

*Dipl.-Volkswirtin Charlotte Articus, hat an der Universität Trier ihre Diplomarbeit zum Thema „Small Area-Verfahren zur Schätzung regionaler Mietpreise“ verfasst, die mit dem vom Statistischen Bundesamt ausgelobten Gerhard-Fürst-Preis 2013 in der Kategorie „Master-/Bachelorarbeiten“ ausgezeichnet wurde. Der folgende Beitrag stellt diese Diplomarbeit näher vor.*

# Small-Area-Verfahren zur Schätzung regionaler Mietpreise

## 1 Einleitung

Die Höhe von Wohnungsmieten variiert regional sehr deutlich – beispielsweise liegen die Preise in Frankfurt am Main und seiner Umgebung weit über denen, die für ähnliche Wohnungen in ländlichen Regionen in Rheinland-Pfalz oder Thüringen gezahlt werden müssen. Ein durchschnittliches Mietniveau für ein Bundesland oder sogar für Gesamtdeutschland ist daher für viele Verwendungszwecke wenig aussagekräftig und es besteht ein Interesse an Informationen auf regionaler Ebene. Regionale Mietpreisniveaus flächendeckend zu messen ist allerdings schwierig.

Ein vielversprechender Ansatz ist die Verwendung von Daten, denen Preisinformationen aus Wohnungsinseraten in Zeitungen und Online-Portalen zugrunde liegen. Die Verwendung solcher Angebotsmieten hat die Vorteile, dass sie ohnehin verfügbar sind und dass aufgrund der ohne große Kosten erzielten hohen Fallzahlen eine regional tief gegliederte Auswertung möglich ist. So publiziert das Bundesinstitut für Bau-, Stadt- und Raumforschung im Bundesamt für Bauwesen und Raumordnung regelmäßig auf Basis solcher Daten errechnete Durchschnittsmieten auf Kreisebene.<sup>1</sup>

Angebotsmieten erfassen allerdings lediglich die von Vermietern angestrebten Mieten im Fall der Neu- oder Wiedervermietung einer Wohnung. Es ist davon auszugehen, dass Bestandsmieten – also Mieten in manchmal schon jahrelang bestehenden Mietverhältnissen – teilweise deutlich niedri-

ger sind.<sup>2</sup> Angebotsmieten geben damit zwar wertvolle Hinweise über die aktuelle Situation auf dem Mietmarkt, zeichnen aber ein unvollständiges Bild über die Preise, die Mieter in einer Region üblicherweise für Wohnraum bezahlen müssen. Für den regionalen Vergleich von Lebenshaltungskosten zum Beispiel sind sie damit kaum geeignet.

Um das Preisniveau auf dem gesamten regionalen Mietmarkt zu erfassen, müssen also zusätzliche Informationen – etwa aus der Befragung von Mieterhaushalten – herangezogen werden. Bei der Auswertung der Ergebnisse solcher Erhebungen ergibt sich aber das zentrale Problem jeder regionalisierten Auswertung von Stichprobendaten: Werden die verfügbaren Informationen aus der Stichprobe räumlich disaggregiert, führt das zu sehr kleinen Teilstichprobenumfängen für die betrachteten Regionen. Mit klassischen statistischen Verfahren ermittelte Schätzwerte weisen dann hohe Standardfehler auf und sind somit wenig verlässlich. Je stärker regional gegliedert wird, desto gravierender ist dieses Problem. Daher werden kleinräumige Auswertungen solcher Erhebungen in der Regel nicht bereitgestellt.

Vor diesem Hintergrund wird in der vorgestellten Diplomarbeit die Anwendung von Small-Area-Verfahren für die Schätzung regionaler Mietpreise geprüft. Diese Schätzverfahren sind speziell darauf ausgerichtet, auch für regionale Einheiten, für die nur wenige Beobachtungen vorliegen – sogenannte small areas –, verlässliche Statistiken anzugeben. Bisher werden diese Verfahren in Deutschland nicht

<sup>1</sup> Siehe Bundesinstitut für Bau-, Stadt- und Raumforschung (Herausgeber): „Synopse Immobilienpreisbeobachtung in Deutschland 2010. Anforderungen – Datengrundlagen – Verfahren – Produkte“, Bonn 2010, Seite 42 ff.

<sup>2</sup> Siehe Lippe, P. von der/Breuer, C. C.: „Konzept für ein wirtschaftliches System periodischer regionaler Preisindizes – Möglichkeiten zur Gewinnung regionaler Daten über Mieten und Immobilienpreise“ in Möller, J./Hohmann, E./Huschka, D. (Herausgeber): „Der weiße Fleck – Zur Konzeption und Machbarkeit regionaler Preisindizes“, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg 2010, Seite 167 ff.

genutzt, um das Potenzial vorhandener Mietpreisdaten aus Haushaltsbefragungen auszuschöpfen.<sup>3</sup>

