Verbesserungen der Regressionsanalyse durch den Kollokationsansatz konnten darüber hinaus auch für den Teilmarkt unbebauter Grundstücke in Hannover festgestellt werden. Die mittlere Verbesserung der Informationsausschöpfung nach den ebenfalls durchgeführten 500 Simulationsläufen betrug ca. 18 % (Zaddach und Alkhatib 2012a). Für die Teilmärkte Ein- und Zweifamilienhäuser sowie Reihen- und Doppelhäuser in Stadt und Landkreis Osnabrück konnte

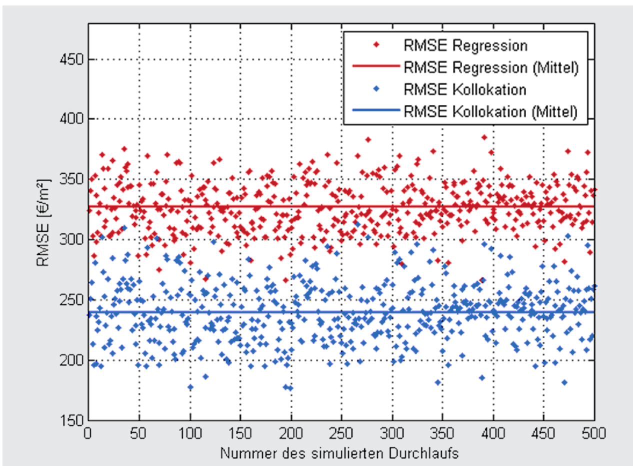

Abb. 6: Darstellung der RMSE der Modellansätze für 500 Simulationsdurchläufe

nach den Simulationen eine zusätzliche Informationsausschöpfung von 27 % erzielt werden (Zaddach und Alkhatib 2012b). Die Ergebnisse der drei Studien belegen, dass die vorgestellte Methode im Vergleichswertverfahren teilmarktunabhängig sinnvoll eingesetzt werden kann und die klassische Regressionsanalyse um einen weiteren Auswerteschritt erweitert. Unter dem Gesichtspunkt eines auch wirtschaftlich vertretbaren Aufwandes bei der Erhebung von Grundstücksdaten kann das Modell der Kollokation damit sinnvoll angewandt werden. Zwar werden die Unterschiede in den Kaufpreisen größtenteils durch einen deterministischen Trendansatz erklärt, doch lassen sich die nicht berücksichtigten Einflüsse teilweise mit der stochastischen Modellkomponente des Signals erfassen und verbessern so den Modellansatz bzw. dessen Erklärungsgehalt. Von Vorteil ist die erstmalige Einführung einer nach Wertermittlungsanforderungen optimierten Regressionsanalyse als ersten Trendansatz der Kollokation, da auf diese Weise die Vorteile der seit den 1970er Jahren als Standard bewährten Vorgehensweise um einen verfeinerten Auswerteansatz ergänzt werden. Als wichtigster Aspekt ist die Erfassung von Anteilen zu nennen, die im Rahmen der Datenerhebung nicht oder nur schwer zu erfassen sind und somit keinen Eingang in die klassische Regressionsanalyse finden können. Die mit Hilfe der Prädiktion im Kollokationsmodell ermittelten Vergleichswerte weichen im Mittel um lediglich knapp 15 % von den Originalkaufpreisen ab, was zwar die in den 1970er Jahren durchgeführten Untersuchungen von Ziegenbein und Hawerk (1978) um knapp 5 % übersteigt, sich jedoch bedingt durch die Wahl des Teilmarktes ergeben kann. Während die damaligen Untersuchungen sich mit der Auswertung unbebauter Grundstücke beschäftigt haben, werden im vorliegenden Fall Kauffälle von Eigentumswohnungen untersucht.

