Wilfried Mann

Die Regressionsanalyse zur Unterstützung der Anwendung des Normierungsprinzips in der Grundstücksbewertung

Sonderdruck

zfv 5/2005 – 130. Jahrgang

Die Regressionsanalyse zur Unterstützung der Anwendung des Normierungsprinzips in der Grundstücksbewertung

Wilfried Mann

Zusammenfassung

Bei der Verkehrswertermittlung wendet der Sachverständige im Rahmen des Vergleichswertverfahrens das Normierungsprinzip an. Normierungsfaktoren für wesentliche Einflussmerkmale auf den Kaufpreis lassen sich im Prozess einer multiplen Regressionsanalyse ableiten. Durch eine partielle Auflösung der Regressionsgleichung werden multiple Marktzusammenhänge zweidimensional darstellbar. Die Ergebnisse dienen dem Sachverständigen als Wissensbasis. Normierungsfaktoren können in den Geschäftsstellen der Gutachterausschüsse für Grundstückswerte zur automatisierten Ableitung z.B. von Boden- und Marktrichtwerten herangezogen und nach deren Veröffentlichung auch zur Verkehrswertableitung aus Richtwerten verwendet werden.

Summary

During the determination of the current market value, the market expert operates with the standardisation principle within the scope of the reference value method. Standardisation factors for essential interventions in regard to the purchase price can be derived through the process of multiple regression analysis. Market contexts can be presented two-dimensionally because of the partial solution of the regression equation. Experts use the results as base for their knowledge. Standardisation factors can be used by the advisory committees for real estate values for an automated deduction of ground and market indicatory values. After their publication they can also be used for current market value deductions from indicatory values.

1 Einführung

Bei der Bewertung von bebauten und unbebauten Grundstücken bedient sich der Sachverständige der klassischen Wertermittlungsmethoden, dem Sachwert-, dem Ertragswert- sowie dem Vergleichswertverfahren. Anhand seiner Sachkenntnis und Erfahrung wird er den Verkehrswert aus den Rechenergebnissen dieser Verfahren unter Berücksichtigung der Lage auf dem Grundstücksmarkt ableiten.

Hierbei spielt das Vergleichswertverfahren eine besondere Rolle. Allgemein gilt, dass ein Verkehrswert aus dem Vergleich von Kaufpreisen für Immobilien ermittelt wird, die hinsichtlich ihrer preisbestimmenden Merkmale mit dem zu bewertenden Objekt übereinstimmen. Da ein direkter Vergleich in der Regel ausscheidet – mehrere Grundstücke mit vollständig gleicher Qualität bei übereinstimmenden allgemeinen Wertverhältnissen gibt es i.d.R. nicht – kommt in der Praxis der indirekte Vergleich zur Anwendung.

Bei der Anwendung des indirekten Vergleichswertverfahrens muss sich der Sachverständige über einen marktgängigen Maßstab Gedanken machen, mit dem er die Objekte vergleichen will[1]. Weiter wird der Experte mehr oder weniger intuitiv über Informationen aus dem Grundstücksmarkt einen Bewertungsrahmen, eine Bewertungsschablone entwickeln. Anhand eines multiplen Abgleichs der Vergleichspreise mit dem Bewertungsobjekt nach Lage, Ausstattung, Baujahr usw. wird er dann einen Verkehrswert ableiten.

Es ist nicht einfach, dieses Expertenwissen in Zahlen so auszudrücken, dass es auf Kaufpreise angewendet werden kann. In der Praxis der Grundstücksbewertung hat sich bei der Systematisierung dieses Problems das so genannte »Normierungsprinzip« entwickelt.

2 Normierungsprinzip

Das Normierungsprinzip sieht in der Anwendung vor, dass ein Kaufpreis mit seinen wertbildenden Merkmalen (Kreis) so umgerechnet wird, als wäre ein Preis für ein Norm-Objekt gezahlt worden (Vollkreis). Das arithmetische Mittel führt dann in der Regel zum Normwert (Abb. 1).

Deckt sich das Norm-Objekt mit den Merkmalen des Bewertungsobjektes, dient der Normwert als Basis für den Verkehrswert. Andererseits kann der Normwert auch zum

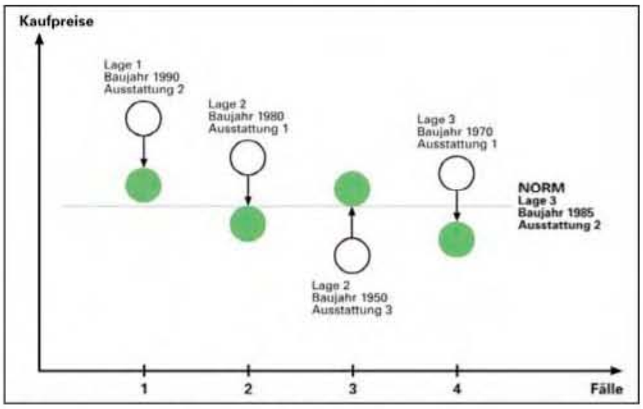

Abb. 1: Prinzipskizze zum Normierungsprinzip

1 Für die weiteren Überlegungen in diesem Fachbeitrag wird der Maßstab Kaufpreis pro m² Wohn-/Nutzfläche eingeführt. Dieser Maßstab bedingt neben der Marktüblichkeit eine höchstmögliche Menge von wertrelevanten Ausprägungen.

Richtwert führen, wenn die typisierten Merkmale des Richtwertgrundstücks Grundlage sind.

Dieses Prinzip erlaubt dem Sachverständigen, seine intuitiven Markterkenntnisse in strukturierten Mustern abzulegen. Die Aufgabe besteht im Prinzip darin zunächst kognitiv herauszufinden,

1. welche Merkmale den Kaufpreis signifikant beeinflussen. Diese **qualitative Fragestellung** wird als Antwort nur relativ wenige Einflussgrößen zulassen, da ein Experte vor Ort in seiner intuitiven Wertabschätzung begrenzt ist.
2. wie groß die Unterschiedsausprägungen sind zwischen den als einflussreich erkannten, wesentlichen Merkmalen.
   Diese **quantitative Fragestellung** führt in der Regel zu Ergebnissen in Prozent, die ein Sachverständiger auf einen Bezugswert anwenden kann.

Vergleicht man diese Lösungsstrukturen mit Ergebnissen, die aus der Kaufpreissammlung mithilfe mathematisch-statistischer Methoden abgeleitet werden können, so wird erkennbar, dass die multiple lineare Regressionsanalyse adäquate Ergebnisse liefert. Zum einen erlaubt die Varianzanalyse einen Schluss auf signifikante Merkmale und beantwortet somit die Frage nach der Menge der erforderlichen Merkmale, zum anderen enthält die Regressionsgleichung alle Merkmale zur Beantwortung der quantitativen Fragestellung. Das Bewertungsmuster des Sachverständigen kann somit durch Auswertungsergebnisse aus der Kaufpreissammlung mithilfe mathematisch-statistischer Methoden gestützt werden.

Hierbei ist allerdings zu bedenken, dass das Ergebnis, als Regressionsgleichung einer linearen multiplen Regressionsanalyse dargestellt, von einem Grundstückssachverständigen in der Regel nicht marktorientiert interpretiert werden kann. Um dennoch die statistischen Ergebnisse dem Sachverständigen so aufzuschlüsseln, dass er überprüfbare Normierungsfaktoren oder -funktionen erhält, schlägt der Verfasser vor, die Regressionsgleichung weiter aufzulösen. Die Ergebnisse der sog. »partiellen Modellauflösung«, die besonders die Skalierung der unabhängigen Variablen berücksichtigt, passen sich den Bewertungsschablonen der Sachverständigen weitgehend an. Sie können als Bausteine in den Prozess der Regressionsanalyse eingefügt und auch bei der automatisierten Anwendung zum Beispiel bei der Ableitung von Boden- und Marktrichtwerten[2] verwendet werden.

3 Multivariates Auswerteverfahren

Teile dieses iterativen Prozesses, in dem die multiple lineare Regression einen zentralen Platz einnimmt, wurden bereits im Jahr 1978 von Ziegenbein[3] beschrieben. Verschiedene auf dem Markt angebotene Softwareprodukte unterstützen die Regressionsanalysen und haben dazu geführt, dass viele Geschäftsstellen der Gutachterausschüsse mit diesem Handwerkszeug arbeiten. Die folgenden Ausführungen gehen auf die allgemeinen Bedingungen der Regressionsanalyse ein und beleuchten kritisch die stochastischen Merkmale von Kaufpreisen. Ein Schwerpunkt ist die ausführliche Darstellung der partiellen Auflösung der Regressionsgleichung und der Ergebniseinbettung in den Gesamtprozess der multivariaten Kaufpreisauswertung bei großen Teilmarktstichproben.

3.1 Grundsätze der Regressionsanalyse

Die lineare multiple Regressionsanalyse, als Sonderfall des Ausgleichungsmodells nach Gauß-Markow (Ausgleichung nach vermittelnden Beobachtungen), basiert auf folgenden klassischen Voraussetzungen:

- lineares Modell
- unabhängige Beobachtungen
- Varianzhomogenität der unabhängigen Beobachtungen
- Normalverteilung der Ziel- und Einflussgrößen.

