



## Bodenwertermittlung mit statistischen Methoden

Jens Kolbe · Rainer Schulz · Martin Wersing · Axel Werwatz

Online publiziert: 18. Oktober 2019  
© Der/die Autor(en) 2019

**Zusammenfassung** Die Besteuerung von Immobilien auf der Grundlage von Marktwerten erfordert eine regelmäßige Bewertung einer Vielzahl von Immobilien. Vor dem Hintergrund der jüngsten Diskussion über die Reform der Grundsteuer in Deutschland wird argumentiert, dass eine wertorientierte Steuer nicht mit vertretbarem Aufwand umgesetzt werden kann. In mehreren anderen Ländern werden Massenbewertungssysteme auf Basis von statistischen Methoden für die Grundsteueranlagung eingesetzt. In diesem Beitrag zeigen wir, wie dies in Deutschland grundsätzlich geschehen könnte; mittels Transaktionsdaten zu deren Erhebung die lokalen Sachverständigenkommissionen gesetzlich verpflichtet sind. Wir diskutieren die Regressionsverfahren zur Schätzung von Bodenwerten aus solchen Daten und veranschaulichen sie, indem wir sie auf Daten aus Berlin anwenden. Wir kommen zu dem Ergebnis, dass die Methoden in der Lage sind, Bodenwertschätzungen zu erstellen, die gut mit fachspezifischen Einschätzungen übereinstimmen.

---

J. Kolbe · A. Werwatz

Institut für Volkswirtschaftslehre und Wirtschaftsrecht, Technische Universität Berlin,  
Straße des 17. Juni 135, 10623 Berlin, Deutschland

J. Kolbe

E-Mail: [j.kolbe@tu-berlin.de](mailto:j.kolbe@tu-berlin.de)

A. Werwatz

E-Mail: [axel.werwatz@tu-berlin.de](mailto:axel.werwatz@tu-berlin.de)

DFG Forschergruppe FOR2569 „Agricultural Land Markets – Efficiency and Regulation“,  
Humboldt-Universität zu Berlin, Berlin, Deutschland

R. Schulz · M. Wersing (✉)

University of Aberdeen Business School, Edward Wright Building,  
Dunbar Street, Aberdeen, AB24 3QY, Großbritannien

E-Mail: [martin.wersing@abdn.ac.uk](mailto:martin.wersing@abdn.ac.uk)

R. Schulz

E-Mail: [r.schulz@abdn.ac.uk](mailto:r.schulz@abdn.ac.uk)

**Schlüsselwörter** Bodenrichtwerte · Automatisierte Bewertungsverfahren · Nichtparametrische Regression · Semi-parametrische Regression

**JEL Classification** C14 · C21 · H10 · H20 · R32 · R51 · R52

## Land value appraisal using statistical methods

**Abstract** The taxation of property based on market values requires frequent appraisals for a large number of properties. In light of the recent property tax reform discussion in Germany, it has been argued that a value-based tax therefore cannot be implemented at a reasonable cost. In several other countries, however, mass appraisal systems based on statistical methods are used for property tax assessments. In this paper, we show how this could in principle be done in Germany, using transactions data that local surveyor commissions are obliged to collect by law. We discuss the regression techniques for estimating land values from such data and illustrate them by applying them to data from Berlin, Germany. We find that the methods are capable of producing land value estimates that match up well with expert based assessments.

**Keywords** Land value · Mass appraisal · Nonparametric regression · Semiparametric regression

## 1 Einführung

Marktwerte von unbebauten Grundstücken sind wichtig für Immobilienmarktteilnehmer aus dem privaten und dem öffentlichen Sektor. Privatwirtschaftliche Unternehmen benötigen solche Werte, um zu entscheiden, ob ein Entwicklungsprojekt wirtschaftlich realisiert werden kann. Akteure aus dem öffentlichen Sektor benötigen solche Werte, um die Kosten alternativer Verwendung, etwa dem Verkauf auf dem Markt oder der Nutzung für ein soziales Wohnungsbauprogramm, einschätzen zu können. Der Bau einer Ringstraße oder der Bau oder Ausbau eines Flughafens kann voraussetzen, dass Eigentümer bestimmter Grundstücke enteignet werden, was wiederum Marktwerte für die Berechnung der Entschädigungssumme nötig macht.

In den genannten Fällen werden Bodenwerte nur für eine relativ kleine Anzahl von Grundstücken benötigt. Dies ist bei einer weiteren Anwendung des öffentlichen Sektors anders: der Besteuerung von Immobilien auf der Grundlage von Marktwerten. Für die Besteuerung unbebauter Grundstücke erfordert dies, dass Bodenwerte vorhanden sind. Diese sind aber auch für bebaute Grundstücke notwendig, sobald diese auf Basis von Ertrags- und Sachwerten besteuert werden sollen. Der Bodenwert wird auch dann benötigt, wenn der Boden und das Gebäude mit unterschiedlichen Steuersätzen besteuert werden soll. Eine reine Grundsteuer, die das Gebäude unversteuert lässt, ist ein extremer Fall von unterschiedlichen Steuersätzen. Graduelle Steuersätze haben theoretisch attraktive Eigenschaften (Brueckner 1986; Oates und

Schwab 2009), wurden bisher aber nur in wenigen Regionen der Welt angewendet (Bourassa 2009; Franzsen 2009).<sup>1</sup>

In Deutschland ist die Grundsteuer zu einem aktuellen Thema geworden, nachdem das Bundesverfassungsgericht (BVerfG) im April 2018 entschied, dass die bisherige Besteuerung auf der Grundlage von Einheitswerten verfassungswidrig ist und dass bis Ende 2019 eine Lösung für dieses Problem gefunden werden muss. Das reformierte Grundsteuerverfahren muss ab 2025 umgesetzt werden. Die Einheitswerte, die die Bemessungsgrundlage für die Grundsteuer bilden, werden mit dem Ertrags- oder Sachwertverfahren ermittelt. Allerdings stützt sich die Ermittlung auf veraltete Informationen (ab 1964 im Westen und ab 1935 in den Beitrittsländern). Obwohl die Informationen für Bewertungen alle sechs Jahre aktualisiert werden sollten, geschah dies nicht (Fuest et al. 2018; Kreuziger et al. 2013). Dementsprechend spiegeln die Einheitswerte nicht die aktuellen relativen Marktwerte wider (Beirat BMF 2010, S. 1), was zum Urteil des BVerfG führte.

Das Urteil gibt dem Gesetzgeber Spielraum für das Verfahren zur Ermittlung der Bemessungsgrundlage der Grundsteuer, dies allerdings unter der Bedingung, dass die *relativen* Wertverhältnisse von Immobilien reflektiert werden. Von verschiedenen Seiten wurden Vorschläge zur Ermittlung der Bemessungsgrundlage vorgebracht. Diese reichen von einer reinen Bodenwertsteuer (auch für bebaute Grundstücke), über ein automatisiertes Vergleichsverfahren für bebaute Grundstücke, über ein Verfahren, welches eine wertorientierte Bodenkomponente mit einer auf pauschalisierten Herstellungskosten beruhenden Gebäudekomponente kombiniert, bis zu einem wertfreien Verfahren, welches die physikalischen Dimensionen einer Immobilie zu einem numerischen Indikator verdichtet (Beirat BMF 2010, S. 5 ff; Fuest et al. 2018, S. 12).<sup>2</sup> Ein wiederkehrendes Thema in der Diskussion der verschiedenen Vorschläge ist die Frage, ob eine auf dem Marktwert beruhende Bemessungsgrundlage angesichts vermeintlich hoher Umsetzungskosten überhaupt machbar ist. Mehrere Autoren führen die Veranlagungskosten als Argument gegen eine wertorientierte Grundsteuer an. Fuest et al. (2018, S. 9) nehmen beispielsweise an, dass eine „hinreichend präzise“ Bewertung 500 € pro Immobilie kosten würde, die alle fünf Jahre anfallen würde. Homburg (2018, S. 175) ist weniger zurückhaltend und erklärt, dass Marktwertermittlung Milliarden kosten und doch nur „scheingenaue Bewertungen“ liefern würde. Hey (2017, S. 35) ist etwas optimistischer in Bezug auf die Umsetzung eines wertorientierten Ansatzes, obwohl die erforderlichen Daten nicht in jeder Gemeinde nach den gleichen Standards und mit der gleichen Transparenz erhoben und analysiert werden.

In der Diskussion spielen computergestützte Massenbewertungsverfahren kaum eine Rolle. Dies ist bemerkenswert, da in vielen Ländern solche Verfahren erfolgreich für die Grundsteueranlagung eingesetzt werden. In dieser Arbeit zeigen wir anhand einer Anwendung auf den Berliner Grundstücksmarkt, wie ein solches Massenbewertungsverfahren funktioniert und wie genau die Bewertungen sind. Die

<sup>1</sup> Die Wertzuwächse der Immobilieneigentümer in Städten wie London haben kürzlich das Interesse an Bodensteuern wiederbelebt, vgl. o.V. (2018a, 2018b).

<sup>2</sup> McCluskey und Franzsen (2013) geben einen Überblick über wertfreie Ansätze, die in anderen Ländern verwendet werden.

Ergebnisse basieren auf Kolbe et al. (2015). Wir sind uns bewusst, dass Berlin über ein sehr effektives System zur Erfassung und Speicherung von Immobilientransaktionsdaten verfügt und dass eine solche Infrastruktur nicht in allen Teilen Deutschlands vorhanden ist. Bis zur vom BVerfG festgelegten Frist kann daher eine flächendeckende Implementierung computergestützter Massenbewertung nicht umgesetzt werden. Allerdings stehen wir den pauschalen Einschätzungen kritisch gegenüber, die behaupten, dass Massenbewertungen – selbst wenn technisch möglich – nicht sinnvoll sind, da teuer und ungenau.

Der Rest des Papiers ist wie folgt gegliedert. In Abschn. 2 wird erläutert, wie die Immobilienbewertung in anderen Ländern durchgeführt wird, wobei wir uns auf diejenigen Länder konzentrieren, die computergestützte Massengutachten einsetzen. Abschn. 3 stellt Methoden zur statistischen Massenbewertung für zwei Datenszenarien vor: (i) Daten aus Transaktionen von unbebauten Grundstücken und (ii) Daten aus Transaktionen von erschlossenen Grundstücken (d.h. Einfamilienhäusern und Eigentumswohnungen). Wir veranschaulichen die Methoden, indem wir sie auf die Transaktionsdaten in Berlin anwenden. Während eine Diskussion über statistische Methoden notwendigerweise etwas technisch ist, legen wir den Schwerpunkt auf Beschreibung der zugrundeliegenden Mechanismen der Bodenwertschätzer. Abschn. 4 schließt mit einem Fazit ab.