Aufbauend auf den Ergebnissen und der Modellformulierung der Kollokation soll in einem nächsten Schritt die Fortpflanzung der Unsicherheiten der Eingangsgrößen auf das Bewertungsergebnis im Mittelpunkt stehen. Einen besonderen Schwerpunkt bildet dabei die Übertragung der sogenannten Bayes-Theorie – hier in Form einer Bayesischen Kollokation – als Alternativmethode zur klassischen multiplen linearen Regressionsanalyse, wie sie bislang in der Immobilienwertermittlung als Standardmethode Verwendung findet. Von besonderem Vorteil ist in diesem Zusammenhang die Möglichkeit, die gegebene Unsicherheit von Eingangsgrößen im Modell berücksichtigen zu können. Dieser Weg ist der klassischen multiplen linearen Regressionsanalyse verschlossen. Als Ergebnis werden detaillierte Kenntnisse des Unsicherheitshaushaltes erwartet: Wo entstehen Unsicherheiten, welcher Art sind diese und wie beeinflussen sie in ihrer Kombination die Verkehrswerte. Es ist davon auszugehen, dass diese Art von Unsicherheitsmaßen für die Bewertungsergebnisse die Markttransparenz erheblich erhöhen kann.

Literatur

Cressie, N. A. C.: Statistics for Spatial Data. New York, Chichester, Toronto, Brisbane, Singapore: John Wiley & Sons, Inc., 1993.

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

Kleiber, W.: Verkehrswertermittlung von Grundstücken: Kommentar und Handbuch zur Ermittlung von Marktwerten (Verkehrswerten), Versicherungs- und Beleihungswerten unter Berücksichtigung der ImmoWertV. 6. Auflage. Köln: Bundesanzeiger Verlagsges., 2010.

Koch, K.-R.: Statistische Grundlagen zur Untersuchung von Immobilienwerten. In: Schmalgemeier, H. (Hrsg.): Statistische Methoden in der Grundstückswertermittlung. Band 16, Stuttgart: Verlag Konrad Wittwer, 1995, S. 7–12.

Moritz, H.: Advanced Physical Geodesy. Band 13, Sammlung Wichmann. Karlsruhe: Herbert Wichmann Verlag, 1980.

Möser, M., et al. (Hrsg.): Handbuch Ingenieurgeodäsie: Auswertung geodätischer Überwachungsmessungen. Heidelberg: Herbert Wichmann Verlag, 2000.

Niemeier, W.: Ausgleichungsrechnung: Statistische Auswertemethoden. 2. Auflage. Berlin und New York: Walter de Gruyter, 2008.

Pelzer, H.: Multiple Regression. In: Brückner, R. (Hrsg.): Mathematische Statistik bei der Ermittlung von Grundstückswerten. Band 65, Wissenschaftliche Arbeiten der Lehrstühle für Geodäsie, Photogrammetrie und Kartographie an der Technischen Universität Hannover, 1976, S. 153–168.

Pelzer, H.: Ein indirektes Vergleichswertverfahren unter Anwendung statistischer Methoden. ZfV – Zeitschrift für Vermessungswesen, Jg. 103, 1978, Nr. 6, S. 245–254.

Uhde, C.: Mathematische Modelle zur Analyse von Grundstücksmärkten. Band 118, Wissenschaftliche Arbeiten der Lehrstühle für Geodäsie, Photogrammetrie und Kartographie an der Technischen Universität Hannover, 1982.

Zaddach, S.; Alkhatib, H.: Use of Collocation Method as an Improved Sales Comparison Approach. Proceedings of the FIG Working Week 2012, Rome, Italy, May 6–10, 2012a.

Zaddach, S.; Alkhatib, H.: Enhancement of the Classical Regression Analysis by Means of Least Squares Collocation Method. Proceedings of the FIG Commission 9 Workshop – Mass Appraisal Techniques, Paphos, Cyprus, September 14–16, 2012b.

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

Ziegenbein, W.; Hawerk, W.: Erfahrungen bei der Prädiktion von Grundstückswerten. ZfV – Zeitschrift für Vermessungswesen, Jg. 103, 1978, Nr. 6, S. 254–261.

Anschrift der Autoren

Dipl.-Ing. Sebastian Zaddach
Geodätisches Institut der Leibniz Universität Hannover
Flächen- und Immobilienmanagement
Nienburger Straße 1, 30167 Hannover
Tel.: 0511 762-17201, Fax: 0511 762-2468
zaddach@gih.uni-hannover.de

Dr.-Ing. Hamza Alkhatib
Geodätisches Institut der Leibniz Universität Hannover
Geodätische Auswertemethoden
Nienburger Straße 1, 30167 Hannover
Tel.: 0511 762-2464, Fax: 0511 762-2468
alkhatib@gih.uni-hannover.de