Die Bedingungen sind:

- Regressionsforderung: Kleinste-Quadrate-Prinzip nach Gauß
- Voraussetzung: Normalverteilung der Residuen (Rest-Abweichungen).

Die Regressionsgleichung folgt folgender Vorschrift:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \dots \beta_k X_k + \varepsilon . \qquad (1)$$

Hierbei ist
Y abhängige Variable,
X_i unabhängige Variablen (Regressoren),
k Anzahl der unabhängigen Variablen,
β_0 Achsenabschnitt (Intercept Parameter),
β_i Parameter der unabhängigen Variablen,
ε Abweichung (Residuum).

Ein F-Test testet die Hypothese, dass alle oder einige Parameter gleich 0 sind, d. h.:

Nullhypothese:
H_0: $\beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$ oder
Alternativhypothese:
H_A: $\beta_1 \neq 0$ oder $\beta_2 \neq 0$ oder … $\beta_k \neq 0$.

Prüft man diese Voraussetzungen anhand der teilmarkttypischen Datenstrukturen in der Kaufpreissammlung, dann kann man feststellen:

2 Marktrichtwerte sind Richtwerte, die im Maßstab Kaufpreis pro Wohnfläche für die Hauptteilmärkte unbebaute Baulandgrundstücke, bebaute Objekte und Eigentumswohnungen abgeleitet und in Düsseldorf erstmals zum 31.12.1998 in einer Karte veröffentlicht wurden.

3 Ziegenbein, Werner: Multiple Regression, in: Brückner (Hrsg.) 1978, Anwenderseminar.

- *Linearität* liegt in der Regel nicht vor, kann aber als Polynom, durch Einsetzen von z. B. quadratischen Gliedern einer Ausprägung in die Regressionsgleichung, erreicht werden.
- *Stochastische Unabhängigkeit* kann mithilfe von Tests auf Autokorrelation (Korrelations-Matrix) überprüft, aber aufgrund der »Kaufpreisrealität« nur bedingt erreicht werden.
- *Varianzhomogenität* ist nur bedingt gegeben. Zu beachten ist, dass die unabhängigen Variablen in der Regel verschiedenartigen Messstrukturen/-skalen entnommen werden. Um eindeutige Rechengrößen zu erzeugen, können so genannte Dummy-Variablen (Ja-Nein-Schein-Variable) eingesetzt werden, durch die die ordinal- und nominalskalierten Variablen auch homogen verarbeitet werden können.
- *Normalverteilung* liegt in der Regel nicht vor, da Kaufpreise eher logarithmisch verteilt sind. Bei großen Stichproben kann allerdings unter der Prämisse des Zentralen Grenzwertsatzes Normalverteilung unterstellt werden. Normalverteilung ist auch explizit für die Auflösung im Rahmen des allgemeinen Ausgleichungsprinzips nicht notwendig. Für eine korrekte Modellierung ist dagegen die Normalverteilung der Residuen von entscheidender Bedeutung.

Kritische Anmerkungen zur Anwendung der Regressionsanalyse bei Kaufpreisen

Zusammenfassend bleibt kritisch zu bemerken, dass die Anwendung der klassischen Regressionsanalyse auf Kaufpreise Fragen offen lässt.

Es bleibt unklar, ob es sich bei Kaufpreisen generell um statistische Daten handelt. Reuter (1994) schreibt hierzu: »Kaufpreise sind historische Ereignisse. Der Kauf eines Grundstücks ist als Experiment nicht wiederholbar. Die Zufallsvariable Kaufpreis ist eine irreale Hypothese.«

Weiter ist die Verkehrswertableitung unter Einsetzen von Werten in eine Regressionsgleichung in Frage zu stellen. Grundsätzlich gilt: Aus einer Grundgesamtheit wird eine Zufallsstichprobe gezogen. Im Rahmen von Stichprobenberechnungen können dann Rückschlüsse auf Grundgesamtheitsparameter, als Kennwerte der Merkmalsverteilungen, gezogen werden. Nichtzufallsstichproben erlauben keine Verallgemeinerungen. Generell ist der Rückschluss auf den Einzelfall verboten (Abb. 2).

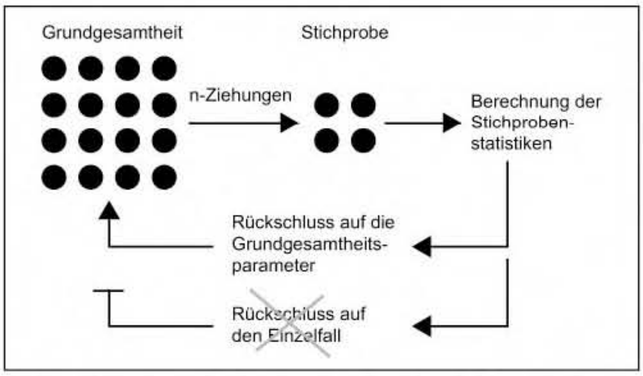

Abb. 2: Von der Grundgesamtheit zur Stichprobe und deren Rückschlüsse[4]

Neben dieser mathematisch-statistisch begründeten Infragestellung ist für den Sachverständigen eine Interpretation des so errechneten Wertes nicht möglich; er ist somit aus der Sicht des Experten abzulehnen.

Die Auflösung der Regressionsgleichung im Rahmen der partiellen Modellauflösung eröffnet dagegen die Möglichkeit einer sachverständigen Beurteilung und genügt dem statistischen Prinzip der Übertragbarkeit von Grundsatzerkenntnissen aus der Stichprobenberechnung auch bei geschwächter stochastischer Interpretation.

3.2 Partielle Modellauflösung

Um zu Transparenz und Nachvollziehbarkeit der Ergebnisse aus der multiplen linearen Regressionsanalyse beizutragen, wird das Ergebnis aus der Regressionsgleichung weiter aufgelöst. Zuvor sollte das Regressionsmodell mit bester Anpassung und plausiblen Ergebnissen entwickelt sein.

Aus Gleichung (1) abgeleitet, wobei die jetzige Schreibweise die Parameterschätzungen meint, die den Bedingungen der Regressionsanalyse unterworfen wurden, ergibt sich:

$$\hat{Y} = \beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 + \beta_3 \bar{x}_3 + \beta_4 \bar{x}_4 + \ldots + \beta_k \bar{x}_k \,. \qquad (2)$$

Die $\bar{x}_{1,2\ldots k}$ können den unterschiedlichsten Skalen zugeordnet sein. Deshalb wird Gleichung (2) nach der Skalierung geordnet und mit neuer Schreibweise wie folgt weiter aufgelöst. Hierzu werden die Buchstaben

v für verhältnisskalierte,
o für ordinalskalierte und
n für nominalskalierte

Ausprägungen der unabhängigen Variablen eingeführt.

$$\begin{aligned}\hat{Y} = {} & \beta_0 \\ & + {}_v\beta_1 \cdot {}_v\bar{x}_1 + {}_v\beta_2 \cdot {}_v\bar{x}_2 + \ldots + {}_v\beta_u \cdot {}_v\bar{x}_u \\ & + {}_o\beta_1 \cdot {}_o\bar{x}_1 + {}_o\beta_2 \cdot {}_o\bar{x}_2 + \ldots + {}_o\beta_p \cdot {}_o\bar{x}_p \\ & + {}_n\beta_1 \cdot {}_n\bar{x}_1 + {}_n\beta_2 \cdot {}_n\bar{x}_2 + \ldots + {}_n\beta_m \cdot {}_n\bar{x}_m \end{aligned} \qquad (3)$$

Hierbei ist u Anzahl aller verhältnisskalierten Variablen,
p Anzahl aller ordinalskalierten Variablen,
m Anzahl aller nominalskalierten Variablen.

Verhältnisskalierte Variablen können als Exponentialfunktion ($Y = a^X$) durchaus die beste Modellanpassung liefern. Im linearen Regressionsmodell kann eine optimale Anpassung auch durch ein Polynom ($Y = X + X^2 + X^3 + \ldots$) erfolgen. Somit können sich zusammengehörige Gruppen im Regressionsansatz befinden, die nur einen Werteinfluss beschreiben:

4 Schnell/Hill/Esser 1995, Abb. 6-3, Grundgesamtheit und Stichprobe, S. 260.

Beispiel: $Y_{(Kaufpreis)} = \beta_0 + \beta_1 \cdot X_{(GFZ)} + \beta_2 \cdot X_{(GFZ^2)}$

Eine vergleichbare Gruppenzuordnung findet sich bei den ordinalskalierten Variablen, die jeweils DUMMY-codiert sind:

Beispiel: $Y_{(Kaufpreis)} = \beta_0 + \beta_1 \cdot X_{(D\_Wolage\ 1)} + \beta_2 \cdot X_{(D\_Wolage\ 2)} + \beta_3 \cdot X_{(D\_Wolage\ 3)} + ...$

Diese Zusammenhänge sind bei der Auflösung zu berücksichtigen und führen, ausgehend von Gleichung (3), zu Gleichung (4) .