## 2 Internationaler Kontext

In den Vereinigten Staaten, wo die Besteuerung von Immobilien auf der Grundlage von Zeitwerten erfolgt, wird unterschieden zwischen Honorargutachtern (*fee appraisers*), welche den Marktwert bestimmter Immobilien im Geschäftsverkehr individuell bewerten, und Schätzern (*assessors*), die Massenschätzungen für Hunderttausende von Immobilien vornehmen. Seit etwa 1968 verwenden Schätzer/Assessoren statistische Methoden wie die Regressionsanalyse für ihre Massenbewertungen (Almy und Ferguson 2010; Back 1970). Heute haben mindestens 15 Länder statistische Massenbewertungssysteme für den Einsatz in der Vermögensbesteuerung eingeführt (Almy 2014; Almy und Ferguson 2010; Bidanset 2014). Dazu gehören Australien, Kanada, Ägypten, Mauritius, Neuseeland, Südafrika und die Vereinigten Staaten. Beispiele aus Europa sind Dänemark, Finnland, Lettland, Litauen, Russland, Schweden, Nordirland im Vereinigten Königreich und die Niederlande.

Basierend auf den Erfahrungen in Kanada, den Niederlanden und den Vereinigten Staaten schätzt Almy (2014), dass die Kosten pro Objekt für ein qualitativ hochwertiges statistisches Massenbewertungssystem bei etwa 20€ pro Bewertung liegen. Dies steht im krassen Gegensatz zu den von Fuest et al. oben genannten 500€ und würde eine wesentlich kostengünstigere Implementierung einer wertorientierten Grundsteuer ermöglichen. Bei durchschnittlichen Steuereinnahmen von 2000€ pro Immobilie würden die Bewertungskosten nur 1% betragen.

Die internationale Erfahrung zeigt auch, dass statistische Massenschätzungen für steuerliche Zwecke hinreichend genaue Schätzungen der Immobilienwerte liefern können. Hefferan und Boyd (2010, S. 155) untersuchen zum Beispiel die Situation in Australien und kommen zu dem Schluss, dass „[...] immer ausgefeiltere com-

putergestützte Bewertungstechniken in den letzten Jahren tatsächlich dazu beigetragen haben, die Einheitlichkeit und Konsistenz [der Bewertungen] zu unterstützen“. Hefferan und Boyd (2010, S. 155) behaupten: „Die Einspruchsraten in Australien liegen deutlich unter den international akzeptierten 2%, wobei viele Jurisdiktionen auf weniger als 1% Einwände stoßen“. Im Rahmen unserer Arbeit ist Australien ein besonders interessantes Beispiel, da es eines der wenigen Länder der Welt ist, das eine Bodenwertsteuer hat und statistische Massenschätzungen durchführt.<sup>3</sup>

In Deutschland wurde vorgeschlagen, dass für die Berechnung der Bodenwertkomponente einer reformierten Grundsteuer Bodenrichtwerte (BRW) herangezogen werden können (Beirat BMF 2010, S. 2). Tatsächlich müssen die BRW bereits mindestens alle zwei Jahre von unabhängigen Gutachterkommissionen (Gutachterausschuss für Grundstückswerte, GAA) veröffentlicht werden. Trotz detaillierter Richtlinien kann man jedoch mit Fug und Recht sagen, dass die Qualität der BRW in hohem Maße vom Wissen und der Expertise der Gutachter abhängig ist. Angesichts der föderalen Struktur Deutschlands wurde daher argumentiert, dass die Schätzung der BRW harmonisiert werden müsse, um einen einheitlichen Qualitätsstandard für Steuerzwecke zu gewährleisten (Löhr 2011).

Im Folgenden wird diskutiert, wie statistische Methoden die Massenbewertung von Bodenwerten transparent unterstützen können.<sup>4</sup>

### 3 Statistische Methoden und Ergebnisse

Wir beginnen unsere Diskussion über die Methodik mit zwei konkreten Beispielen, die veranschaulichen, was wir aus den Daten lernen wollen und welche Mindestvoraussetzungen die Daten erfüllen müssen.

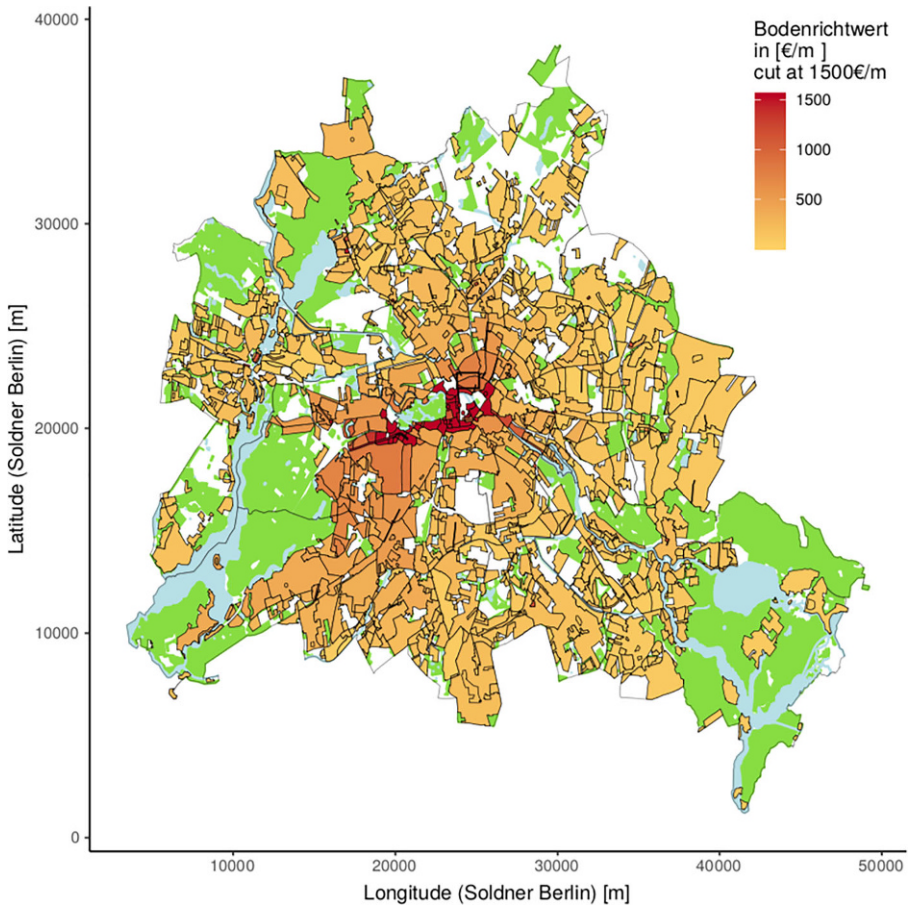
Abb. 1 zeigt die Bodenwertkarte für Berlin auf Basis der vom Berliner Gutachterausschuss veröffentlichten Bodenrichtwerte. Die Bodenwerte gelten für das Jahr 2009, da sie auf Informationen bis einschließlich dieses Jahres basieren. Flächenwerte innerhalb des ersten Perzentils der Werteverteilung werden in hellgelb dargestellt. Die Intensität steigt auf hellrot für Bodenwerte, die innerhalb des zehnten Perzentils der Verteilung liegen. Die Farbintensität nimmt, wie zu erwarten, vom Stadtzentrum aus gesehen kontinuierlich ab.<sup>5</sup>

Der zentrale Aspekt der Karte ist, dass sie eine Schätzung des zu erwartenden Grundstückspreises an jedem beliebigen Standort in Berlin liefert. Wir werden eine solche Schätzung als  $\hat{\theta}_{BRW}(l_1, l_2)$  bezeichnen, wobei  $l_1$  und  $l_2$  die Breite und Länge der Position sind. Das „Dach“ auf  $\theta$ , dem Symbol für den zu erwartenden Grundstückspreises, ist die übliche Weise in der Statistik, um eine aus Stichprobendaten berechnete Schätzgröße zu markieren. Das Subskript „BRW“ weist auf die Datenquelle hin.

<sup>3</sup> Litauen ist ein europäisches Beispiel für ein solches Land (Baranska 2013).

<sup>4</sup> McCluskey et al. (2013) geben einen Überblick darüber, wie statistische Massenbewertungssysteme funktionieren.

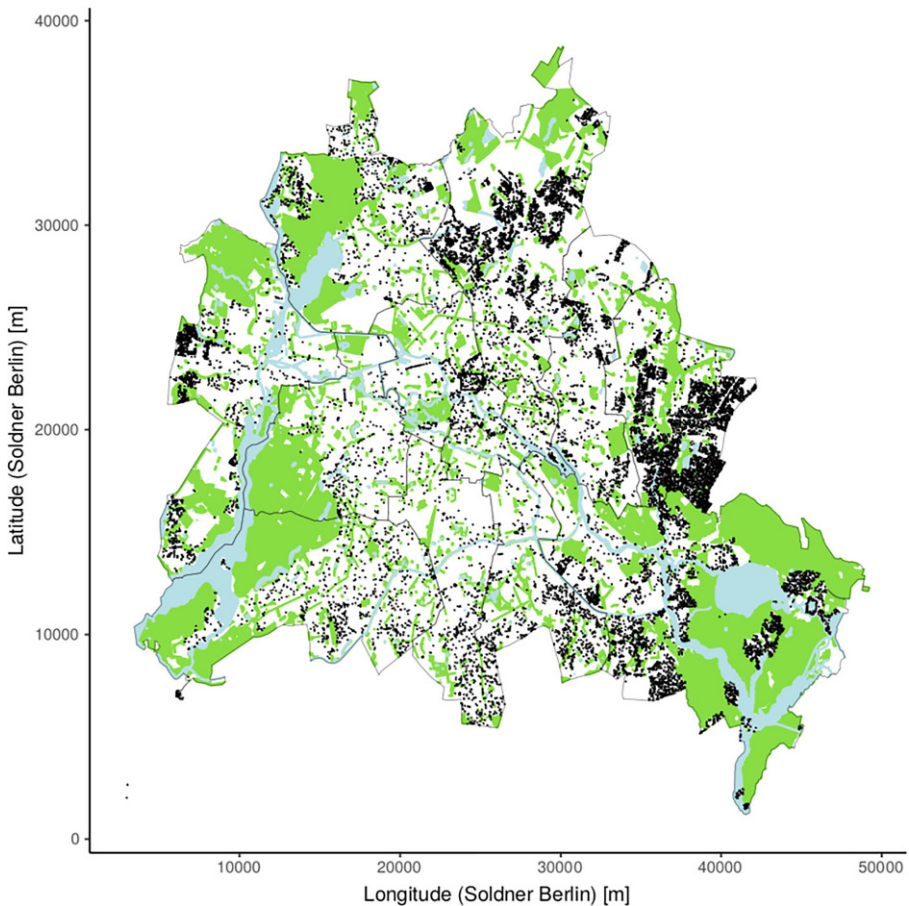
<sup>5</sup> Auf der Gesamtebene bestätigt die Bodenwertkarte somit die Vorhersage des monozentrischen Stadtmodells, dass der Pachtgradient mit zunehmender Entfernung vom Stadtzentrum sinkt.



