



Die Berücksichtigung von außergesetzlichen Merkmalen bei der Mietspiegelerstellung – Kausalität versus Vorhersage

Göran Kauermann · Michael Windmann 

Eingegangen: 13. Mai 2022 / Angenommen: 3. Mai 2023 / Online publiziert: 14. Juni 2023
© Der/die Autor(en) 2023

Zusammenfassung Das neue Mietspiegelgesetz erlaubt die Berücksichtigung von sogenannten außergesetzlichen Merkmalen wie Mietdauer und Vermieterotyp bei der Erstellung von Mietspiegeln. Diese außergesetzlichen Merkmale dürfen in zukünftigen Mietspiegeln bei deren Erstellung und Modellwahl Einfluss finden, nicht aber im konkreten Mietspiegelmodell. Diese gesetzliche Vorgabe lässt viel Spielraum, der in diesem Beitrag aus statistischer Sicht beleuchtet wird. Anhand von konkreten Daten werden die Konsequenzen quantifiziert und aufgezeigt.

Schlüsselwörter Mietspiegel · außergesetzliche Merkmale · Regressionsmietspiegel

The consideration of extra-legal features when creating the rent index—causality versus prediction

Abstract The new rent index law allows the consideration of so-called extra-legal features such as rental period and landlord type when creating rent indexes. These extra-legal features may influence the model selection of future rent indexes, but not the specific rent index model. This legal requirement leaves a lot of leeway, which is examined in this article from a statistical point of view. The consequences are quantified and shown on the basis of concrete data.

✉ Göran Kauermann · Michael Windmann
Institut für Statistik, Ludwig-Maximilians-Universität München, Ludwigstraße 33, 80539 München, Deutschland
E-Mail: goeran.kauermann@lmu.de

Michael Windmann
E-Mail: michael.windmann@lmu.de

1 Einleitung

Mietspiegel wurden vor etwa 50 Jahren als marktregulierendes Instrument durch das 1974 beschlossene Wohnraumkündigungsschutzgesetz eingeführt. Sie geben dabei die sogenannte ortsübliche Vergleichsmiete an, die durch Merkmale wie Wohnungsgröße und Beschaffenheit der Wohnung beeinflusst wird. Das Bürgerliche Gesetzbuch (BGB) regelt dabei in § 558 Abs. 2 BGB die verpflichtenden Merkmale wie Art, Größe, Ausstattung und Lage der Wohnung. Mietspiegel werden generell unterschieden in einfache Mietspiegel und qualifizierte Mietspiegel, als auch in Regressions- und Tabellenmietspiegel. Wir betrachten hier ausschließlich qualifizierte Regressionsmietspiegel. Der Mietspiegelreport der Gesellschaft für Immobilienforschung (Sebastian und Memis 2021) liefert seit einigen Jahren einen immer detaillierter werdenden Bericht über die Mietspiegel in den deutschen Städten und gibt darüber hinaus eine Beurteilung der Qualität der Mietspiegel. Eine generelle Diskussion über die Qualität von Mietspiegeln aus statistischer Sicht findet sich auch in Freund et al (2013) oder Kauermann und Windmann (2016). Die ökonomische Rolle von Mietspiegeln wird in Lerbs und Sebastian (2015) beleuchtet.

In diesem Artikel wollen wir uns der Frage widmen, wie sogenannte außergesetzliche und gesetzliche Merkmale gemeinsam im Mietspiegel verarbeitet werden können. Wir zeigen aus Sicht der Statistik Möglichkeiten und Folgen der Berücksichtigung außergesetzlicher Merkmale auf. Dem Mietspiegelreformgesetz (Bundesgesetzblatt 2021) folgend sind **gesetzliche Merkmale** nach § 2 Abs. 1 dabei wie folgt definiert: *Wohnwertrelevante gesetzliche Merkmale sind die in § 558 Absatz 2 Satz 1 des Bürgerlichen Gesetzbuchs genannten Merkmale Art, Größe, Ausstattung, Beschaffenheit und Lage einer Wohnung, soweit sie für die Mietpreisbildung relevant sind oder im Erstellungsstadium des Mietspiegels relevant sein können.* Demgegenüber stehen **außergesetzliche Merkmale**, die definiert sind als: *Außergesetzliche Merkmale sind Merkmale in Bezug auf die Wohnung oder das Mietverhältnis, die in § 558 Absatz 2 Satz 1 des Bürgerlichen Gesetzbuchs nicht genannt sind, aber dennoch für die Mietpreisbildung relevant sind oder im Erstellungsstadium des Mietspiegels relevant sein können.* Hierunter fallen beispielsweise die Mietdauer oder der Vermietertyp. Bisher waren auf Grund der Definition von Mietspiegeln (§ 558 Abs. 2 Satz 1 BGB) außergesetzliche Merkmal bei der Mietspiegelerstellung nicht nutzbar (siehe auch Bundesinstitut für Bau-, Stadt- und Raumforschung 2020, S. 16f). Das neue Gesetz lässt den Mietspiegelerstellern hier aber mehr Spielraum und nach § 14 gilt: *Wird die ortsübliche Vergleichsmiete nach der Regressionsanalyse ermittelt, so sind wohnwertrelevante gesetzliche Merkmale daraufhin zu untersuchen, ob sie einen statistisch signifikanten Einfluss auf den Mietpreis haben mit dem Ziel, den Zusammenhang zwischen der Miethöhe und den gesetzlichen wohnwertrelevanten Merkmalen möglichst gut zu beschreiben. Außergesetzliche Merkmale können insbesondere zur Wahl des Regressionsmodells und bei der Bemessung von Spannen nach § 16 Absatz 3 herangezogen werden.*

Diese gesetzliche Vorgabe lässt viel Spielraum, auch was die statistischen Implikationen anbetrifft. Malottki et al (2018) beschreiben ausführlich Notwendigkeit, Relevanz und Umsetzung der Berücksichtigung von außergesetzlichen Merkmalen bei der Erstellung von Mietspiegeln. Die statistischen Auswirkungen auf Mietspie-

gelmodelle werden dort nur kurz angesprochen. Wir greifen diesen Punkt weiter auf und diskutieren aus statistischer Sicht unterschiedliche Möglichkeiten der Berücksichtigung von außergesetzlichen Merkmalen in Regressionsmietspiegeln. Wir vergleichen anhand von realen Daten aus München die entsprechenden Konsequenzen in Bezug auf Verzerrung, Bestimmtheitsmaß und Vorhersagegüte. Wir betrachten darüber hinaus Aspekte der Modellselektion.