Konkret werden in der Diplomarbeit durchschnittliche Mietpreise auf Kreisebene (Ebene 3 der Systematik der Gebiets-einheiten für die Statistik – NUTS 3) für 13 der 16 deutschen Bundesländer geschätzt. Die Ergebnisse werden zudem eingehend evaluiert, um zu prüfen, ob Small-Area-Verfahren für die Regionalisierung von Mietpreisen geeignet sind. Im Folgenden werden die angewendeten Verfahren und zentrale Ergebnisse präsentiert. Kapitel 2 führt kurz in die Small-Area-Statistik ein, indem das spezifische Problem beschrieben wird, für das diese modernen statistischen Verfahren eine Lösung bieten sollen. Zudem wird der Lösungsansatz skizziert. Im Anschluss daran wird mit dem sogenannten Fay-Herriot-Modell das Small-Area-Modell, das in der Diplomarbeit angewendet wird, vorgestellt und erläutert. Kapitel 3 stellt die Anwendung auf regionale Mietpreise in Deutschland dar, indem zunächst die Daten und das spezifizierte Modell und anschließend die Ergebnisse präsentiert werden. Der Beitrag schließt mit einem Fazit, das die Verfahren für den vorliegenden Anwendungskontext abschließend bewertet.

## 2 Theorie: Small-Area-Statistik

### Problemstellung und Lösungsansatz

Wenn eine Erhebung durchgeführt wird, interessieren häufig nicht nur Informationen über die gesamte Population, sondern es besteht auch Interesse an verlässlichen Schätzungen für regional oder inhaltlich abgegrenzte Teilgesamtheiten. Eine Grundgesamtheit  $U$  mit  $N$  Elementen wird also nach räumlichen oder sachlichen Kriterien in  $m$  überschneidungsfreie Teilgesamtheiten  $U_i$  mit Umfang  $N_i$  zerlegt. Der Laufindex  $i = 1, \dots, m$  bezeichnet dabei die jeweils betrachtete Teilgesamtheit. Es wird nun eine Stichprobe  $S$  vom Umfang  $n$  gezogen. Der in Teilgesamtheit  $U_i$  realisierte Teil der Stichprobe heißt  $S_i$ , der zugehörige Stichprobenumfang ist  $n_i$ . Typische Beispiele für solche Teilgesamtheiten sind Gemeinden und Kreise oder auch nach bestimmten Merkmalen abgegrenzte demografische Gruppen. Ziel ist nun, teilgesamtheit-spezifische Parameter zu schätzen.<sup>4</sup> Das können beispielsweise regionsspezifische Mittelwerte für eine bestimmte Variable sein – etwa das Durchschnittseinkommen oder eben durchschnittliche Mietpreise. Diese zu schätzenden Mittelwerte werden im Folgenden als  $\mu_i$  bezeichnet.

Ein naheliegender Ansatz wäre es, aus allen Beobachtungen, die in einer Substichprobe  $S_i$  vorliegen, ein geeignet gewichtetes Mittel zu errechnen und dieses als Schätzwert für Region  $i$  zu verwenden. Da bei diesem Vorgehen nur Informationen herangezogen werden, die unmittelbar aus der betrachteten Region stammen, nennt man einen sol-

chen Schätzer auch *direkten Schätzer*. Es handelt sich um eine unverzerrte Schätzfunktion für den wahren Mittelwert in der Grundgesamtheit. Der so ermittelte Schätzwert wird im Folgenden mit  $\hat{\mu}_i^{Dir}$  bezeichnet. Es ergibt sich allerdings aufgrund der beschriebenen Zerlegung der Stichprobe  $S$  regelmäßig das Problem, dass die Stichprobenumfänge in den Teilgesamtheiten sehr klein sind. In diesem Fall führen direkte Schätzverfahren zu einer oft inakzeptabel hohen Varianz der Schätzung und die Schätzergebnisse sind damit wenig verlässlich. Üblicherweise werden kleinräumige Ergebnisse von Stichprobenerhebungen aus diesem Grund nicht veröffentlicht.

Die Small-Area-Statistik bietet eine Bandbreite spezieller Ansätze, um auch in solchen Fällen zuverlässige Schätzungen für Subpopulationen zu gewinnen. Dazu werden sogenannte *indirekte Verfahren* angewendet, die bei der Schätzung des Parameters für eine bestimmte Region zusätzliche Informationen aus anderen Quellen nutzen, beispielsweise aus anderen Regionen. Im vorliegenden Fall werden diese zusätzlichen Informationen über explizite Modelle in die Schätzung einbezogen, die eine über verschiedene Regionen konstante Beziehung zwischen Hilfsvariablen und interessierender Variable unterstellen.<sup>5</sup> Letztendlich wird so die effektive, das heißt die tatsächlich in die Schätzung einbezogene, Stichprobe vergrößert. In der Small-Area-Literatur wird diese Strategie als „borrowing strength“<sup>6</sup> bezeichnet. Ziel ist es, die Varianz der Schätzung deutlich zu minimieren. Allerdings hat das einen Preis: Ein Schätzer, der bestimmte Modellannahmen nutzt, liefert nur dann unverzerrte Ergebnisse, wenn die getroffenen Annahmen erfüllt sind. Damit besteht in der Small-Area-Statistik gewissermaßen ein Zielkonflikt zwischen Unverzerrtheit und Streuung der Schätzergebnisse, der optimal gelöst werden sollte. Inwieweit eine Verbesserung der Schätzung gelingt, hängt letztendlich von der Verfügbarkeit geeigneter Hilfsinformationen sowie von der Güte der verknüpfenden Modelle ab. Eine umfassende Darstellung small-area-statistischer Verfahren ist 2003 von Rao veröffentlicht worden.<sup>7</sup> Der erste deutsche Überblicksartikel ist im Jahr 2013 erschienen.<sup>8</sup>