$$\begin{aligned}
\hat{Y} = \beta_0 & \\
& + {}_v\beta_{11} \cdot {}_v\overline{x}_{11} + {}_v\beta_{12} \cdot {}_v\overline{x}_{12} + \ldots + {}_v\beta_{1u} \cdot {}_v\overline{x}_{1u} && \text{(verhältnisskaliert, Gruppe 1)} \\
& + {}_v\beta_{21} \cdot {}_v\overline{x}_{21} + {}_v\beta_{22} \cdot {}_v\overline{x}_{22} + \ldots + {}_v\beta_{2u} \cdot {}_v\overline{x}_{2u} && \text{(verhältnisskaliert, Gruppe 2)} \\
& + \ldots \\
& + {}_v\beta_{z1} \cdot {}_v\overline{x}_{z1} + {}_v\beta_{z2} \cdot {}_v\overline{x}_{z2} + \ldots + {}_v\beta_{zu} \cdot {}_v\overline{x}_{zu} && \text{(verhältnisskaliert, Gruppe z)} \\
& + \ldots \\
& + {}_o\beta_{11} \cdot {}_o\overline{x}_{11} + {}_o\beta_{12} \cdot {}_o\overline{x}_{12} + \ldots + {}_o\beta_{1p} \cdot {}_o\overline{x}_{1p} && \text{(ordinalskaliert, Gruppe 1)} \\
& + {}_o\beta_{21} \cdot {}_o\overline{x}_{21} + {}_o\beta_{22} \cdot {}_o\overline{x}_{22} + \ldots + {}_o\beta_{2p} \cdot {}_o\overline{x}_{2p} && \text{(ordinalskaliert, Gruppe 2)} \\
& + \ldots \\
& + {}_o\beta_{r1} \cdot {}_o\overline{x}_{r1} + {}_o\beta_{r2} \cdot {}_o\overline{x}_{r2} + \ldots + {}_o\beta_{rp} \cdot {}_o\overline{x}_{rp} && \text{(ordinalskaliert, Gruppe r)} \\
& + \ldots \\
& + {}_n\beta_1 \cdot {}_n\overline{x}_1 + {}_n\beta_2 \cdot {}_n\overline{x}_2 + \ldots + {}_n\beta_m \cdot {}_n\overline{x}_m && \text{(nominalskaliert)} \qquad (4)
\end{aligned}$$

Hierbei ist
- u Anzahl der verhältnisskalierten Variablen in der Gruppe,
- z Anzahl der verhältnisskalierten Gruppen,
- p Anzahl der ordinalskalierten Variablen in der Gruppe,
- r Anzahl der ordinalskalierten Gruppen,
- m Anzahl der nominalskalierten Variablen.

Somit gilt: $k_{(Anzahl\ aller\ Ausprägungen)} = \sum_{i=1}^{z} u_i + \sum_{i=1}^{r} p_i + m$.

Die Berechnung des Schätzwertes $\hat{Y}$ ergibt sich nach der Regressionsgleichung aus der Summe der arithmetischen Mittelwerte $\overline{x}_i$ der jeweiligen Ausprägungen, multipliziert mit den jeweiligen Parameterschätzern β_i und addiert zum Achsenabschnitt β_0. Gleichzeitig ist der Schätzwert $\hat{Y}$ auch das arithmetische Mittel der abhängigen Variablen Y.

$$Y = \hat{Y} = \beta_0 + \sum_{i=1}^{k} \beta_i \cdot \overline{x}_i \qquad (5)$$

Wenn das Produkt aus β_i und $\overline{x}_i$ mit P_i bezeichnet wird, dann ergibt sich

$$Y = \hat{Y} = \beta_0 + \sum_{i=1}^{k} P_i \ . \qquad (6)$$

Das Ziel der weiteren Modellauflösung besteht darin, den Unterschied zwischen den jeweiligen Ausprägungen/Ausprägungsgruppen zu ermitteln.

Bei den nominalskalierten Variablen gilt es festzustellen, wie groß der Unterschied zwischen dem vorhandenen Merkmal und dem Nichtvorhandensein ist. Das heißt, es ist die Regressionsgleichung dergestalt aufzulösen, dass z.B. für das Merkmal »als Abbruchobjekt erworben« im ersten Rechengang der Wert $X_{(Abbruch)} = 0$ und im zweiten der Wert $X_{(Abbruch)} = 1$ gesetzt wird.

Aus dem Vergleich beider Rechenergebnisse lassen sich dann Unterschiede in Euro oder in Prozent ableiten, die sachverständig überprüft werden können.

Das gleiche Prinzip lässt sich auf die ordinalskalierten und auch verhältnisskalierten Variablen anwenden, wobei die Ergebnisse zu Gruppenunterschieden bzw. zu zweidimensionalen Funktionsgleichungen führen.

3.2.1 Auflösung der nominalskalierten Werte

Wie zuvor beschrieben, wird die Regressionsgleichung zunächst bei der Annahme DUMMY = 0, der Einfluss ist **nicht vorhanden**, aufgelöst.

Die zu untersuchende nominalskalierte Ausprägung sei ${}_nX_1$ mit ${}_n\beta_1 \cdot {}_n\overline{x}_1 = {}_nP_1$ (vgl. Gleichung (4) und (6)).

Es errechnet sich der Wert, mit ${}_nY_{NO\,1}$ bezeichnet, wenn die Ausprägung ${}_nX_1$ nicht vorhanden wäre, zu

$${}_nY_{NO\,1} = \hat{Y} - {}_n\beta_1 \cdot {}_n\overline{x}_1 = \hat{Y} - {}_nP_1 \ . \qquad (7)$$

Bei der Annahme DUMMY = 1, der Einfluss ${}_nX_1$ ist **vorhanden**, gilt

$${}_n\beta_1 \cdot {}_n\overline{x}_1 = {}_n\beta_1, \text{ da } {}_n\overline{x}_1 = 1 \ .$$

Somit errechnet sich der Wert ${}_nY_1$ zu

$${}_nY_1 = {}_nY_{NO\,1} + {}_n\beta_1 \ . \qquad (8)$$

Alle m nominalskalierten Ausprägungen sind analog Gleichung (7) und (8) aufzulösen. Abb. 3 verdeutlicht das Prinzip.

3.2.2 Auflösung der ordinalskalierten Werte

Ordinalskalierte Ausprägungen entstammen den unterschiedlichen Ausprägungen einer Variablen, zum Beispiel der Wohnlage. Das heißt, alle Wohnlagezuordnungen

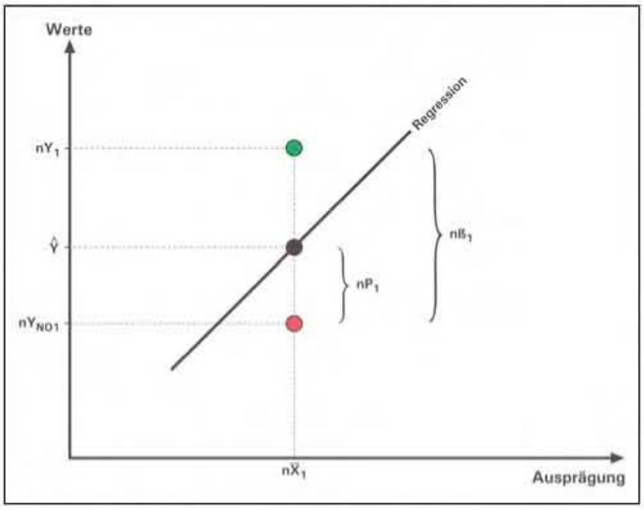

Abb. 3: Prinzip zur Auflösung nominalskalierter Ausprägungen

sind zunächst in der Summe zu untersuchen. Die erste Gruppe der ${}_{o}X_{11\ldots1p}$ sei nach Gleichung (4)

$$ {}_{o}\beta_{11}\cdot{}_{o}\overline{X}_{11}+{}_{o}\beta_{12}\cdot{}_{o}\overline{X}_{12}+\ldots+{}_{o}\beta_{1p}\cdot{}_{o}\overline{X}_{1p} $$
(ordinalskaliert, Gruppe 1)

oder in anderer Schreibweise nach Gleichung (6)

$$ {}_{o}P_{11}\cdot{}_{o}P_{12}+\ldots+{}_{o}P_{1p} \qquad \text{(ordinalskaliert, Gruppe 1)}\,. $$

Bei der Annahme, dass der Gesamteinfluss der Wohnlage = 0 ist, ergibt sich

$$ {}_{o}Y_{NO\,1}=\hat{Y}-({}_{o}P_{11}+{}_{o}P_{12}+\ldots+{}_{o}P_{1p})\,. \tag{9} $$

Bei der Annahme, ein Einfluss der Gruppe ist vorhanden, oder ${}_{o}\overline{X}_{11}$ oder ${}_{o}\overline{X}_{12}$ oder $\ldots{}_{o}\overline{X}_{1p}=1$ gilt

$$ {}_{o}\beta_{11}\cdot{}_{o}\overline{X}_{11}={}_{o}\beta_{11} \text{ oder } {}_{o}\beta_{12}\cdot{}_{o}\overline{X}_{12}={}_{o}\beta_{12} \text{ oder} $$
$$ {}_{o}\beta_{13}\cdot{}_{o}\overline{X}_{13}={}_{o}\beta_{13} \text{ usw.} $$

Somit errechnen sich für jede Ausprägung der Wohnlage die Werte ${}_{o}Y_{11,12\ldots1p}$ zu

$$ \begin{aligned} {}_{o}Y_{11}&={}_{o}Y_{NO\,1}+{}_{o}\beta_{11}\\ {}_{o}Y_{12}&={}_{o}Y_{NO\,1}+{}_{o}\beta_{12}\\ \text{''}\quad&\quad\text{''}\qquad\text{''}\\ {}_{o}Y_{1p}&={}_{o}Y_{NO\,1}+{}_{o}\beta_{1p}\,. \end{aligned} \tag{10} $$

Alle r ordinalskalierten Gruppen sind für Merkmal nicht vorhanden analog Gleichung (9) und für Merkmal vorhanden je Ausprägung gemäß Gleichung (10) aufzulösen (Abb. 4).