Der Artikel ist wie folgt gegliedert. Abschn. 2 gibt den theoretischen Rahmen und zeigt auf, welche Möglichkeiten es aus statistischer Sicht gibt außergesetzliche Merkmale zu berücksichtigen. Abschn. 3 betrachtet die Berücksichtigung von außergesetzlichen Merkmalen bei der Modellwahl und Abschn. 4 gibt eine abschließende Diskussion des Themas.

2 Mietspiegelmodelle

2.1 Gesetzliche und außergesetzliche Merkmale

Wir bezeichnen im folgenden mit y die Nettomiete pro Quadratmeter und mit x den Vektor der gesetzlichen Merkmale, also Wohnfläche, Lage, Ausstattung etc. Mit z notieren wir den Vektor der außergesetzlichen Merkmale, die unter anderem die Mietdauer oder der Vermietertyp sind. Wir betrachten hier der Einfachheit halber ein lineares Regressionsmodell, wobei Erweiterungen auf nicht-lineare Strukturen direkt möglich sind, in der Notation aber umständlich wirken, so dass wir darauf hier verzichten. Das vollständige lineare Modell, welches gesetzliche und außergesetzliche Merkmale enthält, ergibt sich damit zu

$$y = \beta_0 + x\beta_x + z\beta_z + \epsilon. \quad (1)$$

Wir unterstellen nachfolgend, dass das Modell (1) das datengenerierende Modell ist. Das bedeutet, dass keine weiteren erklärenden Variablen verfügbar sind und die Reststreuung ϵ unabhängig von den Einflussgrößen x und z ist. Basierend auf Daten (y_i, x_i, z_i) mit $i = 1, \dots, n$ kann nun der Kleinste-Quadrate-Schätzer (KQ-Schätzer) bzw. bei Annahme einer Normalverteilung für die Residuen der Maximum-Likelihood Schätzer (ML-Schätzer) bestimmt werden, der bekannterweise

$$\hat{\beta} = ((\mathbf{X}, \mathbf{Z})^T (\mathbf{X}, \mathbf{Z}))^{-1} ((\mathbf{X}, \mathbf{Z})^T \mathbf{y}) \quad (2)$$

ist. Dabei ist $\mathbf{y} = (y_1, \dots, y_n)^T$ und \mathbf{X} die Designmatrix mit Zeilen $(1, x_i)$ und \mathbf{Z} entsprechend die Designmatrix mit Zeilen z_i , wobei $i = 1, \dots, n$ der Beobachtungsindex ist. Die geschätzten Koeffizienten können nun für ein Mietspiegelmodell herangezogen werden, was wir nachfolgend als *Mietspiegelmodell 0* bezeichnen wollen. Für eine Wohnung mit gesetzlichen Merkmalen x_* und außergesetzlichen Merkmalen z_* ergibt sich in diesem Modell die ortsübliche Vergleichsmiete zu

$$\hat{y}_{*(0)} = \hat{\beta}_0 + x_* \hat{\beta}_x + z_* \hat{\beta}_z. \quad (3)$$

Das Modell kommt in der Praxis nicht zur Anwendung, da die ortsübliche Vergleichsmiete hier neben den gesetzlichen Merkmalen x auch außergesetzliche Merkmale z enthält, was im qualifizierten Mietspiegel aufgrund gesetzlicher Vorgaben nicht erlaubt ist. Modell (1) kann aber als „Benchmark“ dienen, sprich als das bestmögliche Modell betrachtet werden.

Es ist plausibel anzunehmen, dass in erhobenen Mietspiegeldaten gesetzlichen Merkmale x und außergesetzlichen Merkmale z miteinander korreliert sind. Art und Beschaffenheit des Fußbodenbelags einer Wohnung hängen vom Einzugsdatum und damit von der Mietdauer ab. Damit hängen Fußbodenbelag und Mietdauer zusammen. Ebenso sind Wohnungen, die über Genossenschaften vermietet werden, tendenziell einfacher ausgestattet als Wohnungen im gehobenen Segment. Das impliziert, dass Ausstattung mit dem Vermietertyp assoziiert ist. Diese Assoziationen können einen *omitted variable Bias* zur Folge haben (Wooldridge 2015). Bisherige Mietspiegelmodelle haben dabei die außergesetzlichen Merkmale gänzlich unberücksichtigt gelassen. Dies war gesetzlich in dieser Form vorgeschrieben, was aber explizite Konsequenzen für das resultierende Mietspiegelmodell hat. Konkret bedeutet die Nichtberücksichtigung von außergesetzlichen Merkmalen, dass man die Einflussgrößen z ignoriert und damit das vereinfachte Modell

$$y = \beta_0 + x\beta_x + \tilde{\epsilon} \quad (4)$$

betrachtet, wobei $\tilde{\epsilon} = z\beta_z + \epsilon$. Im Modell (4) ist das Residuum $\tilde{\epsilon}$ nicht unabhängig von x , sondern es gilt

$$\frac{\partial \tilde{\epsilon}}{\partial x} = \frac{\partial z}{\partial x} \beta_z. \quad (5)$$

Diese Korrelation wird bei der Schätzung der Parameter in Modell (4) ignoriert, sofern zum Beispiel ein Kleinste-Quadrate Schätzer (KQ) oder ein Maximum-Likelihood Schätzer (ML) von Modell (4) verfolgt wird. Hierbei wird Unkorreliertheit von den Residuen und den Kovariablen unterstellt. Der entsprechende Schätzer sei als $\hat{\beta}_x$ bezeichnet. Er ist dabei nicht konsistent für β_x in Modell (1), sondern konsistent für

$$\tilde{\beta}_x = \beta_x + \underbrace{(\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Z})}_{\text{Bias}} \beta_z. \quad (6)$$

Man erkennt, dass die Schätzer im Modell (4) mit nur gesetzlichen Merkmalen verzerrt sind, wenn x und z korreliert sind. Der Bias ist dabei durch Gl. (6) explizit berechenbar.

Auch wenn der Schätzer $\hat{\beta}$ verzerrt ist, so weist er eine Optimalität auf, weil sich die vorhergesagte Miete als bester *linearer Prädiktor* aus den gesetzlichen Merkmalen ergibt. Wir bezeichnen das Modell als *Mietspiegelmodell 1* und die ortsübliche Vergleichsmiete für eine Wohnung mit Ausstattung x_* ergibt sich zu

$$\widehat{y}_{*(1)} = \widetilde{\beta}_0 + x_* \widetilde{\beta}_x. \quad (7)$$

Man beachte, dass Modell (7) ein klassisches Mietspiegelmodell ist und damit der Mehrzahl der auf dem Markt befindlichen Regressionsmietspiegel entspricht.