**Abb. 1** Expertenbasierte Landwertkarte (BRW) für Berlin. Zeigt eine Karte von Bodenrichtwerten (BRW logarithmiert) für Berlin. Stichtag ist der 1. Januar 2010. Quelle: Geoportal Berlin/Bodenrichtwerte 01.01.2010

Die Abb. 2 zeigt die Standorte von 24.519 marktüblichen Transaktionen von unbebauten Grundstücken, die in den Jahren 1996–2009 in Berlin stattfanden. Die Transaktionsdaten stammen ebenfalls vom Berliner GAA. Die meisten Transaktionen von unbebauten Grundstücken fanden in den Wohngebieten am Rande Berlins statt. In der dicht bebauten Stadt fanden dagegen wenige Transaktionen von unbebauten Grundstücken statt.

Die in Abb. 1 dargestellten Expertenwerte basieren auf den in Abb. 2 visualisierten Daten, aber die GAA-Gutachter werden auch andere Immobilienmarktinformationen berücksichtigt haben bei der Berechnung ihrer Bodenrichtwerte. Im folgenden Abschnitt werden wir untersuchen, wie man eine Landwertkarte (wie in Abb. 1) aus Daten, wie den in Abb. 2 gezeigten, *allein* schätzen kann. Das heißt, wir werden das Problem der Schätzung des erwarteten Bodenwertes an einem bestimmten Standort betrachten, wenn nur geokodierte Daten von Transaktionen von unbebautem Land



**Abb. 2** Kauffälle innerhalb Berlins. Zeigt den Standort von 24.519 unbebauten Grundstücken an, die zwischen 1996 und 2009 verkauft wurden. Durchgezogene Linien kennzeichnen die Grenzen der 12 Berliner Bezirke (Stand 2010)

zur Verfügung stehen. Der Verkauf von unbebauten Grundstücken enthält offensichtlich unmittelbare Informationen über deren Marktbewertung. Wie aus Abb. 2 hervorgeht sind Transaktionen unbebauter Grundstücke in der Regel jedoch nicht für das Stadtzentrum verfügbar. In diesem, dem wertvollsten Teil einer Stadt finden sich oft nur Verkäufe von bebauten Grundstücken. Die beobachteten Marktpreise solcher Immobilien enthalten Informationen über den Wert des zugrundeliegenden Grundstücks *und* der darauf errichteten Struktur. Die statistische Analyse zur Gewinnung von Bodenwerten aus Transaktionen von bebauten Grundstücken muss in diesem Fall daher einen Weg finden, um die Komponenten Grundstücks- und Gebäudewert zu trennen. Wie dies geschehen kann, ist das Thema von Abschn. 3.3.

### 3.1 Statistische Bodenwertschätzung mittels nichtparametrischen Kern-Regressionen

#### 3.1.1 Erforderliche Daten

Im folgenden Abschnitt gehen wir davon aus, dass die Transaktions-Daten nur Informationen über (a) den Preis und (b) den Ort des Verkaufs eines unbebauten Grundstücks enthalten. Kurz gesagt wird angenommen, dass die Daten aus insgesamt  $n$  Beobachtungen der folgenden Form bestehen:

$$\{p_i, l_{1i}, l_{2i}\}, i = 1, \dots, n$$

wobei  $p_i = \ln(P_i)$  der logarithmierte Preis pro Quadratmeter des Grundstücks und  $l_{1i}$  und  $l_{2i}$  die exakten Geokoordinaten bestehend aus Breite und Länge sind. Warum man typischerweise mit logarithmierten Preisen („Log-Preisen“) und nicht mit den Preisen selbst arbeitet, wird im Folgenden erläutert.

#### 3.1.2 Das Regressionsmodell

Aus statistischer Sicht wird eine Bodenwertkarte wie die in Abb. 1 als Schätzung eines Regressionsmodells der folgenden Form betrachtet:

$$p_i = \theta(l_{1i}, l_{2i}) + \epsilon_i, \quad (1)$$

wobei  $\theta(l_{1i}, l_{2i})$  der erwartete (logarithmische) Grundstückswert am Standort  $l_{1i}, l_{2i}$  ist. Aus anwendungsorientierter Sicht ist  $\theta(l_{1i}, l_{2i})$  das Ziel der Bodenwertermittlung an einem Standort. Der Fehler  $\epsilon_i$  erfasst Abweichungen des Log-Preises eines bestimmten Grundstückes (Grundstück  $i$ ) von seinem erwarteten Wert  $\theta(l_{1i}, l_{2i})$ . Es wird erwartet, dass sich die einzelnen Fehler über alle  $n$  Beobachtungen ausgleichen (zu Null addieren) und dass an allen Standorten die gleiche Varianz herrscht. Letztere Annahme ist für die regressionsbasierte Schätzung von Bodenwerten nicht zwingend erforderlich, wird aber typischerweise gemacht, wenn auch eine Einschätzung der Schätzgenauigkeit mittels oberer und unterer Schranke um die Bodenwertschätzung herum von Interesse ist. Die Verteilung der Log-Preise erfüllt diese Annahme der „Homoskedastizität“ tendenziell besser als die Verteilung der Preise selbst, was ein Grund dafür ist, dass Forscher in der Regel Log-Preise als abhängige Variable ihres Regressionsmodells verwenden.

#### 3.1.3 Die Schätzmethode

Durch die Variation der Standorte  $l_1, l_2$  entsteht eine Fläche mit erwarteten Bodenwerten an verschiedenen Standorten, die grafisch durch eine farbige Karte wie diejenige in Abb. 1 dargestellt werden kann. Es ist die „Regressionsoberfläche“ in statistischer Hinsicht. Ein Standardverfahren zum Schätzen einer solchen Oberfläche an einem beliebigen gewünschten Standort aus Daten, wie den in Abb. 2



dargestellten, ist die Kernregression. Dabei wird lediglich angenommen, dass die Regressionsoberfläche keine Sprünge aufweist. Im Hinblick auf den gegenwärtigen Kontext heißt dies, dass sich die Bodenwerte im Raum kontinuierlich und nicht abrupt ändern. Abgesehen von dieser Anforderung an ihre Glattheit ist die Form der geschätzten Bodenwertkarte von vornherein nicht eingeschränkt. Insbesondere wird den Daten in diesem „nichtparametrischen“ Verfahren keine spezifische Formel auferlegt. So kann sich die Schätzung „frei“ an die Informationen in den Daten darüber anpassen, wie sich die beobachteten Preise von Standort zu Standort unterscheiden. Dies geschieht durch die Bildung lokaler Durchschnittswerte der Log-Preise.

Konkret ist der Kernregressionsschätzer nach Nadaraya (1964) und Watson (1964) definiert an einem beliebigen Standort  $l_1, l_2$  als

$$\hat{\theta}_{\text{NKR}}(l_1, l_2) = \sum_{i=1}^n \frac{K\left(\frac{l_1-l_{i1}}{h_1}, \frac{l_2-l_{i2}}{h_2}\right)}{\sum_{j=1}^n K\left(\frac{l_1-l_{j1}}{h_1}, \frac{l_2-l_{j2}}{h_2}\right)} \cdot p_i \tag{2}$$

oder, kürzer, als

$$\hat{\theta}_{\text{NKR}}(l_1, l_2) = \sum_{i=1}^n W_{i,h_1,h_2}(l_1, l_2) \cdot p_i \tag{3}$$

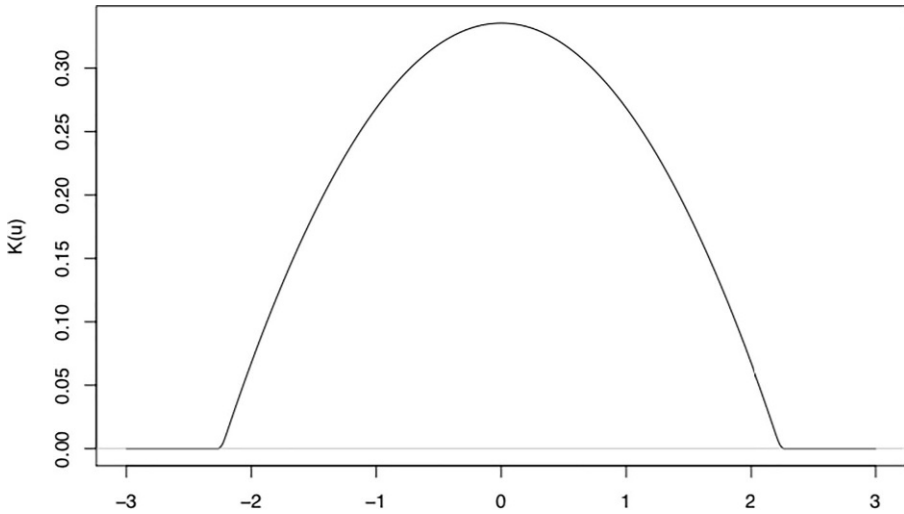
Die zweite Version in Gl. 3 hebt hervor, dass der Kernregressionsschätzer an der Stelle  $l_1, l_2$  tatsächlich als gewichteter (lokaler) Durchschnitt der logarithmierten Preises  $p_i$  mit den Gewichten  $\{W_{i,h_1,h_2}\}_{i=1,\dots,n}$  gesehen werden kann. Die erste Version in (2) zeigt genauer, wie diese Gewichte mit Hilfe einer Kernfunktionen  $K(\bullet, \bullet)$  gebildet werden, die der Methode ihren Namen gibt. Auf diesen zentralen Aspekt der Methode gehen wir detailliert im nächsten Absatz ein.