Im vorliegenden Fall wird ein modellbasiertes Verfahren der Small-Area-Statistik angewendet. Small-Area-Modelle lassen sich grundsätzlich in zwei Klassen einteilen, wobei das Differenzierungskriterium die Struktur der Hilfsinformationen sowie der zur Verfügung stehenden Stichprobeninformationen über die Zielgröße ist. Liegen Mikrodaten vor, also Angaben auf Ebene der einzelnen Erhebungseinheiten, kommen Unit-Level-Modelle zum Einsatz. Das entsprechende Standardmodell wurde 1988 von Battese und anderen vorgeschlagen und wird daher auch als Battese-Harter-Fuller-Modell bezeichnet.<sup>9</sup> Liegen die Informationen – wie im vorliegenden Fall – ausschließlich aggregiert auf Ebene der *areas* vor, können Area-Level-Modelle genutzt werden. Das entsprechende Standardmodell ist das so-

5 Siehe Rao, J. N. K.: „Small Area Estimation“, Hoboken 2003.

6 Ghosh, M./Rao, J. N. K.: „Small Area Estimation: An Appraisal“ in *Statistical Science*, Band 9/1994, Ausgabe 1, Seite 55 ff., hier: Seite 56.

7 Siehe Rao, J. N. K. (Fußnote 5).

8 Siehe Münnich, R./Burgard, J. P./Vogt, M. (Fußnote 4).

9 Siehe Battese, G. E./Harter, R. M./Fuller, W. A.: „An error components model for prediction of county crop areas using survey and satellite data“ in *Journal of the American Statistical Association*, Band 83/1998, Ausgabe 401, Seite 28 ff.

3 In einer ähnlichen Anwendung für Portugal nutzen Pereira und Coelho Standardmodelle der Small-Area-Statistik, um regionale Immobilienpreise zu schätzen. Siehe Pereira, L. N./Coelho, P. S.: „Estimation of House Prices in Regions with Small Sample Sizes“ in *The Annals of Regional Science*, Band 50/2013, Ausgabe 2, Seite 603 ff.

4 Siehe Münnich, R./Burgard, J. P./Vogt, M.: „Small Area-Statistik: Methoden und Anwendungen“ in *ASTA Wirtschafts- und Sozialstatistisches Archiv*, Band 6/2013, Ausgabe 3/4, Seite 149 ff.

nannte Fay-Herriot-Modell. Es wird in der vorgestellten Diplomarbeit angewendet und daher im nächsten Abschnitt ausführlich dargestellt.

## Das Fay-Herriot-Modell

Das Fay-Herriot-Modell wurde bereits 1979 von Fay und Herriot im Kontext einer Schätzung von Median-Einkommen für kleine räumliche Einheiten in den Vereinigten Staaten eingeführt.<sup>10</sup> Es hat sich seitdem zum Standardmodell der Small-Area-Statistik entwickelt. Es ist gegeben durch

$$\begin{aligned}\hat{\mu}_i^{Dir} &= \mathbf{x}_i^T \boldsymbol{\beta} + v_i + e_i \quad \text{für } i = 1, \dots, m \\ v_i &\sim N(0, \sigma_v^2) \\ e_i &\sim N(0, \sigma_{e,i}^2)\end{aligned}$$

Das angegebene Modell setzt sich aus zwei Bestandteilen zusammen: Es enthält (1) die Regressionsgleichung  $\hat{\mu}_i^{Dir} = \mathbf{x}_i^T \boldsymbol{\beta} + e_i$ . Dabei bezeichnet  $\hat{\mu}_i^{Dir}$  den aus der Teilstichprobe  $S_i$  errechneten direkten Schätzwert für Region  $i$  und  $e_i$  ist der zugehörige Stichprobenfehler mit bekannter (Design-) Varianz  $\sigma_{e,i}^2$ . Mit diesem Teil des Modells wird die Annahme eines über alle Regionen konstanten Zusammenhangs zwischen den Hilfsinformationen  $\mathbf{x}_i$  und der interessierenden Variable getroffen. Die zugehörigen Regressionsparameter  $\boldsymbol{\beta}$  stellen damit sogenannte *fixe Effekte* dar. Außerdem enthält es (2) einen zusätzlichen *zufälligen Effekt*  $v_i$  mit Erwartungswert 0 und Modellvarianz  $\sigma_v^2$ . Dieser zufällige Effekt erfasst den Teil der Variation zwischen den Regionen, der sich durch die Unterschiede in den Hilfsvariablen, das heißt durch den fixen Teil des Modells, nicht erklären lässt. Es wird angenommen, dass  $e_i$  und  $v_i$  voneinander unabhängig sind.