Mit dem Mietspiegelreformgesetz erlaubt der Gesetzgeber erstmals, dass außergesetzliche Merkmale zur Wahl des Regressionsmodell herangezogen werden können (Bundesgesetzblatt 2021). Eine Möglichkeit dies im Modellierungsschritt zu berücksichtigen ist, dass zunächst die außergesetzlichen Merkmale mitberücksichtigt werden, aber dann für die konkrete Prognose der ortsüblichen Miete vernachlässigt werden. Statistisch kann dies wie folgt formuliert werden. Das volle, datengenerierende Modell (1) wird mit allen verfügbaren Daten geschätzt. Für die Berechnung der ortsüblichen Vergleichsmiete einer Wohnung dürfen per gesetzlicher Vorgabe aber nur die gesetzlichen Merkmale herangezogen werden, d. h. nur die x_* . Es müssen also Annahmen getroffen werden, wie bei der Prognose mit dem Term $z_* \widehat{\beta}_z$ aus dem geschätzten Modell (3) umgegangen wird. Für die Prognose nehmen wir an, dass die außergesetzlichen Merkmale z_* unbekannt sind. Statistischer Terminologie folgend kann man dies als ein Problem von fehlenden Daten (*missing data*) auffassen. Damit kann das geschätzte Modell (1) nicht direkt zur Vorhersage herangezogen werden, da die Kovariablen z_* unbekannt sind. Eine Lösung des Vorhersageproblems ist nun, den Wert z_* aus den Daten zu schätzen. Das einfachste Vorhersagemodell hierzu ist das arithmetische Mittel $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$ als Schätzer für die als unbekannt angenommen außergesetzlichen Merkmale heranzuziehen. In diesem Modell wird bei der Prognose einer konkreten Wohnung für alle außergesetzlichen Merkmale ein „durchschnittlicher“ Wert angenommen. Damit ergibt sich die ortsübliche Vergleichsmiete durch

$$\widehat{y}_{*(2)} = \widehat{\beta}_0 + x_* \widehat{\beta}_x + \bar{z} \widehat{\beta}_z. \quad (8)$$

Dabei kann man $\widehat{\beta}_0$ und $\bar{z} \widehat{\beta}_z$ zum neuen Intercept $\widehat{\beta}_0$ zusammenfassen. Numerisch identische Ergebnisse erhält man, indem man die geschätzten Werte $\widehat{\beta}_x$ nimmt, dann aber für das Modell noch einmal den Intercept entsprechend nachschätzt. Wir bezeichnen diesen Ansatz nachfolgend als *Mietspiegelmodell 2*. Dieser Ansatz ist mathematisch identisch mit dem in Malottki et al (2018) in Abschn. V.2. (Kontrolle der Merkmale) beschriebenen Verfahren. Durch die Verwendung von \bar{z}_* als Schätzwert des Populationsmittels der außergesetzlichen Merkmale z hat die empirische Verteilung der Werte z_* in der realisierten Stichprobe direkten Einfluss auf die vorhergesagten Werte der ortsüblichen Vergleichsmiete $\widehat{y}_{*(2)}$ als Schätzer für $E(y_* | x_*, z_*)$. Weist die Stichprobe einen Bias bezüglich der außergesetzlichen Merkmalen z auf, so sind auch die vorhergesagten Werte der ortsüblichen Vergleichsmiete verzerrt. Voraussetzung für die Anwendung des Mietspiegelmodells 2 ist also das Ziehen einer unverzerrten Stichprobe.

Die Sichtweise, dass wir für eine Prognose die außergesetzlichen Merkmale als fehlende Werte betrachten und diese durch den Mittelwert schätzen, führt zu einem weiteren Ansatz. Klarerweise ist der Mittelwert die einfachste Form, um fehlende Werte zu ersetzen. Man kann aber auch versuchen den als unbekannt angenommenen

Wert z aus den verfügbaren Daten zu schätzen, ähnlich wie es im Bereich der multiplen Imputation verfolgt wird. Das bedeutet, dass wir ein Vorhersagemodell von z aus den Variablen x und im Prinzip auch aus y bestimmen müssen. Wir vernachlässigen dabei die Werte von y und unterstellen ein (generalisiertes) Regressionsmodell in der Form

$$E(z | x) = h(\eta_x) \quad (9)$$

wobei η_x ein linearer Prädiktor und $h()$ als Linkfunktion verstanden werden muss, so dass der Wertebereich von beiden Seiten der Gleichung identisch ist. Ist z eine binäre Größe, so bietet sich zum Beispiel der logit-Link an. Man beachte, dass z üblicherweise multidimensional ist und damit Gl. (9) elementweise gilt. Der lineare Prädiktor kann nun beliebige Strukturen annehmen, wobei wir hier der Einfachheit halber nur lineare Abhängigkeiten unterstellen. Das bedeutet konkret, dass wir für das Modell (9) unterstellen:

$$\eta_x = \alpha_0 + x\alpha_x. \quad (10)$$

Basierend auf den Daten (x_i, z_i) , $i = 1, \dots, n$ und Modell (9) (bei Unterstellung einer i.i.d. Annahme) können die Koeffizienten α mittels der Maximum-Likelihood-Methode geschätzt werden und die resultierenden Schätzer bezeichnen wir als $\hat{\alpha}$. Für eine Wohnung mit Ausstattungsmerkmalen x_* kann damit der erwartete Wert von z bestimmt werden durch $\hat{z}_* := h(\hat{\alpha}_0 + x_*\hat{\alpha}_x)$. Dieser Wert kann wiederum in das Mietspiegelmodell 0 eingesetzt werden und es ergibt sich die ortsübliche Vergleichsmiete durch die Prognose

$$\hat{y}_{*(3)} = \hat{\beta}_0 + x_*\hat{\beta}_x + \hat{z}_*\hat{\beta}_z. \quad (11)$$

Wir bezeichnen dieses Modell als *Mietspiegelmodell 3*. Dabei berücksichtigen die Mietspiegelmodelle 2 und 3 die außergesetzlichen Merkmale in der Modellierungsphase, nicht aber im konkreten Mietspiegelmodell zur Prognose der ortsüblichen Vergleichsmiete. Das Modell 3 nutzt die außergesetzlichen Merkmale nicht bei der Berechnung der ortsüblichen Vergleichsmiete. Sie fließen aber indirekt über die multiple Imputation ein, indem für diese Wohnung aus den zugehörigen gesetzlichen Merkmalen x_* ein „üblicher“ Einfluss der z_* geschätzt wird. Über diesen indirekten Einfluss hat das Modell 3 den Vorteil, dass das Auftreten von bestimmten Wertekonstellationen von gesetzlichen und außergesetzlichen Merkmalen tendenziell mitberücksichtigt wird (siehe auch Malotki et al 2018, S. 13). So werden z. B. durch die Vorhersage der außergesetzlichen Merkmale für Neubauwohnungen tendenziell kürzere Mietdauern vorhergesagt.