3.2.3 Auflösung der verhältnisskalierten Werte

Das Polynom der verhältnisskalierten Variablen, z.B. der Geschossflächenzahl (GFZ), kann als Schätzfunktion ${}_{v}\hat{Y}_{1}=f({}_{v}\overline{X}_{1})$ aufgelöst werden.

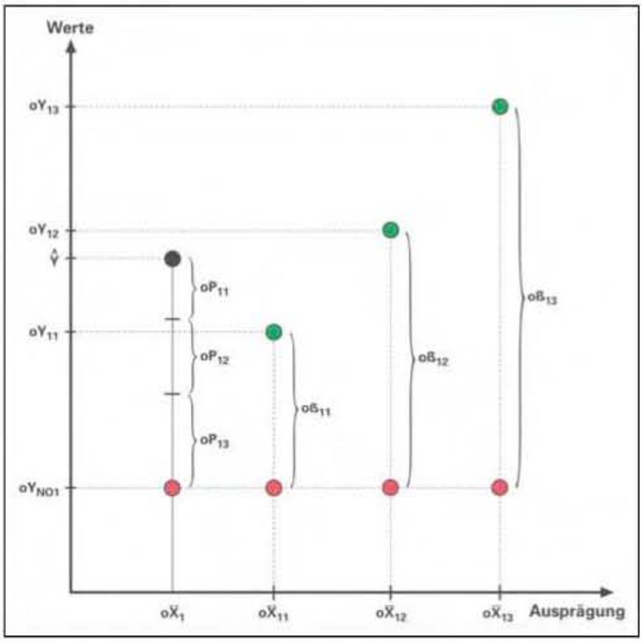

Abb. 4: Prinzip zur Auflösung ordinalskalierter Ausprägungen

Wenn alle anderen Einflussgrößen auf »Null«, d.h. nicht vorhanden, gesetzt sind, gilt für die Gruppe der ${}_{v}X_{11\ldots1u}$ nach Gleichung (4)

$$ {}_{v}\hat{Y}_{1}=\beta_{0}+{}_{v}\beta_{11}\cdot{}_{v}\overline{X}_{11}+{}_{v}\beta_{12}\cdot{}_{v}\overline{X}_{12}+\ldots+{}_{v}\beta_{1u}\cdot{}_{v}\overline{X}_{1u} $$
(verhältnisskaliert, Gruppe 1).

Allgemein, für jede verhältnisskalierte Variable lässt sich einsetzen

$$ \begin{aligned} {}_{v}\hat{Y}_{1}&=\beta_{0}+{}_{v}\beta_{11}\cdot{}_{v}\overline{X}_{11}+{}_{v}\beta_{12}\cdot{}_{v}\overline{X}_{12}+\ldots+{}_{v}\beta_{1u}\cdot{}_{v}\overline{X}_{1u}\\ {}_{v}\hat{Y}_{2}&=\beta_{0}+{}_{v}\beta_{21}\cdot{}_{v}\overline{X}_{21}+{}_{v}\beta_{22}\cdot{}_{v}\overline{X}_{22}+\ldots+{}_{v}\beta_{2u}\cdot{}_{v}\overline{X}_{2u}\\ \text{''}\quad&\quad\text{''}\quad\text{''}\quad\text{''}\qquad\text{''}\quad\text{''}\qquad\quad\text{''}\quad\text{''}\\ {}_{v}\hat{Y}_{z}&=\beta_{0}+{}_{v}\beta_{z1}\cdot{}_{v}\overline{X}_{z1}+{}_{v}\beta_{z2}\cdot{}_{v}\overline{X}_{z2}+\ldots+{}_{v}\beta_{zu}\cdot{}_{v}\overline{X}_{zu}\,. \end{aligned} \tag{11} $$

Einflussgrößen, die intervallskaliert vorliegen, wie z.B. das Jahr des Vertragsabschlusses, können ordinalskaliert – jahrweise gruppiert – oder auch verhältnisskaliert – Datum als Dezimaljahr, aus Jahr, Monat und Tag abgeleitet – verarbeitet werden (Abb. 5).

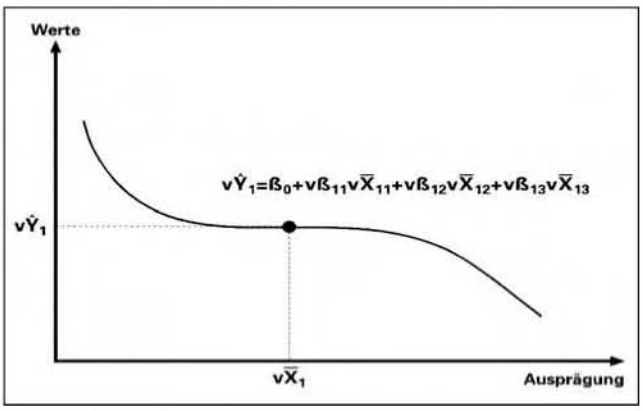

Abb. 5: Prinzip der Darstellung funktionaler Zusammenhänge

3.2.4 Ermittlung von Unterschiedswerten/-faktoren

Aus den zuvor ermittelten Werten mit den Einflüssen bei DUMMY = 0 oder DUMMY = 1 lassen sich Unterschiedswerte der Ausprägungen zum Normalfall bzw. zur typischen Ausprägung der Stichprobe ableiten. Hierzu ist zunächst dieser Normalfall mit der typischen Ausprägung zu definieren. In der Regel ist die Ausprägung mit der höchsten Fallzahl der Normalfall (n = MAX). Bei den nominalskalierten Variablen ist diesem Merkmal dann DUMMY = 0 zugeordnet.

Der Unterschied zwischen den Ausprägungen einer Variablen kann als Faktor errechnet werden, wobei die typische Ausprägung zu 1 gesetzt wird. Um diesen Abweichungsfaktor richtig einzuschätzen, wird dieser im Sinne einer »Korrektur am Kaufpreis« verstanden.

Nominalskalierte Werte

Abgeleitet aus den Gleichungen (7) und (8), ergibt sich für die Ausprägung ${}_{n}X_1$ der Abweichungsfaktor ${}_{n}F_1$ für nominalskalierte Werte zu

$$ {}_{n}F_1 = {}_{n}Y_{NO\,1} \,/\, {}_{n}Y_1 \; . \tag{12} $$

Wenn ${}_{n}Y_1 > {}_{n}Y_{NO\,1}$ ist, bedeutet dies, dass der Faktor ${}_{n}F_1$ kleiner 1 wird und somit als »Korrektur am Kaufpreis« den Ursprungswert reduziert.

Die Rechenvorschrift der Gleichung (12) gilt für alle m nominalskalierten Ausprägungen.

Ordinalskalierte Werte

Wie zuvor definiert, ist die typische Gruppenausprägung der ordinalskalierten Variablen die Ausprägung mit der höchsten Fallzahl. Diese wird aber in der Regel nicht im Regressionsansatz der Analyse vorhanden sein, da bei der Modellbildung Überbestimmungen in den Variablen zu vermeiden sind, um Fehlbeurteilungen auszuschließen.

Für den typischen Wert der Gruppe ${}_{o}X_1$ an der Stelle DUMMY = 0 wird der Faktor mit ${}_{o}F_{TYP\,1}$ benannt und gleich 1 gesetzt. Aus Gleichung (9) abgeleitet gilt dann

$$ {}_{o}F_{TYP\,1} = 1 = {}_{o}Y_{NO\,1} \,/\, {}_{o}Y_{NO\,1} \tag{13} $$

und für die p Ausprägungen einer Gruppe aus Gleichung (10)

$$ \begin{aligned} {}_{o}F_{11} &= {}_{o}Y_{NO\,1} \,/\, {}_{o}Y_{11} \\ {}_{o}F_{12} &= {}_{o}Y_{NO\,1} \,/\, {}_{o}Y_{12} \\ \text{"} \quad & \quad \text{"} \qquad \text{"} \\ {}_{o}F_{1p} &= {}_{o}Y_{NO\,1} \,/\, {}_{o}Y_{1p} \; . \end{aligned} \tag{14} $$

Diese Faktorenermittlung gilt für alle r ordinalskalierten Gruppen.

Verhältnisskalierte Werte

Die verhältnisskalierten Produkte liegen in der Regressionsgleichung additiv im Polynom vor; beschrieben unter 3.2.3 »Auflösung der verhältnisskalierten Werte«. Um die Werte der verhältnisskalierten Variablen bei der Normierung weiter als Faktoren zu verarbeiten, ist zunächst der jeweilige Wert des Einzelfalles ins Verhältnis zum Durchschnittskaufpreis $\hat{Y}$ zu setzen (Abb. 6).