Es handelt sich also um ein einfaches gemischtes lineares Regressionsmodell, das genutzt wird, um die interessierenden Parameter  $\hat{\mu}_i^{Dir}$  zu modellieren. Es wird anhand der unverzerrten, aber mit einer hohen Varianz verbundenen direkten Schätzwerte geschätzt. Bei der Schätzung der festen Modellparameter werden damit die Informationen aus allen Regionen, das heißt aus der gesamten Stichprobe, genutzt. Die Idee dabei ist, dass sich die zufälligen Stichprobenfehler der Schätzwerte bei dieser Betrachtung über alle Regionen ausgleichen, sodass sie die Erfassung des konstanten Zusammenhangs nicht „stören“. Anschließend kann dieser durch die Modellparameter quantifizierte Zusammenhang genutzt werden, um die Schätzung der interessierenden Parameter zu stabilisieren. Bei der Schätzung des Modells stehen nämlich nicht wie in einer klassischen ökonometrischen Anwendung die Modellparameter und ihre Interpretation im Fokus. Ziel ist vielmehr die Prädiktion (das heißt die Schätzung) der interessierenden regionsspezifischen Parameter aus dem Modell.

Unter den getroffenen Annahmen ist der beste<sup>11</sup> Prädiktor für diese Parameter (der sogenannte *Empirical Best Linear Unbiased Predictor*<sup>12</sup>) gegeben durch

$$\hat{\mu}_i^{FH} = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{v}_i.$$

Der Fay-Herriot-Schätzwert  $\hat{\mu}_i^{FH}$  setzt sich also aus einem aus dem festen Teil des Modells  $\mathbf{x}_i^T \hat{\boldsymbol{\beta}}$  ermittelten *synthetischen* Schätzwert und einem regionsspezifischen zufälligen Effekt zusammen. Dabei wird der zufällige Effekt optimal geschätzt durch

$$\hat{v}_i = \hat{\gamma}_i (\hat{\mu}_i^{Dir} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \quad \text{mit} \quad \hat{\gamma}_i = \frac{\hat{\sigma}_v^2}{\sigma_{e,i}^2 + \hat{\sigma}_v^2}.$$

Damit entspricht der zufällige Effekt der Differenz zwischen direktem und synthetischem Schätzwert multipliziert mit dem sogenannten *shrinkage*-Faktor  $\hat{\gamma}_i$ , das heißt dem Anteil der Modellvarianz an der Gesamtvarianz. Dabei kann  $\hat{\gamma}_i \in (0, 1)$  als Maß für das relative Vertrauen, das in den direkten Schätzwert und den aus dem Modell ermittelten synthetischen Schätzwert gesetzt wird, verstanden werden. Dieses Vertrauensmaß bestimmt das kompensatorische Gewicht, das den beiden Komponenten bei der Schätzung gegeben wird. Ist beispielsweise die Modellvarianz groß und die Designvarianz sehr klein, dann ist  $\hat{\gamma}_i$  nahe 1 und es erfolgt eine fast vollständige Korrektur des synthetischen Schätzwertes um die Differenz aus direktem und synthetischem Schätzwert. Das bedeutet letztendlich, dass  $\hat{\mu}_i^{FH}$  beinahe dem anscheinend verhältnismäßig „guten“ direkten Schätzwert entspricht. Ist allerdings die Modellvarianz klein und die Designvarianz groß, dann ist  $\hat{\gamma}_i$  nahe 0 und es erfolgt nur eine sehr kleine Korrektur des aus dem Modell ermittelten synthetischen Schätzwertes.

Setzt man  $\hat{v}_i = \hat{\gamma}_i (\hat{\mu}_i^{Dir} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})$  in  $\hat{\mu}_i^{FH} = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{v}_i$  ein, ergibt sich nach einfacher Umformung die Schreibweise

$$\hat{\mu}_i^{FH} = \hat{\gamma}_i \hat{\mu}_i^{Dir} + (1 - \hat{\gamma}_i) \mathbf{x}_i^T \hat{\boldsymbol{\beta}}.$$

Auch hier lässt sich die eben beschriebene Intuition in Bezug auf einen qualitätsabhängigen Kompromiss zwischen direktem und synthetischem Schätzwert erkennen: Der Fay-Herriot-Schätzer wird als mit dem Vertrauensmaß  $\hat{\gamma}_i$  gewichtetes Mittel zwischen diesen beiden Komponenten dargestellt.

In der vorgestellten Diplomarbeit wurde ein solches Fay-Herriot-Modell für die Schätzung regionaler Mietpreise genutzt. Im folgenden Kapitel werden die dabei verwendeten Daten beschrieben und das spezifizierte Modell vorgestellt.

<sup>10</sup> Siehe Fay, R. E./Herriot, R. A.: "Estimates of income for small places: An application of James-stein procedures to census data" in Journal of the American Statistical Association, Band 74/1979, Ausgabe 366, Seite 269 ff.

<sup>11</sup> Der beste Schätzer ist in diesem Fall derjenige, der den beschriebenen Zielkonflikt zwischen Varianz und Verzerrung optimal löst. Er ist also MSE-optimal.