Da wir uns in diesem Artikel auf die statistischen Möglichkeiten und Auswirkungen der Berücksichtigung außergesetzlicher Merkmale beschränken wollen, ist eine juristische Einordnung und Prüfung aller Ansätze vor dem Einsatz in der Praxis notwendig.

Tab. 1 Gesetzliche Wohnwertmerkmale im Modell zum Mietspiegel für München 2019

Variable	Skala	Ausprägungen/Codierung
Wohnfläche in m^2	metrisch	(WFL)
Baujahr	metrisch	(BJ)
Wohnlage	nominal	durchschnittliche Lage gute Lage (X_1) beste Lage (X_2)
zentrale Lage	binär	1 = zentrale Lage (X_3) 0 = sonst
Gebäudetyp	nominal	sonstiger Hochhaus (X_4) Wohnblock (X_5)
Haustyp	nominal	sonstiger Altbau Typ A (X_6) Altbau Typ B (X_7) Nachkriegsbau (X_8)
Warmwasserversorgung unvollständig	binär	1 = unvollständige WW (X_9) 0 = sonst
Heizung unvollständig	binär	1 = unvollständige HZ (X_{10}) 0 = sonst
Badausstattung	metrische	Score (0,1,2,3) Summe Ausstattungen Bad (X_{11})
offene Küche	binär	1 = offene Küche (X_{12}) 0 = sonst
Küchenausstattung	metrisch	Score (0,1,2,3) Summe Küchenausstattungen (X_{13})
gute Ausstattung	binär	1 = gute Ausstattung (X_{14}) 0 = sonst
Fußboden	nominal	weder einfacher noch guter Boden einfacher Boden (X_{15}) guter Boden (X_{16})
neuer Fußboden	binär	1 = neuer Fußbodenbelag (X_{17}) 0 = sonst

2.2 Empirische Evaluation der Mietspiegelmodelle

Im nachfolgenden Abschnitt wollen wir anhand von konkreten Daten des Mietspiegels für München den Effekt der Berücksichtigung von außergesetzlichen Merkmalen quantifizieren. Die Mietspiegelmodelle 1 bis 3 werden verglichen mit dem Mietspiegelmodell 0, also dem datengenerierenden Modell aus Gl. (1), das als eine Art Benchmark dient. Betrachtet werden einerseits die Unterschiede der Vorhersagegüte der Nettokaltmieten je Quadratmeter y und andererseits der Effekt auf die Koeffizientenschätzer $\hat{\beta}_x$, also die Wirkung auf die gesetzlichen Wohnwertmerkmale.

Für unsere Analyse nutzen wir Daten aus der Erhebung zum Mietspiegel für München 2021. Als außergesetzliche Merkmale dienen die Mietdauer z_1 , in Form der Laufzeit des Mietvertrages in Monaten und der Vermietertyp z_2 , als binäre Variable mit $z_2 = 1$, falls der Vermietertyp eine Genossenschaft oder städtische Gesellschaft ist und $z_2 = 0$ sonst. Die Mietdauer z_1 wird als „glatter“, nicht-parametrischer

Tab. 2 Geschätzte Koeffizienten und Standardabweichungen im Mietspiegelmodell 0 und 1

Variable		$\hat{\beta}_0$	$sd(\hat{\beta}_0)$	$\hat{\beta}_1$	$sd(\hat{\beta}_1)$
Wohnlage					
gute Wohnlage	<i>(X₁)</i>	0,69	0,13	0,69	0,15
beste Wohnlage	<i>(X₂)</i>	2,20	0,28	2,14	0,32
zentrale Lage	<i>(X₃)</i>	1,16	0,14	1,13	0,16
Gebäudetyp					
Hochhaus	<i>(X₄)</i>	-0,37	0,25	-0,55	0,29
Wohnblock	<i>(X₅)</i>	-0,50	0,18	-1,20	0,21
Haustyp					
Altbau A	<i>(X₆)</i>	-1,13	0,30	-1,38	0,35
Altbau B	<i>(X₇)</i>	-1,45	0,32	-2,44	0,36
Nachkriegsbau	<i>(X₈)</i>	-0,45	0,17	-0,60	0,20
Warmwasserversorgung unvollständig	<i>(X₉)</i>	-1,34	0,49	-2,67	0,56
Heizung unvollständig	<i>(X₁₀)</i>	-3,35	0,53	-2,99	0,61
Badausstattung	<i>(X₁₁)</i>	0,33	0,08	0,54	0,09
offene Küche	<i>(X₁₂)</i>	0,52	0,17	0,70	0,20
Küchenausstattung	<i>(X₁₃)</i>	0,30	0,05	0,54	0,06
gute Ausstattung	<i>(X₁₄)</i>	0,96	0,21	1,11	0,24
Fußboden					
einfacher Fußboden	<i>(X₁₅)</i>	-0,27	0,29	-0,82	0,33
guter Fußboden	<i>(X₁₆)</i>	0,37	0,18	1,21	0,20
neuer Fußboden	<i>(X₁₇)</i>	0,61	0,20	1,43	0,23
Vermietertyp	<i>(Z₂)</i>	-3,17	0,25	—	—

Effekt im Regressionsmodell aufgenommen. Ein außergesetzliches Merkmal mit zweifelsfrei hohem Einfluss auf die Miethöhe ist die Unterscheidung zwischen neu abgeschlossenen Mietverträgen sowie Wohnungen mit geänderter Bestandsmiete, das sich direkt aus der gesetzlichen Definition zur ortsüblichen Vergleichsmiete (§ 558 (2) BGB) ergibt. Diese Unterscheidung weist aber eine starke Assoziation mit der Mietdauer auf, da laut § 558 (2) BGB die ortsübliche Vergleichsmiete u. a. aus den Neuverträgen der letzten sechs Jahre gebildet wird. Eine Berücksichtigung neben der Mietdauer enthält nur wenig zusätzliche Information und eine Aufnahme in das Regressionsmodell ist nicht sinnvoll.