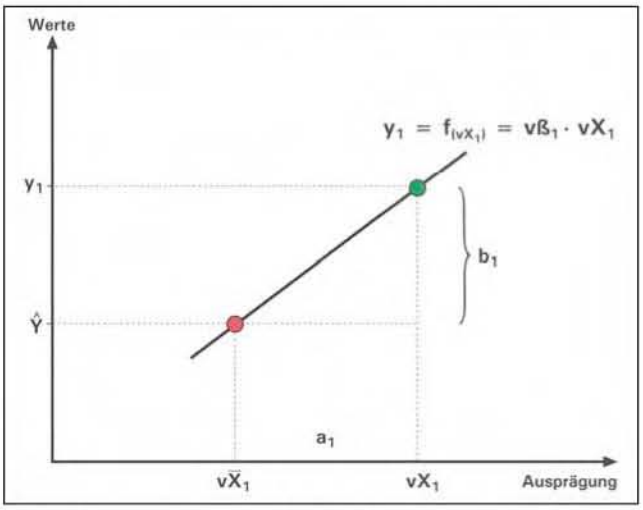

Abb. 6: Prinzip zur Ermittlung der Normierungsfaktoren bei funktionalen Zusammenhängen

Hierbei gilt: $a_1 = {}_{v}x_1 - {}_{v}\bar{x}_1$, wobei $\bar{x} \approx \bar{x}_{(TYP)}$

$$ b_1 = a_1 \cdot {}_{v}\beta_1 $$

$$ y_1 = Y + b_1 $$

Als Korrekturfaktor (${}_{v}F_1$) im Sinne einer »Korrektur am Kaufpreis« errechnet sich

$$ \begin{aligned} {}_{v}F_1 &= \hat{Y} \,/\, y_1 = \hat{Y} \,/\, (\hat{Y} + b_1) \\ &= 1 \,/\, (1 + b_1 \,/\, \hat{Y}) = 1 \,/\, (1 + a_1 \cdot {}_{v}\beta_1 \,/\, \hat{Y}) \; . \end{aligned} \tag{15} $$

Liegen verhältnisskalierte Zusammenhänge als Polynom vor, dann errechnet sich

$$ \begin{aligned} {}_{v}F_1 &= \hat{Y} \,/\, y_{11} \cdot \hat{Y} \,/\, y_{12} = \hat{Y}^2 \,/\, [(\hat{Y} + b_{11}) \cdot (\hat{Y} + b_{12})] \\ &= 1 \,/\, [(1 + a_{11} \cdot {}_{v}\beta_{11} \,/\, \hat{Y}) \cdot (1 + a_{12} \cdot {}_{v}\beta_{12} \,/\, \hat{Y})] \; . \end{aligned} \tag{16} $$

3.2.5 Beispiel

Die Darstellung der Modellauflösung in allgemeiner Schreibweise wirkt sehr kompliziert. Ein Beispiel macht die in der Praxis relativ einfache Handhabung deutlich:

Aufgabenstellung

77 unbebaute Bauland-Kauffälle mit der Zielgröße Kaufpreis in Euro pro m² Wohnfläche und den unabhängigen Variablen

1. Geschossflächenzahl (GFZ),
2. Wohnlage und
3. Gebäudeart (zukünftige freist. Einfamilien- bzw. Reihenhausbebauung)

sind zu untersuchen. Hierbei ist zu ermitteln, wie groß z. B. der Wertabstand zwischen der guten Wohnlage (2) und der mittleren Lage (3) sowie der Unterschied (in %) zwischen freistehenden Einfamilienhaus- und Reihenhausgrundstücken ist.

Lösungsweg

Das optimale Regressionsmodell lautet unter Anwendung der Gleichung (2)

$$Y = \hat{Y} = \beta_0 + \beta_1 \overline{x}_{1\,(GFZ)} + \beta_2 \overline{x}_{2\,(Wohnanlage\ 1)} + \beta_3 \overline{x}_{3\,(Wohnanlage\ 2)} + \beta_4 \overline{x}_{4\,(Wohnanlage\ 4)} + \beta_5 \overline{x}_{5\,(Reihenhaus\text{-}Grundstück)} \,.$$

Die Analyse der Stichprobe führte zu den in Tab. 1 genannten Ergebnissen.

Tab. 1: Ergebnissen der Analyse der Stichprobe

Y (arithm. Mittel)	$\overline{x}_{1,2...5}$ (arithm. Mittel)	Parameter $\beta_{0,1...5}$	Variable	Zeile
1365,2 =			Kaufpreis pro Wohnfläche (EUR/m²)	1
		2129,0	Achsenabschnitt (β_0)	2
	0,6700 ×	−1718,9	Geschossflächenzahl (GFZ)	3
	0,0390 ×	2864,1	DUMMY Wohnlage 1 (sehr gut)	4
	0,0364 ×	604,0	DUMMY Wohnlage 2 (gut)	5
	0,0130 ×	−717,5	DUMMY Wohnlage 4 (einfach)	6
	0,2468 ×	268,2	DUMMY Reihenhausgrundstück	7

Aus diesen Rechengrößen lässt sich das Modell wie folgt partiell auflösen:

Verhältnisskalierte Einflussgröße (Geschossflächenzahl)

Nach Gleichung (11) wird durch Einsetzen entsprechender Zahlen aus den Zeilen 2 und 3

$${}_v\hat{Y}_1 = 2129{,}0 + (-1718{,}9) \cdot GFZ \,.$$

Der GFZ-Einfluss[5] ist somit als degressive, lineare Funktion darstellbar oder als Faktor bei Abweichung von einem typischen Grundstück mit einer GFZ von 0,67 nach Gleichung (15)

$${}_vF_1 = 1 \,/\, (1 + (GFZ - 0{,}67) \cdot (-1718{,}9) \,/\, 1265{,}2) \,.$$

Ordinalskalierte Einflussgröße (Wohnlage)

Nach Einsetzen entsprechender Werte der Zeilen 1, 4, 5 und 6 in die Gleichung (9) – Merkmal nicht vorhanden – wird

$${}_oY_{NO\,1} = 1365{,}2 - [(0{,}0390 \cdot 2864{,}1) + (0{,}0364 \cdot 604{,}0) + (0{,}0130 \cdot (-717{,}5))] = 1240{,}8 \,.$$

Nach Auflösung gemäß Gleichung (10) – Merkmal vorhanden – wird

$${}_oY_{11} = 1240{,}8 + 2864{,}1 = 4104{,}9$$
$${}_oY_{12} = 1240{,}8 + 604{,}0 = 1844{,}8$$
$${}_oY_{13} = 1240{,}8 + (-717{,}5) = 523{,}3 \,.$$

Diese Zwischenergebnisse führen bei Anwendung der Gleichungen (13) und (14) zu folgenden Unterschiedswerten:

Wohnlage 3: $oF_{TYP\,1} = 1240{,}8 \,/\, 1240{,}8 = 1{,}00$
(typ. Merkmal)
Wohnlage 1: $oF_{11} = 1240{,}8 \,/\, 4109{,}9 = 0{,}30$
Wohnlage 2: $oF_{12} = 1240{,}8 \,/\, 1844{,}8 = 0{,}67$
Wohnlage 4: $oF_{13} = 1240{,}8 \,/\, 523{,}3 = 2{,}37.$

Nominalskalierte Einflussgröße (Reihenhausgrundstück)

Nach Einsetzen der Werte der Zeilen 1 und 7 in die Gleichung (7) – Merkmal nicht vorhanden – wird

$${}_nY_{NO\,1} = 1365{,}2 - (0{,}2468 \cdot 268{,}2) = 1299{,}0 \,.$$

Nach Auflösung gemäß Gleichung (8) – Merkmal vorhanden – errechnet sich

$${}_nY_1 = 1299{,}0 + 268{,}2 = 1567{,}2 \,.$$

Aus diesen Zahlen lässt sich der Wertunterschied nach Gleichung (12) wie folgt ermitteln:

Freistehendes Einfamilienhausgrundstück
(typisches Merkmal nach Definition):
$nF_{TYP\,1} = 1{,}00$
Reihenhausgrundstück:
$nF_1 = 1299{,}0 \,/\, 1567{,}2 = 0{,}83.$

Ergebniszusammenfassung

Die errechneten Zahlen machen deutlich, dass z. B. ein Kaufpreis, in guter Wohnlage (2) erzielt, multipliziert mit dem Faktor 0,67, das Niveau der mittleren Wohnlage (3) annimmt.

Die Antwort auf die Frage, wie groß ist der Unterschied U (in %) zwischen freistehenden Einfamilienhaus-

5 Die Funktionen zum GFZ-Einfluss auf den Kaufpreis verlaufen degressiv. Eine steigende GFZ bewirkt also ein Fallen der Preise. Dies ist bei dem Maßstab Kaufpreis pro m² Wohn-/Nutzfläche plausibel und darin begründet, dass bei steigender Wohnfläche, unter Beibehaltung der Grundstücksgröße, die GFZ steigt. Gleichzeitig fällt, unter Beibehaltung des Gesamtkaufpreises bei gleicher Grundstücksgröße, der Kaufpreis pro m² Wohnfläche. Dieser Einfluss wird besonders stark bei Sachwertgrundstücken (Einfamilienhäusern) sichtbar, da hier in der Regel das Wertniveau des Grund und Bodens lageabhängig und nicht in Abhängigkeit von der GFZ entsteht.

und Reihenhausgrundstücken, errechnet sich nach der Vorschrift:

$$U_1 = 100 / {}_nF_1 - 100 \quad \text{zu}$$

$$U_{(\text{Reihenhaus})} = 100 / 0{,}83 - 100 = +20\,\% \cdot$$

Einfamilienreihenhausgrundstücke sind somit um 20 % teurer als freistehende Einfamilienhausgrundstücke.