### 3.1.4 Wie Transaktionen gewichtet werden

Aufgabe der Kernfunktion ist die Gewichtung einer Beobachtung  $i$  bei der Schätzung an einem gewünschten Standort  $l_1, l_2$  an die Entfernung der Beobachtung vom Standort zu koppeln. Dabei soll gelten: je näher die Beobachtung dem Standort, an dem eine Schätzung erfolgen soll, desto höher ihr Gewicht bei der Schätzung. Da der Standort durch die beiden Geokoordinaten  $l_1$  und  $l_2$  beschrieben ist, hat die Kernfunktion auch zwei Argumente: die Nähe der Beobachtung  $i$  zum Standort muss sowohl bezüglich der Breite als auch der Länge der Position bestimmt werden. Dazu existieren verschiedene Vorschläge. Eine beliebte Wahl ist die Spezifizierung der Kernfunktion als Produkt von zwei separaten Kernfunktionen, die nur auf jeweils einer Entfernung basieren:

$$K\left(\frac{l_1-l_{i1}}{h_1}, \frac{l_2-l_{i2}}{h_2}\right) = K_1\left(\frac{l_1-l_{i1}}{h_1}\right) K_2\left(\frac{l_2-l_{i2}}{h_2}\right) \tag{4}$$

Die erste Kernfunktion auf der rechten Seite ist für den Abstand zur Breitenkoordinate  $l_1$  zuständig, die zweite für den Abstand zur Längenkoordinate  $l_2$ . Jeweils wird der Abstand der Beobachtung  $i$  im Raum dabei relativ zu einer Bandweite  $h$



**Abb. 3** Epanechnikov Kernfunktion. Die X-Achse repräsentiert die relative Entfernung vom Zielort. Die Y-Achse repräsentiert das Kerngewicht aus Gl. 5. Die Bandbreite ist auf  $h_j = 1$  eingestellt

betrachtet, die in Breiten- und Längsrichtung unterschiedlich gewählt werden kann. Die Bandweiten  $h_1$  und  $h_2$  definieren dabei ein Rechteck um den Schätzstandort, der durch  $l_1$  und  $l_2$  fixiert ist (siehe auch Abb. 4). Konkret wird zur Bewertung der Distanz folgende, von Epanechnikov (1969) vorgeschlagene Formel, verwendet:

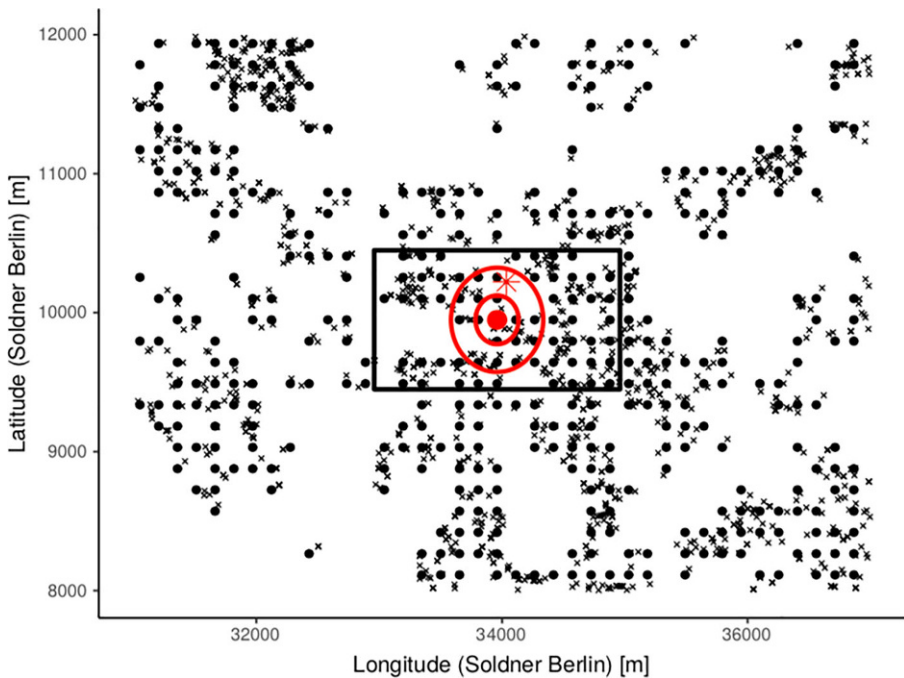
$$K_j \left( \frac{l_j - l_{ij}}{h_j} \right) = I \left( \left| \frac{l_j - l_{ij}}{h_j} \right| \leq 1 \right) * \frac{3}{4} \left\{ 1 - \left( \frac{l_j - l_{ij}}{h_j} \right)^2 \right\} \quad (5)$$

Die Epanechnikov Gewichtsformel besteht aus zwei Teilen. Der erste Teil ist eine Indikatorfunktion, die das Gewicht der Beobachtung  $i$  auf 0 setzt, falls sie in einer der beiden Richtungen außerhalb des durch die Bandweiten definierten Rechtecks liegt. Inuitiv gesprochen werden Beobachtungen, die zu weit entfernt sind, um für den mittleren Bodenwert am Standort informativ zu sein, von der Berechnung damit gewollt ausgeschlossen. Der zweite Teil der Epanechnikov Formel in Gleichung (5) weist den Beobachtungen innerhalb des Rechtecks konkrete, positive Gewichte zu. Dies geschieht gemäß der in Abb. 3 dargestellten Grafik, die die Formel  $\frac{3}{4} \left\{ 1 - \left( \frac{l_j - l_{ij}}{h_j} \right)^2 \right\}$  visualisiert.

Es ist klar, dass das maximale Gewicht an solche beobachteten Transaktionen verteilt wird, die einen relativen Abstand von Null in beide Richtungen aufweisen (also genau am Zielort stattgefunden haben) und dass die Gewichte in dem dargestellten Muster mit zunehmendem Abstand zum Zielort abnehmen.

In Abb. 4 veranschaulichen wir die Gewichtung in einem zweidimensionalen Bild mit Breiten- und Längachsen im Kartenformat.

Nur Datenpunkte innerhalb des durch die Bandweiten definierten Rechtecks um den Zielort erhalten ein positives Gewicht im diesem lokalen Mittelungsverfahren.



**Abb. 4** Illustration der nichtparametrischen Kernregression. Zeigt Rasterpunkte (*Punkte*) und Beobachtungen (*Kreuze*) für ein Stadtviertel in Treptow-Köpenick. Der *rote Punkt* steht für den zu schätzenden Standort. Kerngewichtungen werden durch (*rote*) *Konturlinien* dargestellt. Jede beobachtete Transaktion auf einer Konturlinie erhält das gleiche Gewicht für die Schätzung des Landwerts

Die Höhe des Gewichts gleichgewichtiger Koordinaten wird durch Iso-Höhenlinien dargestellt. Konturlinien, die näher am Zielort (roter Punkt) liegen, sind schmaler und stellen ein höheres Gewicht dar. Folgende Punkte sind eine Zusammenfassung der wesentlichen Kennzeichen der Kernregression für die Bodenwertschätzung:

- Die Kernel-Regression arbeitet an einer bestimmten gewünschten Stelle. Es ist also ein lokales Verfahren, so wie die Bodenbewertung immer lokal funktioniert.
- Der gewünschte Ort kann in den Daten enthalten sein oder auch nicht.
- An der gewünschten Stelle bildet die Kernel-Regression einfach einen gewichteten Stichprobenmittelwert über die Log-Preise der in der Nähe verkauften Grundstücke.
- Die Umgebung wird durch ein Ellipsoid um die Position herum bestimmt. Die Länge und Breite des Ellipsoids wird durch die „Bandbreiten“ gesteuert, die vom Benutzer angegeben werden müssen.
- Der Durchschnitt betrachtet jede Beobachtung, aber diese Beobachtungen außerhalb des Ellipsoids erhalten kein Gewicht. Wie viel Gewicht der (Log-)Preis einer Beobachtung erhält, wird durch die Kernfunktion bestimmt.
- Bei der Bestimmung jedes Gewichts berücksichtigt die Kernfunktion die Unterschiede einer Beobachtung von dem Ort, an dem die Schätzung gewünscht wird, sowohl in Längen- als auch in Breitengradrichtung. Diese Abweichungen werden

in Abhängigkeit von der Bandbreite berechnet. Für jede Beobachtung innerhalb des Nachbarrechtecks sind beide relativen Abstände kleiner oder gleich 1 im Absolutwert.

- Je näher eine Beobachtung am Zielort ist, desto mehr Gewicht erhält sie. Das maximale Gewicht wird somit einer Charge gegeben, die sich genau dort befindet, wo eine Schätzung des erwarteten Log-Preises erforderlich ist.

### 3.1.5 Beispieldaten

Wir veranschaulichen das Vorgehen mit den Transaktionsdaten aus Abb. 2. In Deutschland sind die Gutachterausschüsse gesetzlich verpflichtet und berechtigt, bei allen Immobilientransaktionen Informationen zu verlangen und diese in zu sammeln. Ihre Datenbanken stellen somit eine ausgiebige Quelle für die hier beschriebene regressionsbasierte Bodenwertschätzung dar. Unsere Daten aus Berlin liefern für jede Beobachtung den erforderlichen Input: den (logarithmierten) Transaktionspreis pro Quadratmeter (qm) und die Geokoordinaten der Transaktion. Eine detaillierte Datenbeschreibung finden Sie in Kolbe et al. (2015).

### 3.1.6 Regressionsbasierte Bodenwertschätzungen

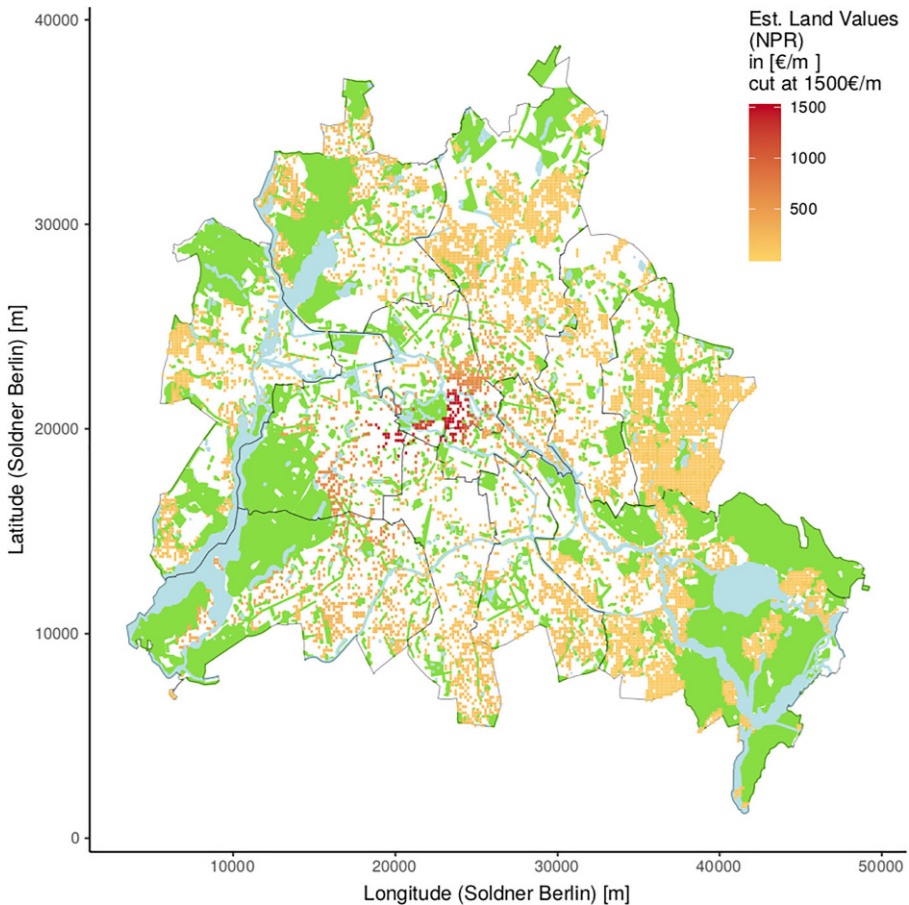
Wir haben den Kernregressionsschätzer von Gl. 3 angewendet, um den Bodenwert auf einem festgelegten Gitter aus Standorten zu schätzen. Dies ermöglicht es uns, eine Bodenwertkarte für Berlin zu erstellen.<sup>6</sup> Um die Bodenwertschätzungen auf der „natürlichen“ Skala zu erhalten, transformieren wir die geschätzten logarithmierten Bodenwerte mit der Formel von Kennedy (1983) zurück in EUR.