Insgesamt stehen in dem Datensatz 1975 Beobachtungen zur Verfügung, die Informationen über die gesetzlichen Merkmale x und die außergesetzlichen Merkmale z enthalten. Die signifikanten Wohnwertmerkmale des Mietspiegels für München aus dem Jahr 2019 bilden die gesetzliche Merkmale x , welche in Tab. 1 aufgeführt sind. Detailliertere Definitionen der Merkmale sind in der Dokumentation zum Mietspiegel für München 2019 (Windmann und Kauermann 2019, S. 30ff) zu finden. Für die nominal skalierten Variablen wurde eine Dummy-Codierung verwendet. Die jeweilige Referenzkategorie ist kursiv dargestellt. Die Merkmale Wohnfläche und Baujahr werden als „glatte“ nicht-parametrische Effekte in das Modell aufgenommen. Das zu

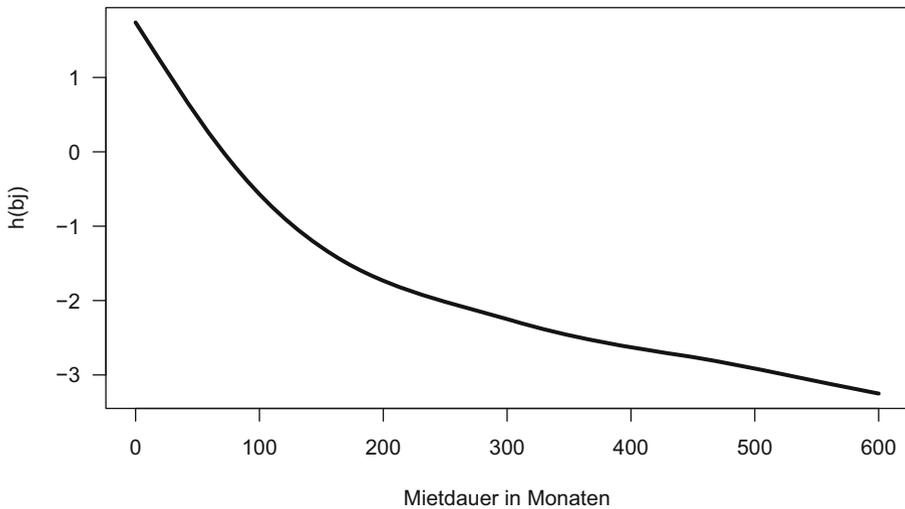


Abb. 1 Geschätzter Effekt $\hat{h}(z_1)$ im Modell 0

schätzende Regressionsmodell mit gesetzlichen und außergesetzlichen Merkmalen hat die Form eines generalisierten additiven Modells (Wood 2017)

$$y = \beta_0 + f(WFL) + g(BJ) + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{17} X_{17} + h(z_1) + z_2 \beta_z + \epsilon \quad (12)$$

wobei $f()$, $g()$ und $h()$ jeweils „glatte“, sprich stetige bzw. differenzierbare Funktionen sind, die aus des Daten geschätzt werden. Das Modell lässt sich mit verfügbarer Software schätzen (Wood et al 2016).

Eine Modellwahl findet bei den betrachteten Mietspiegelmodellen 1 bis 3 an dieser Stelle nicht statt, so dass in allen Modellen dieselben Variablen x und z herangezogen werden. Mögliche Auswirkungen auf die Modellwahl betrachten wir im nachfolgenden Abschnitt. Zur Beurteilung des Einflusses außergesetzlicher Merkmale auf die vorgeschlagenen Mietspiegelmodelle brachten wir drei Größen:

1. Bias (6) der gesetzlichen Merkmale,
2. Bestimmtheitsmaß R^2 für die Modelle sowie die resultierende Residualvarianz und schließlich
3. die Vorhersagegenauigkeit basierend auf einer Kreuzvalidierung.

Es sei drauf hingewiesen, dass weder Bestimmtheitsmaß noch Varianz der kreuzvalidierten Residuen (Vorhersagegenauigkeit) ein Güte- oder Qualitätsmaß für Mietspiegel darstellen. Diese Maßzahlen dienen hier nur dazu die Folgen der drei Ansätze zur Berücksichtigung außergesetzlicher Merkmale aufzuzeigen. Tab. 2 zeigt für die parametrischen Terme der Modelle 0 und 1 die geschätzten Koeffizienten mit den zugehörigen Standardabweichungen. Abb. 1 zeigt den geschätzten Effekt der Mietdauer $\hat{h}(z_1)$ im Modell 0.

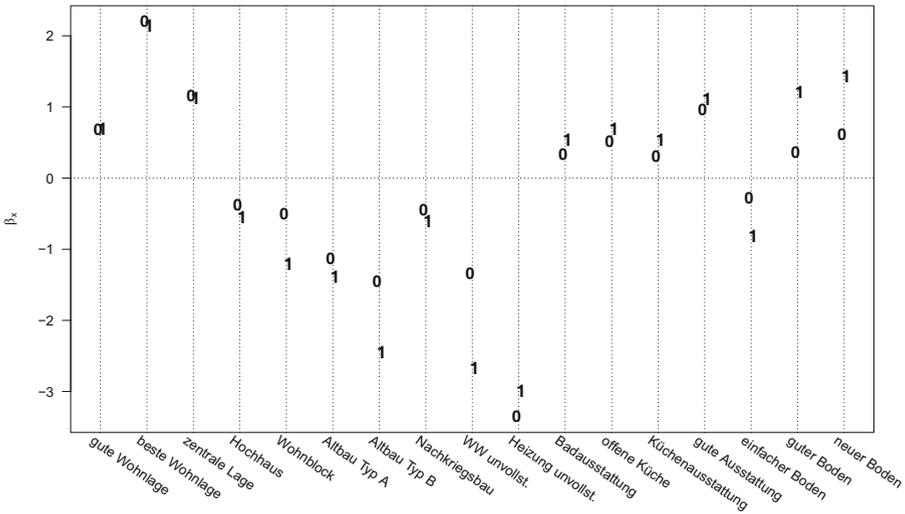


Abb. 2 Geschätzte Koeffizienten $\hat{\beta}_x$ des datengenerierenden Modells ('0') und $\hat{\beta}_x$ des Mietspiegelmodells 1 ('1')

Bias

Abb. 2 zeigt die geschätzten Koeffizienten $\hat{\beta}_x$ der unterschiedlichen Ansätze der Mietspiegelmodelle 1 bis 3 auf die Höhe der Zu- und Abschläge der gesetzlichen Wohnwertmerkmale. Dabei ist es ausreichend nur die Koeffizienten $\hat{\beta}_x$ des Mietspiegelmodells 1 (in Abb. 2 mit '1' dargestellt) mit den Koeffizienten $\hat{\beta}_x$ der Mietspiegelmodelle 2, 3 und dem datengenerierenden Modell 0 (in Abb. 2 mit '0' dargestellt) zu vergleichen, da für die letzten drei Modelle die Schätzungen der Koeffizienten $\hat{\beta}_x$ identisch sind. Abweichungen zwischen $\hat{\beta}_x$ und $\hat{\beta}_x$ stellen eine Schätzung des Bias in Gl. (6) dar. Abb. 3 und 4 zeigen die geschätzten „glatten“ nicht-parametrischen Effekte $\hat{f}(WFL)$ und $\hat{g}(BJ)$ für die Variablen Wohnfläche und Baujahr.