<sup>12</sup> Siehe Rao, J. N. K. (Fußnote 5), hier: Seiten 107, 116.

### 3 Anwendung: Regionale Mietpreise in Deutschland

#### Datengrundlage und Modell

Die Datengrundlage bilden mittlere Bestandsmieten aus der Mikrozensus-Zusatzerhebung zur Wohnsituation. Diese wird alle vier Jahre durchgeführt. Die Ergebnisse werden üblicherweise nicht auf Kreisebene veröffentlicht, weil sie bei einer tiefen regionalen Gliederung die Präzisionsanforderungen nicht erfüllen. In der hier vorgestellten Arbeit wurde allerdings eine vom Statistischen Bundesamt durchgeführte Sonderauswertung der Erhebung von 2010 genutzt. Ausgewiesen waren mittlere Mietpreise für 246 Kreise in 13

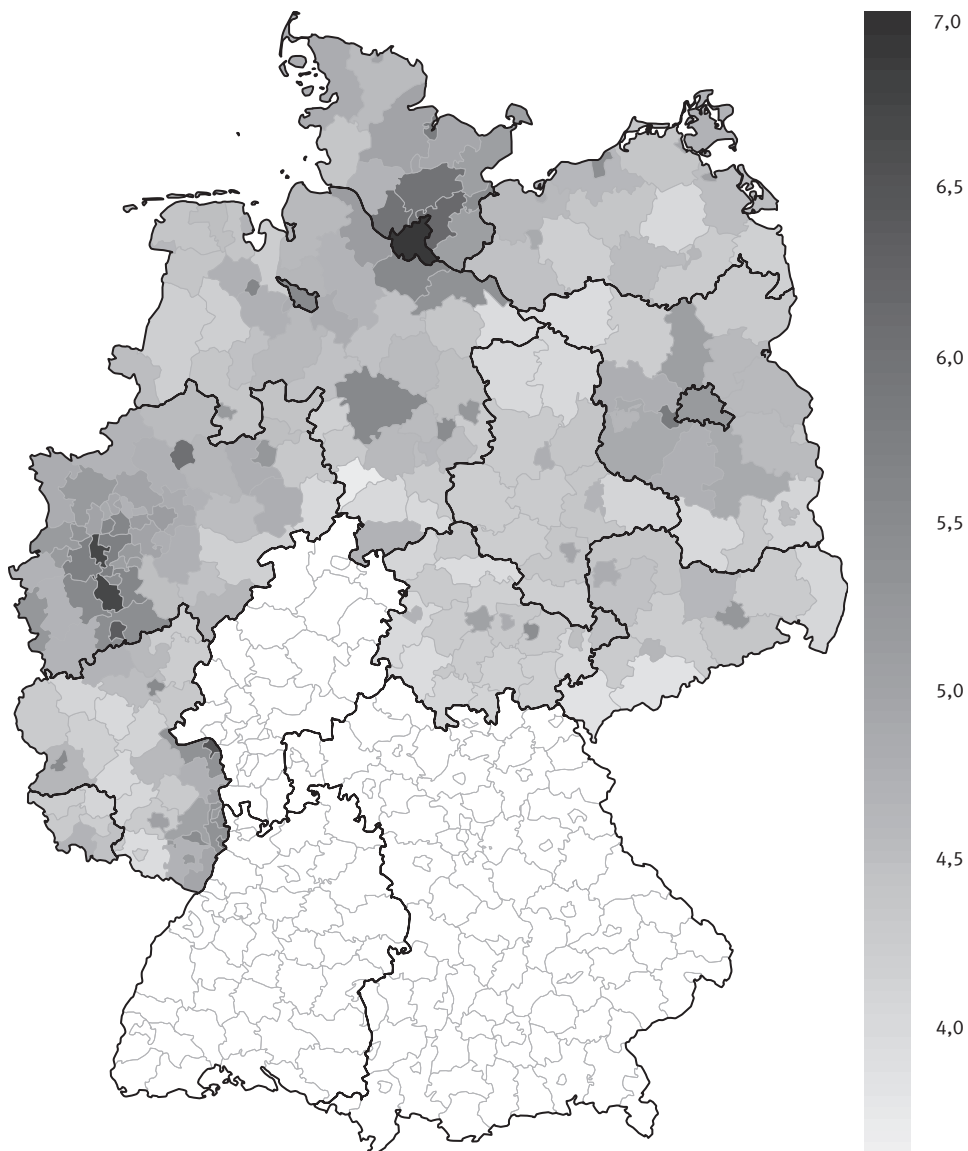
deutschen Bundesländern.<sup>13</sup> Es handelt sich hierbei um die oben eingeführten und mit  $\hat{\mu}_i^{Dir}$  bezeichneten direkten Schätzwerte. Zusätzlich angegeben waren die regionsspezifischen Stichprobenumfänge sowie die geschätzten Varianzen für die zur Verfügung gestellten Schätzwerte.

Hilfsinformationen können aus einem großen Angebot an von der amtlichen Statistik bereitgestellten Regionalindikatoren gewonnen werden.<sup>14</sup> Die Variablenselektion erfolgt über eine literaturbasierte Analyse wichtiger Bestimmungs-

<sup>13</sup> Für Hessen, Baden-Württemberg und Bayern standen keine Informationen zur Verfügung.