Die geschätzte Landwertkarte ist in Abb. 5 dargestellt. Für die Farbgebung verwenden wir das gleiche Schema wie bei der BRW-Karte aus Abb. 1 oben.

Ein Vergleich der beiden Bodenwertkarten zeigt, dass die Farben (und damit die geschätzten Bodenwerte) weitgehend übereinstimmen. Die Korrelation zwischen den Bodenwerten des Gutachterausschusses und den von uns ermittelten Bodenwerten beträgt 0,704.

Unsere Kernregressionsschätzungen basieren ausschließlich auf den Transaktionsdaten aus Abb. 2. Sie konnten daher nur außerhalb der Innenstadt berechnet werden, wo es zu Verkäufen von unbebauten Grundstücken kam. Die BRW-Karte von Abb. 1 hingegen umfasst den gesamten Berliner Raum. Die GAA-Gutachter müssen offensichtlich auch andere Informationen als nur die Transaktionen von unbebauten Grundstücken berücksichtigt haben, um zu ihren flächendeckenden Bodenwertschätzungen zu gelangen. Insbesondere müssen sie Informationen darüber verwendet haben, wie Marktteilnehmer Grundstücke in der Innenstadt bewerten. In diesem Gebiet ist praktisch das gesamte gehandelte Land bebaut. Später, in Abschn. 3.3. unten, wird daher untersucht, wie die regressionsbasierte Schätzung von Grundstückswerten aus Transaktionsdaten von *bebauten* Grundstücken durchgeführt werden kann.

<sup>6</sup> In ähnlicher Weise schätzt McMillen (1996) eine Landwertoberfläche für Chicago mittels Kernel-Regressionen.



**Abb. 5** Kernregressionsbasierte Landwertkarte für Berlin. Zeigt die Konturkarte der geschätzten Landwerte unter Verwendung der Kernregression an. Bandbreite wurde auf  $h_1 = h_2 = 650$  festgelegt

### 3.1.7 Bandbreitenauswahl

Wie gezeigt, liefert die Kernregression eine lokale Schätzung des erwarteten Bodenwertes durch Mittelung über die Log-Preise von Verkäufen in der näheren Umgebung um den Zielort. Eine Schlüsselfrage ist dann, wie groß diese Umgebung gewählt werden sollte. Im der konkreten Anwendung haben wir Bandweiten  $h_1 = 2$  [Meter] und  $h_2 = 1$  [Meter] verwendet, aber bislang keine Begründung für diese spezielle Wahl geliefert.

Ein naheliegender Gedanke könnte sein,  $h_1$  und  $h_2$  so klein wie möglich zu wählen, d. h. auf 0 zu setzen, um sicherzustellen, dass nur die Transaktionen in die lokale Durchschnittsbildung einbezogen werden, die genau an jenem Zielort stattgefunden haben, für den eine Bodenwertschätzung gewünscht wird. Dies würde eine Verzerrung der Bodenwertschätzung verhindern, die entstehen könnte, wenn Transaktionen in der Nähe tendenziell zu einem (leicht) höheren oder niedrigeren Preis stattfinden.

Durch diese Bandweitenwahl kann es jedoch dazu kommen, dass nur sehr wenige oder sogar keine Beobachtungen am Zielort für die Mittelung zur Verfügung stehen. Dies kann die lokale Schätzung sehr unzuverlässig oder gänzlich unmöglich machen. Aus dieser Perspektive sind größere Bandbreiten wünschenswert, da sie sicherstellen, dass mehr Beobachtungen in die lokale Berechnung einfließen. Eine optimale Wahl der Bandbreiten schafft ein Gleichgewicht zwischen diesen beiden Anliegen.

Es hat sich gezeigt, dass die Größe der Bandweiten datengetrieben und objektiv durch die Methode der Kreuzvalidierung bestimmt werden kann. Wie der Name andeutet, verwendet die Kreuzvalidierung in einem iterativen Prozess jeden Datenpunkt einmal nur zum „Validieren“ (und nicht zur Schätzung). Auf diese Weise kann ein objektives Maß der Güte der Schätzung für ein gegebenes Bandweitenpaar erstellt werden. Wird dies für viele verschiedene Bandweitenkombinationen durchgerechnet, kann am Ende jenes Bandweitenpaar bestimmt werden, das die beste „validierte“ Leistung liefert. Dieses rechenintensive Verfahren stellt sicher, dass die Bodenwertschätzungen die lokalen Verhältnisse in den Transaktionen relativ unverzerrt darstellen, aber gleichzeitig auf einer nicht zu kleinen Anzahl von Beobachtungen beruhen.

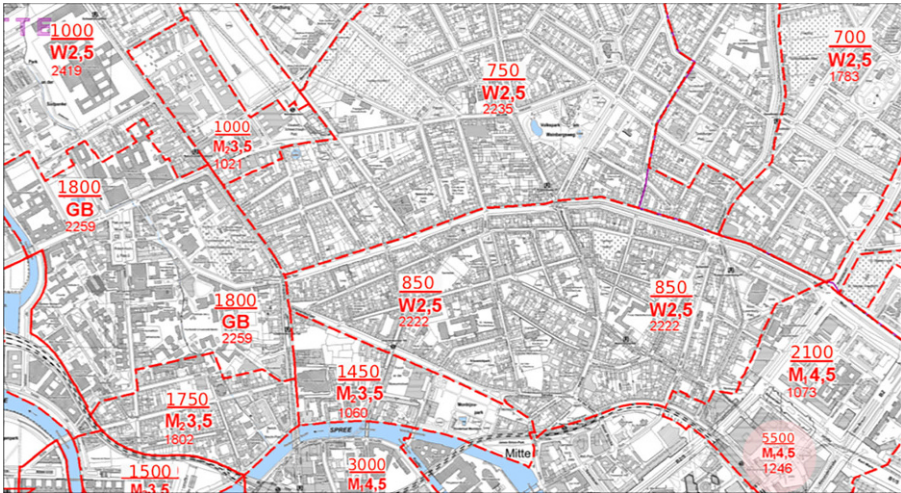
### 3.1.8 Statistische Eigenschaften

Alle praktisch durchführbaren Bewertungsmethoden liefern nur Schätzungen des tatsächlichen Bodenwertes an einem bestimmten Standort. Es ist ein großer Vorteil des hier beschriebenen statistischen Ansatzes, dass er es ermöglicht, die Unsicherheit zu quantifizieren, die ein unvermeidlicher Aspekt jeder Schätzung ist. Da die Kernregressionsmethode im Wesentlichen ein gewichteter lokaler Durchschnitt der Log-Preise ist, ist es relativ leicht möglich, ihre Genauigkeit einzuschätzen. Sie ergibt sich aus dem Verhältnis der Varianz (ein Maß für das „Rauschen“ in den Log-Preisen) und der effektiven Größe der Stichprobe, die in den Durchschnitt eingegangen ist. Einzelheiten sind in Härdle (1990, Kap. 4) enthalten. Dies kann dazu verwendet werden, um Vertrauensintervalle um die Bodenwertschätzung herum zu konstruieren, die den wahren Bodenwert in einem vordefinierten Vertrauensniveau abdecken.

### 3.1.9 Verwandte Literatur und alternative Methoden

Neben der Kernregression gibt es weitere nichtparametrische Schätzverfahren, die ebenfalls auf lokaler Mittelung basieren.<sup>7</sup> Colwell und Munneke (2003) schätzen beispielsweise Standortwerte für Chicago aus Transaktionen von unbebautem Land mittels Glättungssplines. Genau wie Kernregressionen gehen Glättungssplines auch davon aus, dass sich die Landwerte von Standort zu Standort nicht abrupt ändern. Diese Annahme steht jedoch im Widerspruch zu den blockweise gewachsenen Städten, in denen benachbarte Stadtteile durch Straßen scharf abgegrenzt werden und sich in ihrem Charakter deutlich unterscheiden können. Tatsächlich zeigt der von der

<sup>7</sup> Härdle et al. (2004, Kap. 4) geben einen Überblick über alternative nichtparametrische Methoden.



**Abb. 6** Detail der sachverständigen Landwertkarte (BRW). Zeigt einen Bereich etwas nördlich des zentralen Geschäftsviertel und den nördlichen Teil der Museumsinsel. Bezugsdatum für die Karte ist der 1. Januar 2010. Quelle: Geoportail Berlin/Bodenrichtwerte 01.01.2010

Berliner Gutachterausschuss veröffentlichte Ausschnitt der BRW-Karte in Abb. 6, dass die von diesen Experten produzierte Bodenwertfläche nicht glatt ist.

### 3.2 Identifizierung von Bodenwertzonen mittels nichtparametrischer adaptiver Regression

In Kolbe et al. (2015) verwenden wir Adaptive Gewichtsglättungen (Adaptive Weights Smoothing AWS), um eine stückweise konstante Landwertkarte zu schätzen, ähnlich der Karte, die vom GAA Berlin im Bodenrichtwertatlas (BRW Atlas) veröffentlicht werden. AWS baut direkt auf dem Kernregressionsschätzer auf, geht aber iterativ vor und arbeitet mit einer leicht vereinfachten Kernfunktion. Konkret wird die Kernfunktion von Gl. 4 zu

$$K\left(\frac{l_{1i}-l_{1j}}{h_1}, \frac{l_{2i}-l_{2j}}{h_2}\right) = \omega_{ij} \quad (6)$$

Das heißt, die räumliche Entfernung einer Beobachtung  $j$  vom Zielort der Wertermittlung wird nun einfach als Summe der absoluten Differenzen der Breiten- und Längenkoordinaten gemessen, geteilt durch eine gemeinsame Bandweite  $h$ . Im iterativen AWS-Verfahren wird diese Bandweite  $h$  zunächst im Initialisierungsschritt auf eine kleine Zahl gesetzt. Der resultierende anfängliche Bodenwertschätzwert wird als  $\hat{\theta}^0(l_{1i}, l_{2i})$  oder kurz  $\hat{\theta}_i^0$  bezeichnet. Das Superskript „0“ bezeichnet die Initialisierung, d.h. den „nullten“ Schritt des Algorithmus. In dieser anfänglichen Schätzung werden, auf Grund der klein gewählten Bandweite, nur wenige, räumlich sehr nahe Beobachtungen verwendet, um den Initialisierungsbodenwert  $\hat{\theta}_i^0$  an einem Ort  $(l_{1i}, l_{2i})$  zu schätzen.