Es zeigt sich, dass für einige Wohnwertmerkmale die Berücksichtigung der außergesetzlichen Merkmalen nur zu geringen Unterschieden führt, so z.B. bei der Wohnlage und zentraler Lage sowie Bad- und Küchenausstattung. Deutliche Unterschiede gibt es hingegen bei einigen Gebäudetypen (Wohnblock, Altbau Typ B) sowie Qualität und Alter des Fußbodenbelags. Für diese Merkmale ist also eine Abhängigkeit zwischen den außergesetzlichen Merkmalen z und gesetzlichen Merkmalen x vorhanden. Diese Wohnwertmerkmale weisen im Mietspiegelmodell 1, also ohne Berücksichtigung außergesetzlicher Merkmale, betragsmäßig größere Werte der Koeffizienten auf, da sie nicht nur einen „kausalen“ Effekt von x sondern auch die Effekte der außergesetzlichen Merkmale z beinhalten. Insbesondere für den Bodenbelag ist dies durchaus erklärbar. Ein Wechsel bzw. eine Erneuerung des Bodenbelags findet bei einer Mietwohnung üblicherweise bei Mieterwechsel statt, was wiederum mit einem neuen Mietvertrag und damit einer kürzeren Mietdauer einhergeht. Im Mietspiegelmodell 1 ohne die außergesetzlichen Merkmale wird der

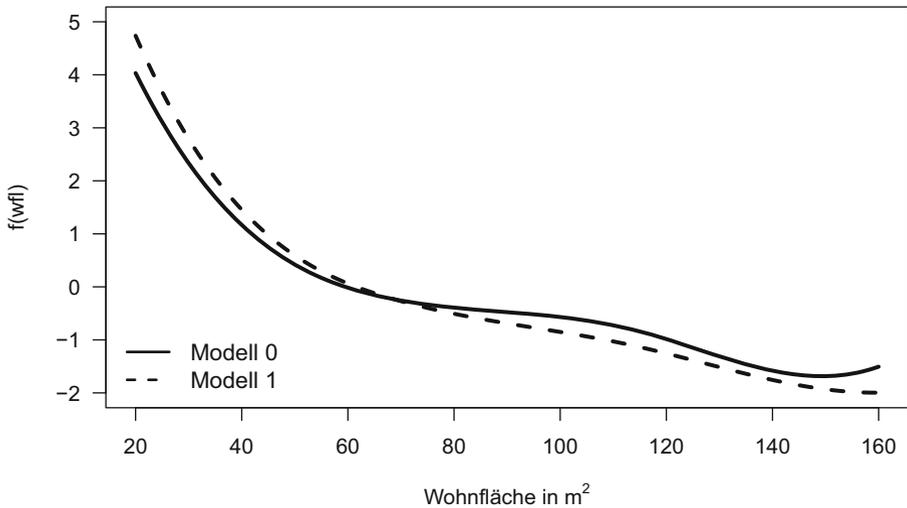


Abb. 3 Geschätzter „glatter“ nicht-parametrischer Effekt der Wohnfläche

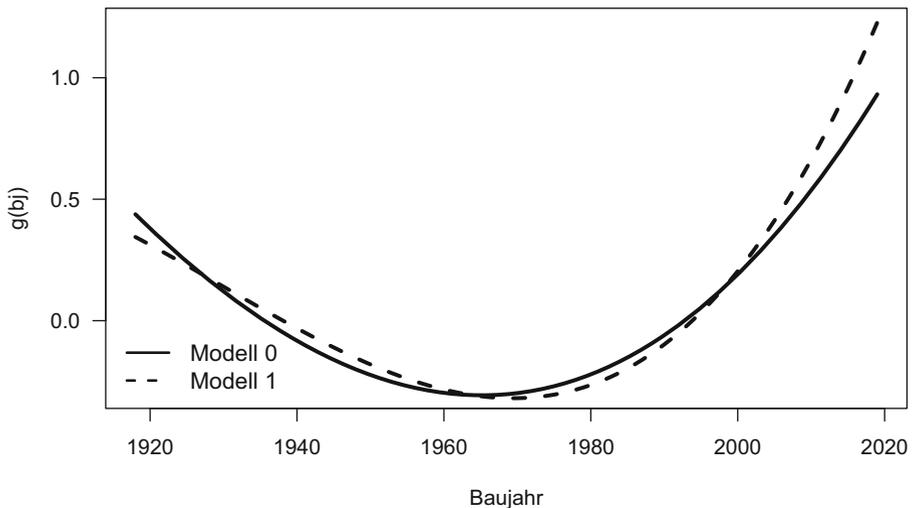


Abb. 4 Geschätzter „glatter“ nicht-parametrischer Effekt des Baujahrs

Effekt vom Bodenbelag überschätzt, da ein direkter Einfluss der Mietdauer unberücksichtigt bleibt.

Die genannten Abhängigkeiten zwischen den gesetzlichen Merkmalen x und den außergesetzlichen Merkmalen z zeigen sich auch im Assoziationsmaß Cramers' V , welches in Abb. 5 gezeigt wird. Dafür wurden die stetigen Merkmale Wohnfläche, Baujahr und Mietdauer kategorisiert. Aus der Grafik erkennt man, dass die Mietdauer z_1 starke Abhängigkeiten zu den Variablen Art des Fußbodens und neuer Boden hat. Bei dem Vermietertyp z_2 „Genossenschaft, städtischer Vermieter“ können hohe



Abb. 5 Cramers'V der gesetzlichen und außergesetzlichen Merkmale x und z

Abhängigkeiten zum Haus- und Gebäudetyyp sowie dem Baujahr beobachtet werden. Die gesetzlichen Merkmale Art des Fußbodens, neuer Boden, Haustyp (konkret Wohnblock) und Gebäudetyyp (konkret Altbau Typ B) haben, wie in Abb. 2 gezeigt, einen hohen Bias der geschätzten Koeffizienten β_x im Mietspiegelmodell 1.