In Analogie ergibt sich für den Lagevergleich: Die Wohnlage (2) ist 49 % teurer als die Wohnlage (3).

3.2.6 Anwendung

Diese zuvor beschriebenen Ableitungsschritte können so programmiert werden, dass die Endergebnisse – die jeweiligen Unterschiedsfaktoren ${}_nF_1$, ${}_nF_2 \ldots {}_nF_m$, sowie ${}_oF_{11}$, ${}_oF_{12} \ldots {}_oF_{rp}$ – direkt ablesbar sind. Hierzu müssen die Ergebnisse aus der Regressionsanalyse in eine Datei geschrieben werden, um diese dann weiter verarbeiten zu können.

Die partielle Modellauflösung führt zu

- **Kurvenfunktionen** bei verhältnisskalierten Werten
- **Gruppenabstufungen** bei ordinalskalierten Werten
- **Unterschiedsabstufungen** bei nominalskalierten Werten.

Die zum Stichprobentyp (NORMFALL) gefundenen Unterschiede können mithilfe der zuvor ermittelten Faktoren oder auch in Prozent umgerechnet, anschließend vom Sachverständigen geprüft und diskutiert werden. Die funktionalen Zusammenhänge lassen sich zweidimensional darstellen und sind somit ebenfalls sachkundig beurteilbar.

Zur Weiterverarbeitung können die Norm- oder Korrekturfaktoren in einem Rechenoperationsblock, dem so genannten »NORM-POOL« abgelegt werden.

Dieser NORM-POOL wird zunächst aus den statistischen Ergebnissen gespeist, kann aber durch das Expertenwissen ergänzt werden. Diese Rechen-(Normierungs-)Logik kann dann bei den verschiedenen Anwendungen auf den Kaufpreis angesetzt werden (Abb. 7).

```
if wolage =  3   then f_wola = 1.00;                  * Wohnlage 3 (Typ);
if wolage =  1   then f_wola = 0.30;                  * Wohnlage 1;
if wolage =  2   then f_wola = 0.67;                  * Wohnlage 2;
if wolage =  4   then f_wola = 2.37;                  * Wohnlage 4;
if gebart = 13   then f_geb  = 1.00;                  * Einfamilienhausgrundstück;
if gebart = 14   then f_geb  = 0.83;                  * Einfamilienreihenhausgrundstück;
f_gfz = 1/(1 + (GFZ – 0,67) · (–1718,9) / 1365,2);    * Basis GFZ = 0,67;
```

Abb. 7: Beispiel eines NORM-POOLs für unbebaute Einfamilienhausgrundstücke[6]

Hierbei bedeuten:

wolage	Wohnlage mit Schlüsselnummer
gebart	Gebäudeart mit Schlüsselnummer
f_geb	Faktor für die Normierung der Gebäudeart
f_wola	Faktoren für die Normierung der Wohnlagen
f_gfz	GFZ-Faktor (analog ${}_vF_1$ nach Gleichung (15))
if/then	Rechenlogik

Die Anwendung der Faktoren an den Ursprungskaufpreis pro m² Wohn-/Nutzfläche (KPNF) führt zum normierten Kaufpreis ($KP_{(NORM)}$) und erfolgt nach der Rechenvorschrift:

$$KP_{(NORM)} = KPNF \cdot f\_geb \cdot f\_wola \cdot f\_gfz$$

3.3 Schritte im multivariaten Auswerteverfahren

Der zeitintensive Prozess des multivariaten Auswerteverfahrens gliedert sich in die folgenden, hier nur kurz dargestellten Schritte[7]. Ziel ist es, die Sachverständigen des Gutachterausschusses, als Nichtmathematiker, frühzeitig in das Verfahren mit einzubinden.

3.3.1 Auswahl der zu untersuchenden Teilmärkte und der abhängigen Einfluss-(Ziel-)Größe

Die Auswahl der Stichprobe steht in engem Zusammenhang mit der zu lösenden Aufgabenstellung. Für die durchzuführenden multiplen linearen Regressionsanalysen sollte der Datenumfang so groß wie möglich sein. Hierbei ist zu beachten, dass die Stichproben natürlich nicht beliebig vergrößert werden können, da immobilienmarktspezifische Eigenschaften der Teilmärkte zu beachten sind. Typische praxisorientierte teilmarktabhängige Preismaßstäbe sind für

- unbebaute Grundstücke:
  Kaufpreis pro Grundstücksfläche
- bebaute Grundstücke:
  Kaufpreis pro Rohertrag und
  Kaufpreis pro Wohn-/Nutzfläche
- Wohnungseigentum:
  Kaufpreis pro Wohnfläche.

3.3.2 Sichten der in der Kaufpreissammlung erfassten unabhängigen Einflussgrößen

Dieser Arbeitsschritt ist im Verfahren nach Zeitaufwand und Qualität nicht zu unterschätzen. Die intensive Beschäftigung mit dem Datenmaterial lässt die innere Datenstruktur und den Modelltyp, d. h. das typische Objekt der Stichprobe je Teilmarkt, erkennen.

6 Baustein im INTEX.GA (Integriertes Expertensystem zur Führung und Auswertung der Kaufpreissammlung und Gutachtenbearbeitung) eine Eigenentwicklung auf Basis der Software SAS (SAS Institute GmbH, Heidelberg).

7 Ausführliche Beschreibung in Mann 2004, Integratives Auswertemodell zur Beschreibung des Immobilienmarktes, S. 33–60.

Auswahl der preisbeeinflussenden Merkmale
Zunächst sollten unabhängig von der Größe der Stichproben relativ viele Einflüsse (Regressoren) ausgewählt werden. Im Laufe des Verfahrens, bei der Strukturierung und statistischen Auswertung, reduziert sich dann die Menge.

Sachgerechte Vorabgruppenbildung
Die in der Kaufpreissammlung vorgegebene Codierung kann durchaus im konkreten Auswertefall aufgegeben werden, sodass z.B. neue Gruppen zu bilden sind.

Datenkontrolle
Eine erste Auszählung der Daten lässt erkennen, ob bei vergleichbarem Auswertestand alle Datenfelder ausgefüllt bzw. erfasst sind. Neben der Prüfung auf Vollständigkeit sollten die umfangreichen Fragestellungen nach Inhalt und Fehlern des Datenmaterials geklärt werden, z.B. ob Definitionen eingehalten sind, logische Zuordnungen stimmen, Erfassungsfehler vorliegen.

Prüfen auf offensichtliche Autokorrelation
Noch vor der statistischen Analyse ist empirisch zu prüfen, ob zwei unabhängige Einflüsse offensichtlich voneinander abhängig (autokorreliert) sind.

Definition der Stichprobe
Aufgrund der bisher gefundenen Verteilungs- und Mengenübersichten ist die Stichprobe so einzugrenzen, dass die untypischen Fälle die nachfolgenden Untersuchungen nicht stören können.

Es ist hierbei darauf zu achten, dass verhältnisskalierte Variablen **nicht** oder **nicht allein** nach mathematisch-statistischen Gesichtspunkten ausgegrenzt werden.

Bilden von Dummy-Variablen
Die Zuordnung für ordinal-/nominalskalierte Variablen erfolgt nach der Maßgabe DUMMY = 1, wenn Merkmal vorhanden, DUMMY = 0, wenn Merkmal nicht vorhanden ist.

Schrittweise Regression
Dieses Spezialverfahren der Varianzanalyse dient der Minimierung der Anzahl der in die Regressionsanalyse einzuführenden Variablen. Wenn in den Untersuchungen große Stichproben mit vielen Regressoren verarbeitet werden, ist diese Vorabprüfung sinnvoll; ansonsten kann dieser Schritt im Verfahren auch ausgelassen werden.

3.3.3 Durchführung der Regressionsanalyse

Dem klassischen Prozess der Regressionsanalyse können durch die Methode der partiellen Modellauflösung weitere Schritte hinzugefügt werden.

Einfache Statistiken
Anzahl, Mittelwert, Standardabweichung, Abweichung des Mittels, Minimum, Maximum, relative Abweichung und Summe für die jeweils zu analysierenden Variablen werden ermittelt. Entscheidend für die Weiterverarbeitung im Verfahren sind die Mittelwerte und bei den Dummy-Variablen auch die Summe, da diese Kenngröße hier die Anzahl der Variablen bzw. Variablengruppen angibt.

Multiple lineare Regression
Zur Beurteilung der Varianzanalyse-Ergebnisse bieten sich an:
- Werte der F-Statistik zur Modellbeurteilung
- Bestimmtheitsmaß als Maß für die Modellerklärung
- Parameter auf logische Vorzeichen und Größenordnungen prüfen
- die unabhängigen Variablen auf Signifikanz prüfen.