Im nächsten Schritt (und allen nachfolgenden Schritten) ist die Bodenwertschätzung immer noch ein lokaler gewichteter Durchschnitt:

$$\hat{\theta}^1(l_{1i}, l_{2i}) = \frac{\sum_{j=1}^n \omega_{ij}^1 p_j}{\sum_{j=1}^n \omega_{ij}^1} \quad (7)$$

Der wesentliche Unterschied besteht darin, dass eine zweite Kernfunktion in die Gewichtung von Beobachtungen einfließt. Das Gewicht einer Beobachtung wird nun als ein Produkt des vertrauten „Distanzkerns“ von Gl. 6 und des sogenannten „Levelkerns“:  $\omega_{ij}^1 = K(\text{distance}_{ij}^1) \times K(\text{level}_{ij}^1)$  gebildet, wobei

$$\text{level}_{ij}^1 = \left( \frac{\hat{\theta}_i^0 - \hat{\theta}_j^0}{\sqrt{2\hat{\sigma}_\epsilon^2}} \right)^2 \cdot \frac{n_i^0}{\lambda} \quad (8)$$

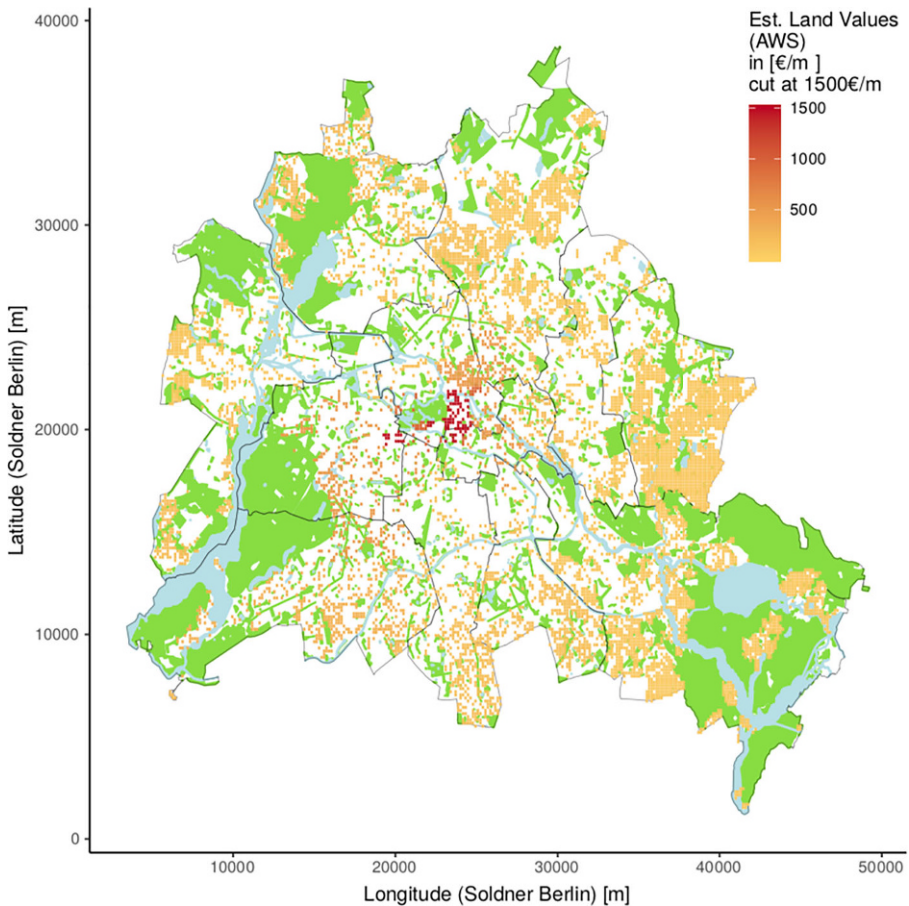
Entscheidend ist die Differenz  $\hat{\theta}_i^0 - \hat{\theta}_j^0$  im Zähler des ersten Terms auf der rechten Seite, welche den Unterschied zwischen den Bodenwertschätzungen an den Standorten  $l_{1j}$ ,  $l_{2j}$  und  $l_{1i}$ ,  $l_{2i}$  in der Iteration 0 wiedergibt. Wenn die beiden Schätzungen nahe beieinander liegen, scheinen an beiden Standorten sehr ähnliche Bodenwerte vorzuliegen. Folglich erhält der beobachtete (logarithmische) Grundstückspreis von Standort  $l_{1j}$ ,  $l_{2j}$  ein erhebliches Gewicht, um in der nachfolgenden Iteration eine Bodenwertschätzung an Standort  $l_{1i}$ ,  $l_{2i}$  zu bilden. Dieses Merkmal macht AWS strukturell anpassungsfähig und ermöglicht es, aus den Daten Bereiche mit homogenem Bodenwerten zu bestimmen, analog zu den vom GAA definierten Bodenrichtwertzonen. Dies wird erreicht, indem die Bandbreite  $h$  in aufeinanderfolgenden Iterationen sukzessive erhöht wird und der Glättungsparameter  $\lambda$  in Gl. 8 als Schwellenwert für die Beurteilung der Nähe von Bodenwertschätzungen von verschiedenen Standorten verwendet wird.

### 3.2.1 Geschätzte Bodenwerte

Da es sich bei AWS um ein lokales, iteratives Verfahren handelt, ist es rechenintensiv. Wir haben daher Landwertschätzungen nicht an einzelnen Koordinaten, sondern auf einem Gitter von vordefinierten Intervallen, d.h. kleinen Feldern der Größe  $152 \times 152$  m, berechnet. Die geschätzten Bodenwerte für jedes dieser Felder sind in Abb. 7 dargestellt. In der wir die gleiche Farbgebung wie in der BRW-Karte von Abb. 1 verwenden.

Der Vergleich beider Karten zeigt, dass die AWS-Schätzfelder in Abb. 7, genau wie der klassische Kernschätzer, nur einen Teil der kontinuierlich schattierten Bereiche der Bodenrichtwerte aus Abb. 1 abdecken, da die AWS-Berechnungen ausschließlich Transaktionen unbebauter Grundstücke verwenden und im Innenstadtbereich daher keine Ergebnisse liefern können. Dort, wo Vergleiche angestellt werden können, stimmen AWS- und BRW-Landwertschätzungen in Bezug auf ihre räumlichen Farbmuster recht gut überein.





**Abb. 7** AWS-basierte Landwertkarte für Berlin. Zeigt die Konturkarte der geschätzten AWS-Landwerte an. Die Bandbreite ist auf  $\lambda^* = 19,9$  eingestellt

### 3.2.2 Glättungsparameterauswahl und -validierung mit BRW

Wie bei jeder lokalen, nichtparametrischen statistischen Methode hängt auch die Leistung von AWS von einem Glättungs- oder Steuerungsparameter ab, der vom Benutzer angegeben werden muss. Im Falle von AWS bestimmt der Parameter  $\lambda$  in Gl. 8, wie sensibel der Schätzer auf lokale Schwankungen des geschätzten Bodenwerts in der vorherigen Iteration reagiert. Wie man  $\lambda$  optimal auswählt, ist Thema der laufenden Forschung. Die Schätzungen in Abb. 7 wurden mit dem von Polzehl und Spokoyny (2006, 2008) vorgeschlagenen Wert  $\lambda^* = 19,9$  berechnet.

Um die Sensitivität unserer Ergebnisse hinsichtlich der Wahl von  $\lambda$  zu untersuchen, haben wir den AWS-Algorithmus für die acht verschiedenen  $\lambda$ -Werte in der obersten Zeile von Tab. 1 durchgeführt. Dabei zeigt sich, dass – gemessen am Bestimmtheitsmaß  $R^2$  – das Verfahren relativ unempfindlich gegenüber der  $\lambda$  Wahl ist.

**Tab. 1** Gütemaße  $R^2$  für unterschiedliche Regressionen von  $BRW_i$  und Grundstückspreisen  $y_i$  AWS-Bodenwerten  $\theta_i$ . Die Regressionen schließen Sie eine Konstante ein. Die Anzahl der für Regressionen in der ersten Zeile verwendeten Beobachtungen beträgt 7222 und 7448 für Regressionen in der zweiten Zeile

|            | $\lambda^*$ | $\lambda$ |        |         |         |         |         |        |
|------------|-------------|-----------|--------|---------|---------|---------|---------|--------|
|            | 19,9        | 3,8415    | 4,4756 | 10,5180 | 16,8410 | 23,2840 | 29,7938 | 36,346 |
| BRW        | 0,7747      | 0,7274    | 0,7390 | 0,7640  | 0,7720  | 0,7828  | 0,7733  | 0,7690 |
| Bodenwerte | 0,6992      | 0,8661    | 0,8526 | 0,7734  | 0,7195  | 0,6764  | 0,6525  | 0,6418 |

Die  $R^2$ -Werte wurden durch Kleinste-Quadrate Regression der für einen bestimmten Wert von  $\lambda$  erhaltenen AWS-Landwerte entweder auf beobachtete Bodenpreise oder auf Bodenrichtwerte berechnet. Die recht hohen  $R^2$ -Werte (Skala von 0 bis 1) zeigen, dass die Übereinstimmung zwischen AWS-Landwertschätzungen und sowohl den Grundstückspreisen als auch den BRW-Landwerten für alle Werte von  $\lambda$  recht gut ist. In einer detaillierteren Analyse, die in Kolbe et al. (2015) zu finden ist, stellten wir fest, dass AWS auch erfolgreich Bereiche homogener Bodenwerte bestimmt (wie im BRW-Atlas). Zusammenfassend zeigt unsere Forschung, dass AWS ein transparentes, statistisches Verfahren ist, welches in der Lage ist, professionell geschätzte Bodenwerte zu reproduzieren sowohl hinsichtlich des mittleren Preisniveaus an einer bestimmten Stelle als auch hinsichtlich der geografischen Struktur.

### 3.3 Weiterführende Literatur

AWS wurde auch von Helbing et al. (2017) zur Schätzung der landwirtschaftlichen Bodenwerte eingesetzt. Eine Alternative zu AWS für die nichtparametrische Regression, wenn die Regressionsfläche Sprünge oder Kanten aufweisen kann, ist die z. B. in Vidakovic (1999) beschriebene Wavelet-Methode.

### 3.4 Schätzung von Grundstückswerten aus Immobilientransaktionen mittels semiparametrischer Regression

Sowohl die Kernregression als auch die adaptive Gewichtsglättung mit AWS schätzen Bodenwerte, indem sie über die Preise unbebauter Grundstücke mitteln und auf diese Weise ausschließlich mit den direktesten und „saubersten“ Marktinformationen über den Wert des Bodens an einem bestimmten Standort arbeiten. Allerdings sind solche Informationen in der Regel nicht für das Zentrum einer Stadt verfügbar, in dem praktisch alle gehandelten Grundstücke bebaut sind. Daten über Transaktionsdaten von bebauten Grundstücken enthalten Informationen über das Bündel von Grundstücken und Gebäuden. Daher müssen für die Schätzung des Bodenwertes an solchen innerstädtischen Standorten die beobachteten Preise von Häusern oder Eigentumswohnungen in ihre Land- und ihre Gebäudekomponente aufgeteilt werden. Auch dies kann durch eine Regressionsanalyse erreicht werden.