Bestimmtheitsmaß R^2

Tab. 3 zeigt die Werte des Bestimmtheitsmaßes R^2 und den Anteil des Bestimmtheitsmaßes R_i^2/R_0^2 des Mietspiegelmodells $i = 1, 2, 3$ am datengenerierenden Modell 0. Darüber hinaus ist die geschätzte Standardabweichung der Residuen für die vier Modelle angegeben. Aus der Schätzung des datengenerierenden Modells 0 zeigt sich, dass die verwendeten gesetzlichen Merkmale x und die außergesetzlichen Merkmale z zusammen etwa 53% der Streuung der Nettomieten erklären können. Die außergesetzlichen Merkmale Mietdauer und Vermieterotyp erklären einen

Tab. 3 Bestimmtheitsmaß R^2 , Standardabweichung ϵ der geschätzten Residuen und Standardabweichung $\tilde{\epsilon}$ aus einer Kreuzvalidierung für das datengenerierende Modell 0 und die Mietspiegelmodelle 1 bis 3

	Datengenerierender Prozess	Modell 1	Modell 2	Modell 3
R^2	0,53	0,39	0,20	0,47
Anteil R^2		74%	38%	89%
$sd(\epsilon)$	2,48	2,85	2,97	2,89
$sd(\tilde{\epsilon})$	2,53	2,90	3,01	2,92

nicht unerheblichen Teil der Heterogenität der Nettomieten. Das zeigt sich durch die niedrigeren R^2 -Werte der Modelle 1 bis 3. Den niedrigsten Wert hat dabei das Mietspiegelmodell 2, also das Modell, das zwar sowohl aus gesetzlichen als auch außergesetzlichen Merkmalen geschätzt wird, bei dem aber dann nur der Achsenabschnitt neu bestimmt wird. Werden hingegen die außergesetzlichen Merkmale durch ein eigenes Modell wie in (9) skizziert durch die gesetzlichen Merkmale geschätzt, so kann das Bestimmtheitsmaß im Vergleich zum üblichen Mietspiegelmodell 1 deutlich gesteigert werden.

Der R^2 -Wert des datengenerierenden Modells 0 dient als Benchmark für den Vergleich der Ansätze der Mietspiegelmodelle 1 bis 3. Es zeigt sich, dass mit dem klassischen Ansatz im Mietspiegelmodell 1 74% der Streuung im Vergleich zum Benchmarkmodell erklärt werden können. Im Mietspiegelmodell 2 sind es mit 38% weniger als die Hälfte des Benchmarkmodells. Das bedeutet, dass der kausale Einfluss der gesetzlichen Merkmale nur einen geringen Teil der Streuung der Nettomieten erklären kann. Das Mietspiegelmodell 3, in dem die außergesetzlichen Merkmale mittels der gesetzlichen Merkmale imputiert werden, erklärt 89% des Benchmarkmodells. Der Informationsverlust durch das Imputieren ist hier eher gering. Die Streuung der Residuen in Form der Standardabweichungen geben ein ähnliches Bild wieder.

Vorhersagegüte durch Kreuzvalidierung

Die obigen Analysen betreffen „in sample“ Größen, also Güteaussagen zum Datensatz, der zum Schätzen (Lernen) des Modells benutzt wurde. Wir erweitern dies nun auf „out of sample“ Analysen und betrachten die Varianz der kreuzvalidierten Residuen, die sich aus einer Leave-One-Out-Analyse ergeben. Sei \hat{y}_i^{-i} die Vorhersage der Nettomiete für die i -te Wohnung in der Stichprobe, wobei diese i -te Wohnung bei der Modellschätzung nicht berücksichtigt wird, dann sind die kreuzvalidierten Residuen $\tilde{\epsilon}_i = y_i - \hat{y}_i^{-i}$, ($i = 1, \dots, n$). Die so definierten kreuzvalidierten Residuen werden auch im Mietspiegel für München (Windmann und Kauermann 2019, S. 57) zur Berechnung der Spannen verwendet. Tab. 4 zeigt für die Mietspiegelmodelle 1 bis 3 und das Benchmarkmodell 0 die Standardabweichung sowie das 1/6 und 5/6-Quantil, was in Mietspiegeln der üblichen 2/3-Spanne entspricht.

Die Ergebnisse sind vergleichbar mit denen des Bestimmtheitsmaßes R^2 aus Tab. 3. Das Benchmarkmodell 0 zeigt, wie auch zu erwarten war, die beste Vorhersagegüte. Die Modelle 1 und 3 weisen eine ähnlich hohe Streuung der kreuzvalidierten Residuen auf und das Modell 2 die stärkste Streuung. Diese Reihenfolge

Tab. 4 Standardabweichung sowie 1/6 und 5/6-Quantile der kreuzvalidierten Residuen $\tilde{\epsilon}$

Modell	$sd(\tilde{\epsilon})$	1/6-Quantil	5/6-Quantil
Benchmarkmodell 0	2,53	-2,05	2,05
Mietspiegelmodell 1	2,90	-2,44	2,42
Mietspiegelmodell 2	3,00	-2,56	2,61
Mietspiegelmodell 3	2,92	-2,14	2,87

zeigt sich auch in der Spannweite der 1/6 und 5/6-Quantile. In den Modellen 0, 1 und 2 weisen die kreuzvalidierten Residuen eine symmetrische Verteilung auf. Im Modell 3 dagegen weisen die positiven Residuen eine größere Spannweite auf als die negativen.

Zusammenfassend kann man festhalten, dass die Berücksichtigung außergesetzlicher Merkmale bei der Mietspiegelerstellung je nach Ansatz einen Einfluss auf die Vorhersagegüte und auch den Bias der geschätzten Effekte der gesetzlichen Parameter hat. Für einige Wohnwertmerkmale x ändert die Berücksichtigung der außergesetzlichen Merkmale demnach die Wirkung auf die ortsübliche Vergleichsmiete, wobei das Nicht-Berücksichtigen zu einer stärkeren Wirkung der Wohnwertmerkmale x führt. Will man außergesetzliche Merkmale berücksichtigen, so zeigt das Modell 2 die schlechtesten statistischen Eigenschaften und es sollte Modell 3 bevorzugt werden. Inhaltlich ist das nachvollziehbar, da Modell 2 die unbekanntesten außergesetzlichen Merkmale durch das arithmetische Mittel ersetzt, wohingegen Modell 3 das Regressionsmodell (9) benutzt. Der Vorteil von Modell 3 ist darüber hinaus, dass es die Effekte β_x unverzerrt schätzt, wohingegen Modell 1 einen Bias aufweist, wie in Abb. 2 und 4 zu sehen ist.

3 Modellwahl

Im Abschn. 2 haben wir den Einfluss der außergesetzlichen Merkmale auf die Vorhersagegüte und den Bias der geschätzten Koeffizienten der gesetzlichen Wohnwertmerkmale betrachtet. In diesem Abschnitt wird der Einfluss verschiedener Ansätze der Modellwahl auf die in den Regressionsmodellen enthaltenen Merkmale und die Vorhersagegüte untersucht. Wir verwenden wieder die Daten des Mietspiegels für München 2021 und ziehen als Benchmarkmodell das Mietspiegelmodell aus Gl. (3) heran.