<sup>14</sup> In der Diplomarbeit wurde überwiegend mit der vom Bundesinstitut für Bau-, Stadt- und Raumforschung herausgegebenen Indikatorensammlung INKAR gearbeitet, die viele regionalstatistische Informationen aus unterschiedlichen gesellschaftlichen Bereichen enthält ([www.bbsr.bund.de/BBSR/DE/Veroeffentlichungen/INKAR/inkar\\_node.html](http://www.bbsr.bund.de/BBSR/DE/Veroeffentlichungen/INKAR/inkar_node.html), abgerufen am 14. Februar 2014).

**Schaubild 1** Mit dem Fay-Herriot-Modell geschätzte durchschnittliche regionale Mietpreise 2010  
EUR je m<sup>2</sup>



Datengrundlage: Mikrozensus-Zusatzerhebung zur Wohnsituation. Ohne Hessen, Baden-Württemberg und Bayern.

2014 - 01 - 0079

faktoren der Preisbildung auf Mietmärkten sowie über einfache sequenzielle Modellselektionsverfahren.<sup>15</sup>

Das auf Basis der Ergebnisse letztendlich ausgewählte Modell<sup>16</sup> enthält sowohl Indikatoren für die Anspannung auf lokalen Mietmärkten [Bevölkerungsentwicklung (PDEV),

Bedeutung des Mietmarktes (RENT) und Leerstandsquote (VACQ)] als auch für die Kaufkraft und Attraktivität einer Region [Beschäftigtenquote (EMPL), Gesamtwanderungssaldo (MIGR) und Baulandpreise (LAND)]. Damit ergibt sich für Region  $i, i = 1, \dots, m$

$$\hat{\mu}_i^{Dir} = \beta_0 + \beta_1 PDEV_i + \beta_2 RENT_i + \beta_3 VACQ_i + \beta_4 EMPL_i + \beta_5 MIGR_i + \beta_6 LAND_i + v_i + e_i.$$

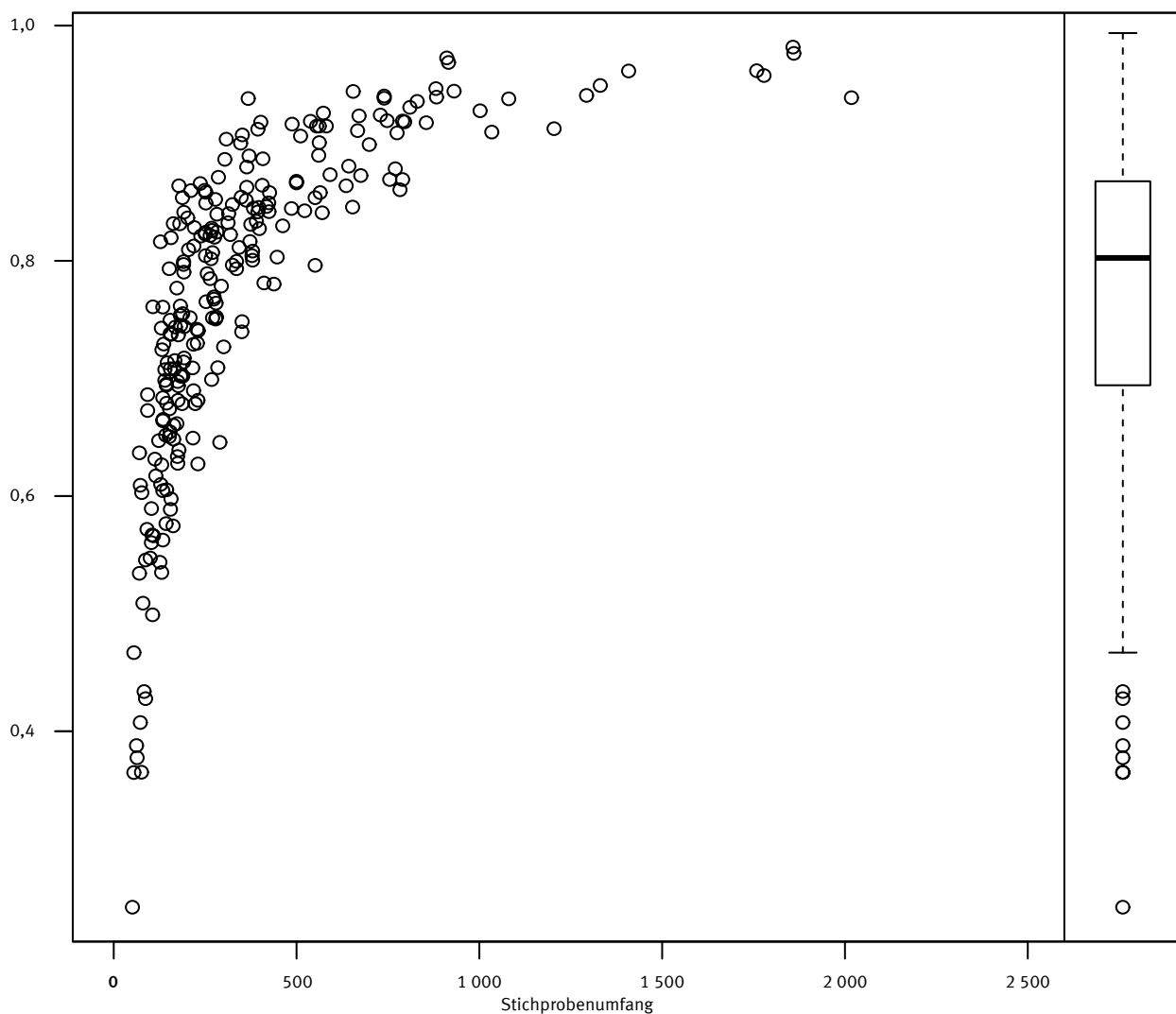
Bei den verwendeten Indikatoren handelt es sich um Daten aus der Bevölkerungsfortschreibung des Bundes und der Länder (PDEV), der Beschäftigtenstatistik der Bundesagentur für Arbeit (EMPL), der Wanderungsstatistik des Bundes und der Länder (MIGR) und der Statistik der Kaufpreise für Bauland (LAND). Die Indikatoren RENT und VACQ beruhen auf Ergebnissen des Zensus 2011.