Prüfen der Residuen auf Normalverteilung
Grobe Abweichungen von der Normalverteilung weisen auf ein nicht exakt definiertes Modell hin.

Liste der Criticals (Ausreißer)
Nach mathematisch-statistischen Gesichtspunkten lassen sich Einzelbeobachtungen als Criticals herausfinden, die zum einen auf Ausreißer (untypische Fälle) schließen lassen, oder auch auf Fälle hinweisen, die einen extremen Einfluss auf das Modell ausüben können (Hebelpunkte).

In der Literatur finden sich Angaben zu ausreißerverdächtigen Kauffällen nach der Faustformel:

$$Y_{\text{Ausreißer}} < (\hat{Y} - 2{,}5s) \text{ oder } Y_{\text{Ausreißer}} > (\hat{Y} + 2{,}5s)$$

Ausreißer werden in der Regel nach dem ersten oder zweiten (optimierten) Modellansatz gefunden. Unter Beachtung der Hebelpunkte überschreiten durchaus Fälle die »4s-Grenze« und verbleiben dennoch im Regressionsansatz. Entscheidend ist, dass trotz in der Stichprobe verbleibender »Ausreißer«, oder besser Criticals, die Residuen normalverteilt sind.

Kontrolle der Datenzuordnungen/Autokorrelation
Durch Korrelationsmatrizen lassen sich unerwünschte Autokorrelationen aufdecken. In der Praxis wird deutlich, dass dieses Verfahren vereinzelt Korrelationen nachweist, die aber aufgrund der »Kaufpreisrealität« durchaus akzeptabel und vertretbar sein können.

Optimierte Regressionsfunktion
Aufgrund der bisher gewonnenen Erkenntnisse werden die Daten neu geordnet. Ausreißer entfallen, bei korrelierten unabhängigen Variablen wird nur eine ins Modell übernommen usw. Im Anschluss hieran wird die Regressionsanalyse erneut durchgeführt. Dieser Prozess verbessert die Ergebnisse in der Regel
- im Bestimmtheitsmaß
- in der Optimierung der Auswahl der signifikanten Einflussgrößen
- bezüglich der Normalverteilung der Residuen.

Ableitung und Verarbeitung der partiellen Einflussfaktoren
Die bisher gefundene Regressionsgleichung wird gemäß der zuvor beschriebenen »partiellen Modellauflösung« aufgelöst und die Normierungsfaktoren in dem so genannten »NORM-POOL« abgelegt. Somit kann jeder Kaufpreis auf die zuvor definierte Norm und somit auf das typische Niveau der Stichprobe umgerechnet werden.

Kontrolle
Mithilfe der Regressionsanalyse, wobei jetzt als Zielgröße die normierten Kaufpreise eingeführt werden, können folgende Regressionsbedingungen überprüft werden:

- Sind die Residuen normalverteilt?
- Sind die normierten Kaufpreise normalverteilt?
- Liegt das Bestimmtheitsmaß bei 0?
- Tragen die Parameterschätzungen mit hoher Wahrscheinlichkeit nicht mehr zur Modellerklärung bei?

Können diese Fragen mit »Ja« beantwortet werden, ist als Schlusskontrolle davon auszugehen, dass auch Kurvenverläufe, Gruppenabstufungen, Unterschiedsabstufungen und die Normierungsfunktion korrekt ermittelt wurden.

Indexermittlung
Große Stichproben, die für die Analyse gefordert werden, bedingen die Selektion von Kaufpreisen, die im Verlauf mehrerer Jahre gehandelt wurden. Die dann im Rahmen der Regression gefundene, in der Regel als Polynom vorliegende partielle Zeitkurve wird die tatsächliche Preisentwicklung nicht marktgerecht widerspiegeln. Das liegt zum einen darin begründet, dass am Ende der Funktionskurve, d.h. heute, die Aussagekraft durch die Spreizung des Konfidenzbandes am geringsten ist. Zum anderen kann der exakte Kurvenverlauf durch ein weit geschwungenes Polynom nur bedingt wiedergegeben werden.

Dieses Problem der Zeitreihenmodulation kann entschärft werden, indem von Jahresmitteln ausgegangen wird. Hierzu wird das gesamte Material der Stichprobe mit allen bisher ermittelten Korrekturwerten so normiert, dass nur der Einfluss durch den Kaufzeitpunkt erhalten bleibt.

Diese Werte lassen sich pro Jahr zusammenfassend auswerten und grafisch darstellen.

Nach sachverständiger Prüfung auf »Jahres-Criticals« führen diese Jahresmittel zum Index für jedes Jahr. Dieses Ergebnis kann anhand anderer Wirtschaftsdaten und oder ggf. durch Städtevergleich auf Plausibilität geprüft, empirisch korrigiert und als »Fix-Größe« erneut in die Regression eingeführt werden.

Iterationen
Die Schritte im Rahmen der »Durchführung der Regressionsanalyse« können beliebig wiederholt werden. Durch die bisher gewonnenen Erfahrungen können auch Schritte zu »Sichten der in der Kaufpreissammlung erfassten unabhängigen Einflussgrößen« wiederholt bzw. ergänzt werden.

3.4 Ergebnisse aus dem multivariaten Auswerteverfahren

Zum Schluss ist zusammenfassend noch einmal zu erinnern, dass die stochastischen Eigenschaften von Kaufpreisen interpretationswürdig sind. Mathematiker und Sachverständige müssen deshalb bei aller Akribie in der Anwendung der Auswertemethoden das Bewusstsein dafür bewahren, dass hohe Genauigkeiten in den Ergebnissen im mathematischen Sinn nicht zu erwarten sind.

3.4.1 Ergebnisse zu den qualitativen Fragestellungen

Die statistischen Tests der Varianzanalyse lassen Erkenntnisse zu signifikanten Einflüssen auf den Kaufpreis und zur Modellanpassung zu. Hierbei sind die wesentlichen statistischen Kenngrößen, wie Anzahl der Fälle, Anzahl der Ausprägungen, Mittelwerte, kleinster/größter Wert, Standardabweichung des Einzelfalles/des Mittels, relative Abweichung in Prozent, Vorzeichen der Parameter der unabhängigen Variablen ($\beta_{1,2,\ldots}$), partielles Bestimmtheitsmaß (B_{YX}), multiples Bestimmtheitsmaß (B oder R^2), Irrtumswahrscheinlichkeit (i.d.R. Signifikanzniveau 5 %) und Testgröße(n) auf Signifikanzniveau (Fisher-/Student-Verteilung) sachgerecht zu interpretieren. Diese stehen in der Regel in sachlogischen Zusammenhängen und der Schwerpunkt sollte nicht nur auf eine statistische Kenngröße, wie zum Beispiel das multiple Bestimmtheitsmaß (R^2), gelegt werden.

Aus Untersuchungen großer Stichproben der Düsseldorfer Kaufpreissammlung wurden die in Tab. 2 genannten Merkmale je Teilmarkt, die den Kaufpreis pro m^2 Wohnfläche beeinflussen, als wesentlich erkannt.

3.4.2 Ergebnisse zu den quantitativen Fragestellungen

Aus der Regressionsgleichung lassen sich die Unterschiedsausprägungen und funktionalen Zusammenhänge nach dem zuvor beschriebenen Verfahren der partiellen Modellauflösung ableiten. Hierbei sind bei großen Stich-

Tab. 2: Wesentliche preisbildende Merkmale

Merkmal	unbebaute Grundstücke	bebaute Grundstücke	Wohnungseigentum
Kaufzeitpunkt	✗	✗	✗
Wohnlage	✗	✗	✗
Gebäudeart	✗	✗	
Geschossflächenzahl	✗		
Sondermerkmal, geplant Wohnungseigentum	✗		
Wiederkaufsrecht der Stadt	✗		
Alter/Modernisierung		✗	✗
Grundstücksgröße			✗
Objektgröße		✗	✗

Tab. 3: Wesentliche statistische Kenngrößen

Kenngrößen	Ergebnisse
Ausreißer	Ausreißer > 4,5s-Grenze Nach Tschebyscheff umfasst bei beliebigen Verteilungen und großen Stichproben der 3s-Bereich noch 88,9 %, der 4s-Bereich noch 93,7 % und der 5s-Bereich noch 96,0 % aller Werte[8].
Genauigkeitsmaße für Unterschiedsausprägungen	Die Genauigkeitsmaße liegen nach Durchführung der Regression als Variationskoeffizient, relative Abweichung in %, für ■ unbebaute Grundstücke bei ±30 % ■ bebaute Grundstücke bei ±27 % ■ Wohnungseigentum bei ±20 %. Die Ergebnisse sind eine Funktion der Anzahl der unabhängigen Variablen bzw. Ausprägungen (k) und der Beobachtungen (n).
Bestimmtheitsmaße	Das Bestimmtheitsmaß (B) als Maß für die gesamte Modellanpassung(-güte) liegt zwischen 0,5 und 0,85. Die partielle Modellauflösung mit anschließender Normierung der Preise auf eine darzustellende Einflussgröße führt zu einem B zwischen 0,15 und 0,65.
Vertrauensbereiche	Die Basis der Untersuchungen liegt auf dem Signifikanzniveau. Das heißt, die Ergebnisse treffen mit 95-prozentiger Wahrscheinlichkeit zu. Die Wahl kleiner Irrtumswahrscheinlichkeiten vergrößert die Stichprobe nach Anzahl, z.B. beim Ausschluss von Ausreißern, lässt aber weniger Ausprägungen (unabhängige Variable) zur Beurteilung zu.