Tab. 5 Bestimmtheitsmaß R^2 und Standardabweichung der Residuen $sd(\epsilon)$ als Maß für Vorhersagegüte bei den vier unterschiedlichen Ansätzen der Modellwahl

Modellwahl	Prädiktion	R^2	$sd(\epsilon)$
ohne z	ohne z , Modell 1	0,38	2,86
mit z	ohne z , Modell 1	0,38	2,86
ohne z	mit z , Modell 2	0,20	2,97
mit z	mit z , Modell 2	0,20	2,98

Bei der Modellwahl, also Wahl der Parameter, werden nur Variablen in das Regressionsmodell aufgenommen, die zum 5%-Niveau signifikant sind. Die Ansätze zur Modellwahl bestehen aus zwei Teilen. Zum einen untersuchen wir, ob bei der Wahl der signifikanten gesetzlichen Parameter x außergesetzliche Merkmale z im Regressionsmodell enthalten sind und zum anderen, ob bei der Schätzung der Koeffizienten der gesetzlichen Parameter $\hat{\beta}_x$ außergesetzliche Parameter z im Regressionsmodell enthalten sind, d. h. ob bei der Parameterwahl bzw. der Koeffizientenschätzung das Mietspiegelmodell 1 (ohne außergesetzliche Merkmale z) oder Mietspiegelmodell 2 (mit außergesetzlichen Merkmalen z) angewandt wird. Tab. 5 zeigt das Bestimmtheitsmaß R^2 und die Standardabweichungen der Residuen ϵ für die vier möglichen Ansätze der Modellwahl.

Aus den R^2 -Werten und der Standardabweichung der Residuen zeigt sich, dass bei der Modellwahl, ob mit oder ohne gesetzliche Merkmale z , fast keine Unterschiede in der Vorhersagegüte entstehen. Dies deutet daraufhin, dass die im Mietspiegel für München enthaltenen gesetzlichen Merkmale x in diesem Fall nicht weiter reduziert werden können.

4 Abschließende Bemerkungen

In diesem Artikel haben wir gezeigt, wie außergesetzliche Merkmale in der Modellierung von Regressionsmietspiegeln berücksichtigt werden können. Anhand einer empirischen Analyse mittels Daten aus dem Mietspiegel für München haben wir die Folgen untersucht, die für die Vorhersagegüte und den Bias bei der Schätzung der Koeffizienten zu erwarten sind.

Die Analyse zeigt, dass der Bias für gesetzliche Wohnwertmerkmale, die eine starke Assoziation zu außergesetzlichen Merkmalen haben, hoch ist. Dies hat zur Folge, dass die Vorhersagegüte im Mietspiegelmodell 2, das bei der Prädiktion den Einfluss außergesetzlicher Merkmale außen vor lässt, deutlich abnimmt. Im Umkehrschluss bedeutet dies, dass die außergesetzlichen Merkmale Mietdauer und Vermietertyp hier einen deutlichen Teil der Streuung der Mieten erklären. Die beste Prognosegüte weist das Modell 3 auf. Dieses Modell stellt insoweit einen Kompromiss dar, da es bei der Berechnung der ortsüblichen Vergleichsmiete nur gesetzliche Merkmale nutzt aber gleichzeitig auch den Informationsgehalt der außergesetzlichen durch ein Imputationsmodell berücksichtigt. Allerdings weist dieses Modell durch den zusätzlichen Imputationsschritt in der Anwendung eine hohe Komplexität auf und es bleibt fraglich in wieweit ein Einsatz in der Praxis möglich und sinnvoll ist. Da wir in diesem Artikel hauptsächlich die statistischen Aspekte und Folgen betrachten wollen, empfiehlt sich eine juristische Prüfung vor Einsatz in der Praxis.

Die Berücksichtigung außergesetzlicher Merkmale bei der Modellwahl führte nur zu geringen Änderungen der signifikanten Merkmale im abschließenden Modell, was darauf schließen lässt, dass fast alle Merkmale in den Daten auch einen kausalen Einfluss auf die Höhe der Mieten haben.

Subsumierend lässt sich festhalten, dass die zusätzlichen Freiheiten, die der Gesetzgeber geschaffen hat, durchaus einen großen Einfluss auf die geschätzten Koeffizienten und folglich auch auf die Prognosefähigkeit der ortsüblichen Vergleichsmiete

haben. Eine generelle statistische Empfehlung, in welcher Form die außergesetzlichen Merkmale genutzt werden sollen, kann kaum gegeben werden. Das Modell 3 ist auf Basis der statistischen Parameter klarer Favorit, es lässt aber außergesetzliche Merkmale auf Umwegen durch Imputation im finalen Modell zu und ist in der Anwendung komplexer. Insgesamt scheint es daher aus unserer Sicht sinnvoll, dass der Gesetzgeber hier nachschärft und klarere Angaben macht, wie (und ob überhaupt) außergesetzliche Merkmale genutzt werden sollen.

Financial interests The authors declare they have no financial interests.

Funding Open Access funding enabled and organized by Projekt DEAL.

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Bundesgesetzblatt (2021) Gesetz zur Reform des Mietspiegelrechts (Mietspiegelreformgesetz – msrg)
- Bundesinstitut für Bau-, Stadt- und Raumforschung (2020) Hinweise zur Erstellung von Mietspiegeln
- Freund S, Hilla V, Missal D et al (2013) Qualifizierte Mietspiegel: Verbreitung, Standardisierungsmöglichkeiten und Qualitätsdefizite. *Wohnungswirtsch Mietr* 66:259–263
- Kauermann G, Windmann M (2016) Mietspiegel heute zwischen Realität und statistischen Möglichkeiten. *AStA Wirtschaft Sozialstat Arch* 10:205–223
- Lerbs O, Sebastian S (2015) Mietspiegel aus ökonomischer Sicht – Vorschläge für eine Neuregulierung. *Beitr Immobilienwirtsch* 10:1–26
- Malottki C, Krapp M, Vaché M (2018) Außergesetzliche Mietpreisdeterminanten im Mietspiegel – Auswirkungen und statistische Behandlung. *Wohnungswirtsch Mietr* 71:665–675
- Sebastian S, Memis HI (2021) gif-Mietspiegelreport 2021 (Version 1.2)
- Windmann M, Kauermann G (2019) Statistische Analyse der Nettomieten. In: *Mietspiegel für München 2019. Statistik, Dokumentation und Analysen*, Landeshauptstadt München, Sozialreferat, Amt für Wohnen und Migration, S 19–70 (Kapitel 2)
- Wood S (2017) Generalized additive models: An introduction with R. CRC, Boca Raton
- Wood S, Pya N, Säfken B (2016) Smoothing parameter and model selection for general smooth models (with discussion). *J Am Stat Assoc* 111:1548–1575
- Wooldridge JM (2015) *Introductory econometrics: A modern approach*. Cengage learning

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.