15 Siehe Fox, J.: "Regression Diagnostics", Newbury Park 1991. Als Modellselektionskriterium wird das *conditional AIC* herangezogen. Es handelt sich dabei um ein von Vaida und Blanchard vorgeschlagenes relatives Maß der Anpassungsgüte für gemischte Modelle. Siehe Vaida, F./Blanchard, S.: "Conditional Akaike Information for Mixed-Effect Models" in Biometrika, Band 92/2005, Ausgabe 2, Seite 351 ff.

16 In der vorgestellten Diplomarbeit werden insgesamt sechs verschiedene Small-Area-Modelle spezifiziert und ausführlich miteinander verglichen, sodass auch die Konsequenzen unterschiedlicher Strategien der Auswahl und Nutzung verfügbarer Hilfsvariablen untersucht werden können. Im Folgenden wird lediglich das Modell vorgestellt, das schließlich für die Schätzung der Mietpreise ausgewählt wurde. Dabei ist anzumerken, dass die Variablen RENT und VACQ, die ursprünglich aus der Mikrozensus-Zusatzerhebung gewonnen wurden, inzwischen durch wesentlich verlässlichere Informationen aus dem Zensus 2011 ersetzt worden sind.

## Schaubild 2 Präzisionsgewinne durch die Anwendung der Small-Area-Verfahren

RMSE der Fay-Herriot-Schätzwerte  
Standardfehler der direkten Schätzwerte



Um die Lesbarkeit zu erhöhen, wurde die X-Achse auf den Wertebereich von 0 bis 2 500 beschränkt. Damit sind vier Regionen mit einem größeren Stichprobenumfang als 2 500 nicht zu sehen.

2014 - 01 - 0080

## Ergebnisse

Die Ergebnisse der Schätzung der fixen Modellparameter enthält Tabelle 1.

**Tabelle 1 Ergebnisse der Schätzung der fixen Modellparameter zur Schätzung regionaler Mietpreise**

	Regressionsparameter $\beta$	Standardfehler	t-Wert	p-Wert
Achsenabschnitt .	2.821	0.413	6.824	0.000
Bevölkerungsentwicklung . . . . .	0.033	0.018	1.909	0.056
Bedeutung des Mietmarktes . . .	0.014	0.002	5.601	0.000
Leerstandsquote .	-0.039	0.014	-2.717	0.007
Beschäftigtenquote . . . . .	0.024	0.007	3.230	0.001
Gesamtwanderungssaldo . . . .	0.021	0.010	2.071	0.038
Baulandpreise . . . .	0.004	0.000	9.041	0.000

Außerdem wird die Modellvarianz  $\sigma_v^2$  aus den Daten geschätzt ( $\sigma_v^2 = 0,085$ ). Anschließend werden die Fay-Herriot-Schätzwerte für die durchschnittlichen Mietpreise in Deutschland aus dem Modell prädiziert. In der vorgestellten Diplomarbeit werden die Ergebnisse mit verschiedenen diagnostischen Instrumenten auf Verletzung der Modellannahmen und Verzerrungen überprüft. Dabei ergibt sich ein Hinweis auf eine leichte Verzerrung der Schätzung für die Regionen mit den höchsten und niedrigsten Mietpreisen: Die Preise in Kreisen mit besonders teuren Mieten werden offensichtlich tendenziell unterschätzt, während es bei Kreisen mit günstigen Mietpreisen zu einer Überschätzung kommt. Dies betrifft etwa 10 der 246 betrachteten Kreise. Die Resultate für alle anderen Regionen sind aber als valide zu bezeichnen.

Schaubild 1 illustriert die Ergebnisse der Schätzung. Die durchschnittlichen Mietpreise bewegen sich zwischen 3,72 Euro je Quadratmeter und 6,95 Euro je Quadratmeter. Die besonders hochpreisigen Städte – Hamburg, Düsseldorf, Köln und Mainz – sind deutlich zu erkennen.