### 3.4.1 Erforderliche Daten

In diesem Abschnitt gehen wir davon aus, dass die Transaktionsdaten Informationen über (a) den Preis, (b) den Standort und (c) den Gebäudecharakter einer Immobilie enthalten. Kurz gesagt, es wird angenommen, dass die Daten  $m$  aus  $n$  Beobachtungen der folgenden Form zu bestehen.

$$\{p_i, l_{1i}, l_{2i}, X_{1i}, \dots, X_{pi}\}, i = 1, \dots, n$$

wobei  $p_i = \ln(P_i)$  der logarithmierte Preis der Immobilie (Gebäude mit Grundstück) ist,  $l_{1i}$  und  $l_{2i}$  wie gehabt die Geokoordinaten ihrer Lage darstellen und  $X_{1i}, \dots, X_{pi}$  Gebäudemerkmale wie Grundfläche oder Alter sind.

### 3.4.2 Semiparametrisches Regressionsmodell

Wir beginnen mit der Annahme, dass der Log-Preis einer Immobilie additiv in den Wert des Gebäudes und den Wert des Grundstücks aufgeteilt werden kann, um das folgende partiell lineare Regressionsmodell zu formulieren:

$$p_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + \theta(l_{1i}, l_{2i}) + \epsilon_i p_i \quad (9)$$

Im Vergleich zu unserem ursprünglichen Regressionsmodell in Gl. 1 enthält das Modell von Gl. 9 zwei Teile, den „Gebäudeteil“ und den „Lageteil“. Im linearen Teil des Modells bilden die Koeffizienten  $\beta_1, \dots, \beta_p$  den Einfluss der zugehörigen Gebäudeeigenschaft auf den Log-Preis der Immobilie ab. Der Lageeinfluss wird wie oben über eine nichtparametrische Funktion  $\theta(l_{1i}, l_{2i})$  der Lagekoordinaten modelliert.

Alle Variablen werden pro Quadratmeter Grundstücksgröße gemessen.

### 3.4.3 Zweistufige Schätzung

Unser eigentliches Ziel bleibt es, die nichtparametrische Bodenwertfläche  $\theta(l_{1i}, l_{2i})$  zu schätzen. Dazu werden aber zunächst im ersten Schritt die Koeffizienten des Gebäudeteils geschätzt. Diese Koeffizientenschätzungen bezeichnen wir als  $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_p$ . Wir verwenden sie, um den Gebäudewert aus dem Immobilienpreis zu entfernen.

$$\widehat{u}_i = p_i - \left( \widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \dots + \widehat{\beta}_p X_{pi} \right). \quad (10)$$

Die daraus resultierenden Abweichungen des Immobilienpreises vom Gebäudewert, die „Residuen“  $\widehat{u}_i$ , werden im zweiten Schritt als abhängige Variable verwendet. In diesem letzten Schritt werden die Residuen  $\widehat{u}_i$  mit einem nichtparametrischen Regressionsverfahren wie Kernel-Regression oder AWS auf die Standortkoordinaten  $l_{1i}$  und  $l_{2i}$  regressiert. Wir veranschaulichen diesen zweistufigen Ansatz noch einmal mit Transaktionsdaten aus Berlin.

### 3.4.4 Beispieldaten

Wir veranschaulichen das Verfahren mit Berliner Daten, die sich aus 27.549 Einfamilienhaustransaktionen und 166.839 Verkäufen von Eigentumswohnungen zusammensetzen, die zwischen 1996 und 2013 stattfanden. Wohnungseigentumstransaktionen stellen eine Quelle für die Berechnung von Bodenwerten in der Mitte Berlins dar, während Einfamilienhaustransaktionen überwiegend außerhalb des Innenstadtrings stattfinden. Daher ist es notwendig Verkaufsinformationen für beide Arten von Immobilien aufzunehmen um einen umfassenden Satz von Bodenwertschätzungen zu erhalten. Die Kehrseite dieser Kombination von Haus- und Wohnungsverkäufen in einer Analyse ist eine Erhöhung der Komplexität des Modells in Gl. 9, weil nun zwei Sätze von Regressionskoeffizienten benötigt werden, einen für jede Art von Wohneigentum, um deren unterschiedliche Spezifikationen abbilden zu können.

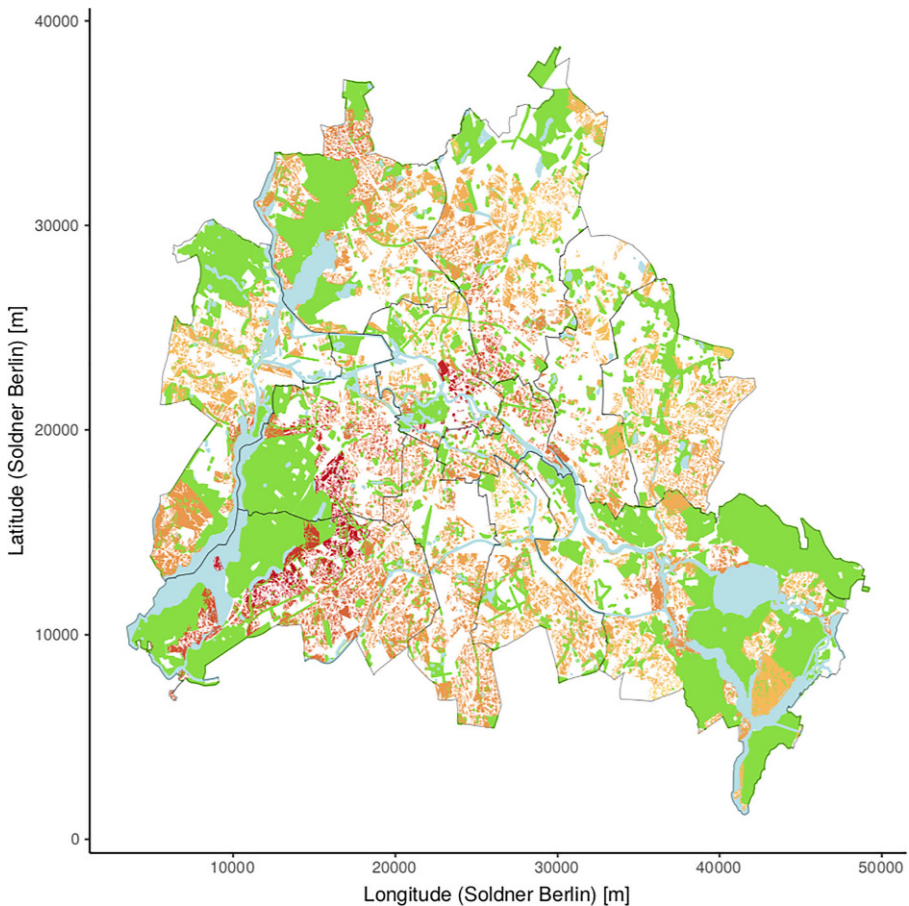
### 3.4.5 Semiparametrische Bodenwertschätzungen für Berlin

Um die Koeffizienten der Gebäudeeigenschaften im ersten Schritt zu schätzen, haben wir den von Yatchew (1997) vorgeschlagenen Schätzer verwendet. Die Grundidee dieses Schätzers besteht darin, dass der Bodenwert  $\theta(l_1, l_2)$  bei der Berücksichtigung von Preisunterschieden für benachbarte Beobachtungen (Immobilien) vernachlässigt werden kann. Daher werden die Daten zunächst geografisch gemäß ihrer Nähe zueinander angeordnet. Im Folgenden werden die Preisunterschiede nahe beieinander liegender Immobilien auf die Unterschiede in deren Gebäudeeigenschaften regressiert. Damit wird der Gebäudeeinfluss (d. h., die Regressionskoeffizienten des Gebäudeteils des Modells) bestimmt und kann gemäß Gl. 10 abgezogen werden. Im zweiten Schritt haben wir dann AWS eingesetzt, um durch lokale Mittelung der Residuen  $\hat{u}_i$  die in Abb. 8 dargestellte Landwertfläche zu berechnen.

In Kolbe et al. (2012) haben wir die Korrelation zwischen den BRW-Werten und unseren semiparametrischen Bodenwertschätzungen, die nur auf Haustransaktionen basieren, berechnet. Für diese Teilmenge der Daten (Einfamilienhäuser) fanden wir eine starke positive Korrelation von 0,845, was darauf hindeutet, dass der in diesem Abschnitt beschriebene zweistufige semiparametrische Regressionsansatz in der Lage ist valide Bodenwerte aus Daten über Immobilientransaktionen zu extrahieren.

Nur wenige Studien haben bislang Bodenwerte aus Daten von Hauspreistransaktionen mittels semiparametrischer Regression geschätzt. Cheshire and Sheppard (1995) und Bryan and Sarte (2009) sind Beispiele hierfür; keine dieser Studien vergleicht die geschätzten Bodenwerte mit von Experten geschätzten Richtwerten.<sup>8</sup> Anglin und Gencay (1996), Clapp (2003) und Case et al. (2004) passen unter anderem semiparametrische Regressionsmodelle an Hauspreise an, um ihre Eignung für Massenschätzungen zu bewerten. Die Ergebnisse dieser Studien zeigen, dass die nichtparametrische Modellierung der Standortkomponente die Genauigkeit von Hauspreisschätzungen im Vergleich zu Standardmethoden wie der (parametrischen) linearen Regression verbessert.

<sup>8</sup> Das Fehlen solcher Benchmarks ist der Grund, warum Bodenwerte überhaupt erst geschätzt werden müssen.



**Abb. 8** Landwertkarte basierend auf semiparametrischer Regression. Zeigt die Konturkarte der Landwerte, die aus den Residuen in Gl. 10 via AWS geschätzt wurden. Die Bandbreite ist auf  $\lambda^* = 19,9$  eingestellt

## 4 Fazit

In mehreren anderen Ländern werden computergestützte Massenschätzungssysteme für die Vermögenssteuerveranlagung eingesetzt, aber dies ist in der Diskussion in Deutschland wenig beachtet worden.<sup>9</sup> In diesem Beitrag zeigen wir, wie dies auch in Deutschland mit den Transaktionsdaten geschehen könnte, zu deren Erhebung die lokalen Gutachterkommissionen gesetzlich verpflichtet sind. Wir haben zwei Arten von Transaktionsdaten berücksichtigt: den geokodierten Verkauf von unbebauten Grundstücken und den geokodierten Verkauf von bebauten Grundstücken. Erstere stellen die reinste Quelle der Grundstücksbewertung durch die Marktteilnehmer dar. Letztere benötigen eine statistische „Trennung“ des Verkaufspreises in eine Gebäu-

<sup>9</sup> Ausnahmen sind Senatorin für Finanzen (2010) und Houben (2017).

de- und eine Grundstückskomponente. In beiden Fällen haben wir gezeigt, wie man moderne nichtparametrische Regressionsverfahren einsetzen kann, um Bodenwerte an einem bestimmten Ort zu schätzen, an dem Verkäufe in der Nähe getätigt wurden. Wir haben in unserer Anwendung dieser Methoden auf Daten aus Berlin festgestellt, dass die daraus resultierenden Bodenwertschätzungen typischerweise gut mit (kostspieligeren) Expertenschätzungen der Bodenwerte übereinstimmen. Im Gegensatz zu diesen Expertenschätzungen sind die von uns verwendeten statistischen Methoden transparent und standardisierbar.