Tab. 4: Ausschnitt zu Normierungsfaktoren

Merkmal des Normalfalles	abweichende Ausprägung	Normfaktor an den Kaufpreis	Korrektur in % am Normalwert
mittlere Wohnlage	sehr gute Wohnlage	0,71	+40 %
	gute Wohnlage	0,87	+15 %
	einfache Wohnlage	1,11	−10 %
modernisiert, neuzeitliche Ausstattung	nicht modernisiert (Baujahr bis 1974)	1,33	−25 %
Mietverhältnis wird nicht übernommen	Übernahme des/der Mietverhältnisse/s	1,11	−10 %

proben die in Tab. 3 genannten typischen statistischen Kenngrößen zu erwarten.

Die qualitativen Ergebnisse lassen sich wie in Tab. 4 und Abb. 8 darstellen. Hierbei ist zu beachten, dass der Normfaktor auf den Kaufpreis bewirkt, dass ausgehend von den vorhandenen Kaufpreisen Norm- oder Richtwerte abgeleitet werden können. Die Korrektur in % am Normalwert ermöglicht die Ableitung von Vergleichswerten aus Richtwerten (Normwerten). Dieser Ansatz ist dem Experten vertraut, da er in der Regel mithilfe von Prozentangaben das Normierungsprinzip anwendet.

Normierter Kaufpreis (KPNF_A) in Abhängigkeit vom Gebäudealter bei bebauten Grundstücken

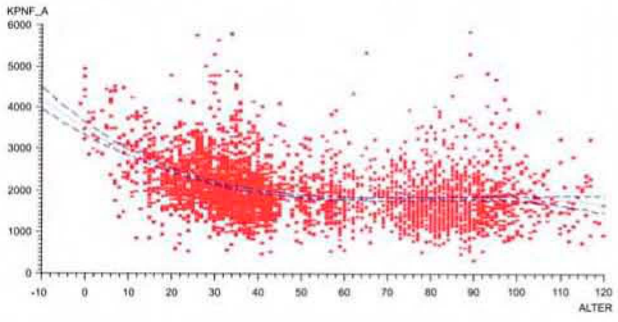

Abb. 8: Darstellung funktionaler Zusammenhänge

4 Ziele und Ausblick

4.1 Verstärkte Nutzung der Kaufpreissammlung als Wissensbasis für Experten

Ein Sachverständiger baut sein Expertenwissen aus verschiedenen Quellen, wie eigene Umfragen oder Bewertungen, Fachliteratur, Markterfahrung u. a. auf. Eine qualifiziert geführte Kaufpreissammlung als Datenbasis wird aber bisher kaum umfassend als Wissensbasis für Experten aufgeschlüsselt. Die erfahrenen Sachverständigen für Grundstücksbewertungen sind in der Regel gegenüber den Ergebnissen aus mathematisch-statistischen Analysen eher skeptisch eingestellt, zumal sie Regressionsgleichungen nicht interpretieren können. Die vorgestellte Analysetechnik bietet hier neue Möglichkeiten an, da die Bewertungsschablonen der Experten und der Ergebnisse

8 Sachs 1999, Ungleichung von Tschebyscheff, S. 122; Ausreißer, S. 364.

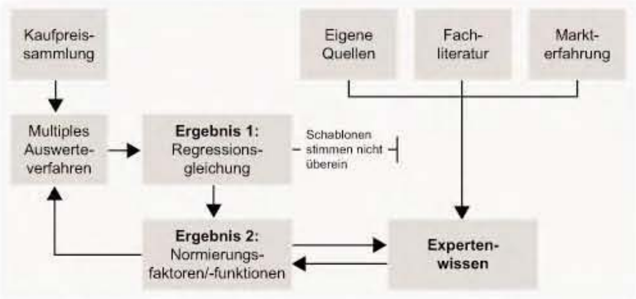

Abb. 9: Statistische Ergebnisse aus der Kaufpreissammlung und Expertenwissen

aus Regressionsanalysen in Übereinstimmung gebracht werden können. Abb. 9 zeigt die Einwirkungsmöglichkeit der Sachverständigen im multiplen Auswerteverfahren und deren Bereicherung beim Aufbau von eigenem Expertenwissen aus dem Material der Kaufpreissammlung.

4.2 Automatisierte Ableitung von Richtwerten aus großen Datenbeständen

Den Mitgliedern der Gutachterausschüsse, die abschließend die Verantwortung für die Veröffentlichung der erforderlichen Daten in den Grundstücksmarktberichten übernehmen, wird es leichter fallen, Ergebnisse zu veröffentlichen, wenn sie diese auf Marktkonformität hin überprüfen konnten. Dies wird durch die Einbindung der Experten im Prozess der multiplen Auswerteverfahren durch eine Diskussion der Normierungsfaktoren oder der funktionalen Zusammenhänge möglich.

Gleichzeitig lassen sich die sachverständig geprüften Korrekturwerte so auf die Kaufpreise anwenden, dass automatisiert Richtwerte abgeleitet, beschlossen und veröffentlicht werden können (Abb. 10).

Abb. 10: Von Kaufpreisen zu Richtwerten

Die in Düsseldorf auf einer Karte publizierten Marktrichtwerte basieren auf diesem Prinzip. Aus jährlich ca. 3500 ausgewerteten Kauffällen, überwiegend für Eigentumswohnungen, für bebaute Grundstücke sowie unbebaute Baulandgrundstücke, werden automatisiert Korrekturwerte an die Kaufpreise angebracht. Unter Anwendung des Normierungsprinzips werden Kaufpreislisten pro Gebiet und pro Teilmarkt erzeugt, um endlich ca. 360 Marktrichtwerte abzuleiten, zu beschließen und zu veröffentlichen (Mann 2005).

Die angewendeten Korrekturwerte werden ebenfalls veröffentlicht, diesmal in %, um dem Anwender die Möglichkeit zu eröffnen, sachverständig aus den Richtwerten

Abb. 11: Von Richtwerten zu Verkehrswerten

Vergleichswerte (Markt- oder Verkehrswerte) abzuleiten (Abb. 11).

Weitere Anwendungen finden sich in der Ableitung und Anwendung von Bodenrichtwerten, Sachwertkorrekturfaktoren, Liegenschaftszinssätzen oder auch bei teilmarktübergreifenden Auswertungen, z. B. typisierte Wertverhältniszahlen zwischen den Teilmärkten »Düsseldorfer Türmchen« oder Rohertragsvervielfältiger für unbebaute Baulandgrundstücke (Gutachterausschuss für Grundstückswerte in der Landeshauptstadt Düsseldorf 2005). Darüber hinaus werden »Immobilienrechner« möglich, die normierte Kaufpreismittelwerte für ausgewählte Teilmärkte auch per Internet ausrechnen können.

Literatur

Bortz, J.: Statistik für Sozialwissenschaftler. Springer-Verlag, Berlin/Heidelberg, 4. Auflage 1993.

Brückner, R. (Hrsg.): Mathematische Statistik bei der Ermittlung von Grundstückswerten – Lehrbriefe und Vorlesungen zum Kontaktstudium 1976 und des Anwenderseminars 1978 des Geodätischen Instituts der Universität Hannover. Wissenschaftliche Arbeiten der Lehrstühle für Geodäsie, Photogrammetrie und Kartographie an der Technischen Universität Hannover, Nr. 65, 1976, Nr. 65a, 1978.

Gutachterausschuss für Grundstückswerte in der Landeshauptstadt Düsseldorf: Grundstücksmarktbericht, 2005.

Kreyszik, E.: Statistische Methoden und ihre Anwendungen. Vandenhoek und Rupprecht in Göttingen, 4. Auflage 1973.

Mann, W.: Marktrichtwerte – Verfahren zur Ableitung und Beschlussfassung von Marktrichtwerten und praktische Erfahrungen. Flächenmanagement und Bodenordnung 4/5/2005, Luchterhand Verlag GmbH, Neuwied 2005.

Mann, W.: Integratives Auswertemodell zur Beschreibung des Immobilienmarktes. Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover Nr. 251, 2004.

Reuter, F.: Zur Problematik statistischer Vergleichsmodelle bei der Ermittlung von Verkehrswerten. Zeitschrift Vermessungswesen und Raumordnung, 56. Jahrgang, S. 338–350, 1994.

Sachs, L.: Angewandte Statistik. Springer-Verlag, Berlin/Heidelberg/New York, 9. Auflage 1999.

Schnell, R., Hill, P. B., Esser, E.: Methoden der empirischen Sozialforschung. R. Oldenbourg Verlag, München/Wien, 5. Auflage 1995.

Schlittgen, R.: Einführung in die Statistik – Analyse und Modellierung von Daten. R. Oldenbourg Verlag, München/Wien, 9. Auflage 2000.

Anschrift des Autors
Dipl.-Ing. Wilfried Mann
Geschäftsführer und stellv. Vorsitzender
c/o Gutachterausschuss für Grundstückswerte in der
Landeshauptstadt Düsseldorf
Brinckmannstraße 5, 40200 Düsseldorf
wilfried.mann@stadt.duesseldorf.de