Ziel der Anwendung der Small-Area-Verfahren ist es, die Varianz der Schätzung regionaler Mietpreise deutlich zu reduzieren. Daher interessiert, ob ein solcher Präzisionsgewinn realisiert werden konnte, sodass die Nutzung der komplexeren statistischen Methoden in diesem Fall gerechtfertigt ist. Im Schaubild 2 wird das Verhältnis zwischen der Wurzel aus dem mittleren quadratischen Fehler [Root Mean Squared Error (RMSE)] der Fay-Herriot-Schätzwerte und dem Standardfehler der direkten Schätzwerte  $\sigma_{e,i}$  gegen die regionsspezifischen Stichprobenumfänge abgetragen. Es ist für die meisten Regionen ein deutlicher Präzisionsgewinn zu erkennen. Wie zu erwarten, ist diese Verbesserung der Schätzung für Regionen mit sehr kleinem Stichprobenumfang besonders groß, während sie für Kreise mit einem Stichprobenumfang über etwa 800 Beobachtungen eher gering ist. Diese Regionen können aber ohnehin kaum noch als *small areas* bezeichnet werden.<sup>17</sup>

<sup>17</sup> Um die Lesbarkeit des Schaubildes 2 zu erhöhen, wurde die X-Achse auf den Wertebereich von 0 bis 2 500 beschränkt. Damit sind vier Regionen mit einem größeren Stichprobenumfang als 2 500 nicht zu sehen.

## 4 Fazit

Im Ergebnis zeigen die Analysen der vorgestellten Diplomarbeit, dass das Anwendungsgebiet der Schätzung regionaler Mietpreise gute Voraussetzungen für einen erfolgreichen Einsatz von Small-Area-Verfahren bietet. Mit den mittleren Bestandsmieten aus der Mikrozensus-Zusatzerhebung zur Wohnsituation liegen recht stabile direkte Schätzwerte vor, die für viele Kreise auf (im Kontext der Small-Area-Statistik) verhältnismäßig großen Stichproben basieren. Zudem sind zahlreiche qualitativ hochwertige Hilfsvariablen verfügbar. Die Anwendung der modellbasierten Small-Area-Verfahren liefert daher insgesamt gute Resultate. Es kann ein teilweise erheblicher Präzisionsgewinn gegenüber den verfügbaren direkten Schätzwerten erzielt werden. Diagnostische Instrumente zeigen, dass die Modelle für einen Großteil der Kreise unverzerrte Schätzwerte liefern.

Auch wenn die Ergebnisse bei einer Betrachtung über alle Kreise hinweg als gut zu beurteilen sind, ergeben sich für bestimmte Kreise Probleme: Die Mieten in besonders teuren Regionen werden tendenziell unterschätzt, während Preise in günstigen Regionen überschätzt werden. Hier sind weitere Arbeiten geplant, um ein Modell zu finden, das auch die Regionen mit besonders hohen und besonders niedrigen Mietpreisen angemessen abbilden kann. [UU](#)

## Auszug aus Wirtschaft und Statistik

### Herausgeber

Statistisches Bundesamt, Wiesbaden

[www.destatis.de](http://www.destatis.de)

### Schriftleitung

Dieter Sarreither,  
Vizepräsident des Statistischen Bundesamtes

Redaktion: Ellen Römer  
Telefon: + 49 (0) 6 11 / 75 23 41

### Ihr Kontakt zu uns

[www.destatis.de/kontakt](http://www.destatis.de/kontakt)

### Statistischer Informationsservice

Telefon: + 49 (0) 6 11 / 75 24 05

### Abkürzungen

WiSta	=	Wirtschaft und Statistik
MD	=	Monatsdurchschnitt
VjD	=	Vierteljahresdurchschnitt
HjD	=	Halbjahresdurchschnitt
JD	=	Jahresdurchschnitt
D	=	Durchschnitt (bei nicht addierfähigen Größen)
Vj	=	Vierteljahr
Hj	=	Halbjahr
a. n. g.	=	anderweitig nicht genannt
o. a. S.	=	ohne ausgeprägten Schwerpunkt
St	=	Stück
Mill.	=	Million
Mrd.	=	Milliarde

### Zeichenerklärung

p	=	vorläufige Zahl
r	=	berichtigte Zahl
s	=	geschätzte Zahl
–	=	nichts vorhanden
0	=	weniger als die Hälfte von 1 in der letzten besetzten Stelle, jedoch mehr als nichts
.	=	Zahlenwert unbekannt oder geheim zu halten
...	=	Angabe fällt später an
X	=	Tabellenfach gesperrt, weil Aussage nicht sinnvoll
oder —	=	grundsätzliche Änderung innerhalb einer Reihe, die den zeitlichen Vergleich beeinträchtigt
/	=	keine Angaben, da Zahlenwert nicht sicher genug
()	=	Aussagewert eingeschränkt, da der Zahlenwert statistisch relativ unsicher ist

Abweichungen in den Summen ergeben sich durch Runden der Zahlen.