Wir sind uns bewusst, dass es mehrere Bereiche gibt, in denen weitere Untersuchungen erforderlich sind. Erstens könnte es Kommunen geben, in denen eine relativ geringe Anzahl von Transaktionen vorliegt. Zusätzliche Informationen wie Angebotspreise könnten insbesondere in diesen Fällen eine sinnvolle Datenerweiterung sein. Das Volumen der Angebotsdaten ist stets weit höher als die Anzahl der tatsächlichen Transaktionen, weil bei weitem nicht jedes Angebot zu einer Transaktion führt. Dies erfordert aber eine Prüfung, ob Angebotspreise die Transaktionspreise valide vervollständigen können.

Zweitens stellt sich die Frage nach den Kosten und der effizienten Organisation des Verwaltungsprozesses. Es gibt Hinweise darauf, dass die steuerliche Bewertung von Immobilien mit Skaleneffekten verbunden sein kann (Sjoquist und Walker 1999) und dass es optimale Betriebsgrößen von Bewertungsbüros geben kann (Krupa 2017).<sup>10</sup> Sehr interessant wäre auch die Analyse der Beziehung zwischen Kosten und Genauigkeit der Wertermittlung mit rigorosen Methoden, wie von Mehta und Giertz (1996) vorgeschlagen. Schließlich konzentriert sich unsere Arbeit auf Marktwerte von unbebautem Land. Statistische Methoden können äquivalent zur Bewertung von bebautem Land verwendet werden, siehe zum Beispiel Schulz et al. (2014).

**Danksagung** Kolbe und Werwatz danken der Deutschen Forschungsgemeinschaft, DFG-Forschergruppe FOR2569 „Agrarlandmärkte – Effizienz und Regulierung“, für die finanzielle Unterstützung.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) 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.

## Literatur

- Almy RR (2014) Valuation assessment of immovable property. Working Papers on Fiscal Federalism, Bd. 19. OECD, Paris
- Almy RR, Ferguson AG (2010) Valuing our world: potential roles for AVMs and CAMA. In: Proceedings of the Union of Pan-American Valuers' Association XXV Congress, Appraisal Institute, Chicago (<http://www.mrcl.com.br/upav/19.pdf>)
- Anglin PM, Gencay R (1996) Semiparametric estimation of a hedonic price function. *J Appl Econ* 11:633–648
- Back K (1970) Land value taxation in light of current assessment theory and practice. In: Holland DM (Hrsg) Land value taxation in light of current assessment theory and practice. Committee on Taxation,

<sup>10</sup> Keine der beiden Studien berücksichtigt explizit Steuerverwaltungen, die computergestützte Massenbewertungssysteme verwenden, obwohl Sjoquist und Walker (1999) erwähnen, dass mehrere der in ihrer Stichprobe enthaltenen Büros diese verwenden.

- Resources and Economic Development, Bd. 5. University of Wisconsin Press, S 37–54, Milwaukee, Wisconsin. <https://catalogue.nla.gov.au/Record/1897902>
- Baranska A (2013) Real estate mass appraisals in selected countries: functioning systems and proposed solutions. *Real Estate Manag Valuat* 21:35–42
- Bidanset PE (2014) Moving automated valuation models out of the box: The global geography of AVMs, Fair and Equitable, S. 3–7. [https://www.iaao.org/media/Topics/AVMs/FE\\_July\\_Bidanset.pdf](https://www.iaao.org/media/Topics/AVMs/FE_July_Bidanset.pdf). Zugegriffen: 14.10.2019
- Bourassa SC (2009) The U.S. experience. In: Dye RF, England RW (Hrsg) *Land value taxation. Theory, evidence, and practice*. Lincoln Institute of Land Policy, Cambridge MA, S 11–26
- Briefing land-value tax: on firmer ground (2018b, 11. Aug.) *The Economist* S 18–20
- Brueckner JK (1986) A modern analysis of the effects of site value taxation. *Natl Tax J* 39:49–58
- Bryan K, Sarte PD (2009) Semiparametric estimation of land price gradients using large data sets. *Econ Q* 95:53–74
- Case B, Clapp JM, Durbin R, Rodriguez M (2004) Modeling spatial and temporal house price patterns: a comparison of four models. *J Real Estate Finance Econ* 29:167–191
- Cheshire P, Sheppard S (1995) On the price of land and the value of amenities. *Economica* 62:247–267
- Clapp JM (2003) A semiparametric method for valuing residential locations: applications to automated valuation. *J Real Estate Finance Econ* 27:303–320
- Colwell PF, Munneke HJ (2003) Estimating a price surface for vacant land in an urban area. *Land Econ* 79:15–28
- Epanechnikov VA (1969) Non-parametric estimation of a multivariate probability density. *Theory Probab Appl* 14:153–158. <https://doi.org/10.1137/1114019>
- Franzen RCD (2009) International experience. In: Dye RF, England RW (Hrsg) *Land value taxation. Theory, evidence, and practice*. Lincoln Institute of Land Policy, Cambridge MA, S 27–50
- Fuest C, Immel L, Meier V, Neumeier F (2018) *Die Grundsteuer in Deutschland: Finanzwissenschaftliche Analyse und Reformoptionen*, ifo Studie, Forschungsgruppe Steuer- und Finanzpolitik, Leibniz-Institut für Wirtschaftsforschung an der Universität München e. V. Studie im Auftrag von Haus & Grund Deutschland – Zentralverband der Deutschen Haus, Wohnungs- und Grundeigentümer e. V. sowie ZIA Zentraler Immobilien Ausschuss e. V. <https://www.ifo.de/DocDL/ifo-studie-2018-fuest-et-al-grundsteuer.pdf>. Zugegriffen: 14.10.2019
- Härdle WK (1990) *Applied nonparametric regression*. Econometric society monographs. Cambridge University Press, Cambridge
- Härdle WK, Müller M, Sperlich S, Werwatz A (2004) *Nonparametric and semiparametric models*. Springer, Berlin
- Hefferan MJ, Boyd T (2010) Property taxation and mass appraisal valuations in Australia and New Zealand. *Prop Manag* 28:149–162
- Helbing G, Shen Z, Odening M, Ritter M (2017) Estimating location values of agricultural land. *Ger J Agric Econ* 66:188–201
- Hey J (2017) *Verfassungsmäßigkeit der Reform der Bemessungsgrundlage der Grundsteuer und der Entwicklung der Grundsteuerhebesätze*, Gutachten im Auftrag der BID Bundesarbeitsgemeinschaft Immobilienwirtschaft Deutschland. Institut für Steuerrecht Universität zu Köln, Köln
- Homburg S (2018) Ein Vorschlag zur Grundsteuerreform. *Wirtschaftsdienst* 98:169–175
- Houben H (2017) *Bewertung für grundsteuerliche Zwecke zwischen Wunsch und Wirklichkeit*. *Steuer Wirtsch* 2:184–199
- Kennedy P (1983) Logarithmic dependent variables in prediction bias. *Oxf Bull Econ Stat* 45:389–392
- Kolbe J, Schulz R, Wersing M, Werwatz A (2012) Location, location, location: extracting location value from house prices. SFB 649 Discussion Paper, Bd. 2012-040 (<http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2012-040.pdf>). Zugegriffen: 14.10.2019
- Kolbe J, Schulz R, Wersing M, Werwatz A (2015) Identifying Berlin's land value map using adaptive weights smoothing. *Comput Stat* 30:767–790
- Kreutziger S, Schaffner M, Stephany R (2013) *Bewertungsgesetz, Kommentar*, 3. Aufl. C.H. Beck, München
- Krupa O (2017) Government consolidation in property tax administration. *State Local Gov Rev* 49:27–36
- Leaders: stuck in the past (2018a, 11. Aug.) *The Economist* S 9
- Löhr D (2011) Reform der Grundsteuer: Zu einem blinden Fleck in der Stellungnahme des Wissenschaftlichen Beirats beim Bundesministerium der Finanzen. *Wirtschaftsdienst* 91:333–338
- McCluskey WJ, Franzsen R (2013) Non-market value and hybrid approaches to property taxation. In: McCluskey WJ, Cornia GC, Walters LC (Hrsg) *A primer on property tax. Administration and policy*. Wiley-Blackwell, Chichester, S 287–305

- McCluskey WJ, Davis P, McCord M, McIlhatton D, Haran M (2013) Computer assisted mass appraisal and the property tax. In: McCluskey WJ, Cornia GC, Walters LC (Hrsg) *A primer on property tax. Administration and policy*. Wiley-Blackwell, Chichester, S 307–338
- McMillen DP (1996) One hundred fifty years of land values in Chicago: a nonparametric approach. *J Urban Econ* 40:100–124
- Mehta S, Giertz F (1996) Measuring the performance of the property tax assessment process. *Natl Tax J* 49:73–85
- Nadaraya EA (1964) On estimating regression. *Theory Probab Appl* 9:141–142
- Oates WE, Schwab RM (2009) *The simple analytics of land value taxation*. In: Dye RF, England RW (Hrsg) *Land value taxation. Theory, evidence, and practice*. Lincoln Institute of Land Policy, Cambridge MA, S 51–71
- Polzehl J, Spokoiny V (2006) Propagation-separation approach for local likelihood estimation. *Probab Theory Relat Fields* 135:335–362
- Polzehl J, Spokoiny V (2008) Structural adaptive smoothing by propagation-separation-methods. In: Chen C, Härdle WK, Unwin A (Hrsg) *Handbook of data visualization*. Springer, Berlin, Heidelberg, S 471–492
- Schulz R, Wersing M, Werwatz A (2014) Automated valuation modelling: a specification exercise. *J Prop Res* 31:131–153
- Senatorin für Finanzen, Freie Hansestadt Bremen (2010) *Grundsteuer auf der Basis von Verkehrswerten, Machbarkeitsstudie*
- Sjoquist DL, Walker MB (1999) Economies of scale in property tax assessment. *Natl Tax J* 52:207–220
- Vidakovic B (1999) *Statistical modeling by wavelets*, 1. Aufl. Wiley, New York
- Watson GS (1964) Smooth regression analysis. *Sankhya Indian J Stat* 26:359–372
- Wissenschaftlicher Beirat beim BMF (2010) *Reform der Grundsteuer, Stellungnahme*. Bundesministerium der Finanzen, Berlin
- Yatchew A (1997) An elementary estimator of the partial linear model. *Econ Lett* 57:135–143