

Projekt Immobilienmarkt



TECHNISCHE
UNIVERSITÄT
DARMSTADT

WiSe 2023/24

**Sozioökonomische Faktoren und ihre Auswirkung
auf Mietpreisunterschiede**

Sebastian Geis | 2737261

David Kazakovic | 2611484

Jamal Pierre Mengste | 2285920

Pascal Gorzelany | 2390718

26.02.2024

Inhaltsverzeichnis

1. Einleitung	5
1.1. Problemstellung	5
1.2. Zielsetzung	6
1.3. Aufbau der Arbeit	7
2. Grundlagen	8
2.1. Wohnungsmarkt allgemein	8
2.2. Wohnungsmarkt Berlin	9
2.3. Demografie Berlin	10
2.4. Charakter Berlins	11
2.5. Auswahl der Variablen	12
3. Daten	14
3.1. Datenaufbereitung	14
3.2. Explorative Datenanalyse - Sozialdatensatz	16
3.3. Explorative Datenanalyse - Mietdatensatz	21
4. Analyse	25
4.1. Konzentrationen von Bevölkerungsgruppen	25
4.1.1. Betrachtung anhand einzelner Variablen	25
4.1.2. Betrachtung anhand einer Clusteranalyse	32
4.2. Regressionsanalyse	38
4.2.1. Miete über Jahre	39
4.2.2. Miete in Brennpunkten	42
4.2.3. Miete in den Clustern	44
4.2.4. Lokationskoeffizient	47
5. Diskussion	51
5.1. Interpretation der Ergebnisse	51
5.2. Limitation der Arbeit	52
6. Fazit	54
Literaturverzeichnis	57
A. Weitere Abbildungen	60
B. Beitragsübersicht der Autoren	62

Abbildungsverzeichnis

1.	Monatlich genehmigte Wohnungen in Deutschland	8
2.	Karte Berlin	15
3.	Sozialdatensatz auf Karte 2015	18
4.	Miete auf Karte 2015	22
5.	Verteilung Kaltmiete	23
6.	Arbeitslosenquote 2010 und 2019	26
7.	LQ Arbeitslosenquote Vergleichsstudie 2007 und 2016	27
8.	Kaufkraft 2010 und 2019	28
9.	Ausländerquote 2010 und 2019	30
10.	LQ Ausländerquote Vergleichsstudie 2007 und 2016	30
11.	Verlauf der Berliner Mauer	32
12.	Elbow Plot Cluster	33
13.	Clustering 2015	35
14.	Zweitstimmen Landtagswahl 2016	36
15.	Cluster über die Zeit	37
16.	Residuen zur Regression der Miete über Zeit	41
17.	Brennpunkte 2015	42
18.	Modell Regression Brennpunkte	44
19.	Modell Regression Cluster	46
20.	Modell Regression LQ	49
21.	Residuen Regression Cluster	49
22.	Restliche Variablen Sozialdatensatz auf Karte 2015	60
23.	Residuen Regression Brennpunkte	60
24.	Residuen Regression Cluster	61

Tabellenverzeichnis

1.	Modifikationen der Rohvariablen	16
2.	Summary Statistics des Sozialdatensatzes	16
3.	Probleme im Sozialdatensatz	19
4.	Korrelationsmatrix Sozialdatensatz	20
5.	Summary Statistics Mietdatensatz	21
6.	Mietdatensatz über die Jahre	23
7.	Interpretation Cluster 2015	33
8.	Regression Miete über Zeit	40
9.	Regression Miete in Brennpunkten	43
10.	Regression Cluster	45
11.	Regression LQ	48

1. Einleitung

Diese Arbeit widmet sich der Untersuchung des Berliner Wohnungsmarktes, einem Brennpunkt sozioökonomischer Dynamiken und städtischer Entwicklungen, um die komplexen Wechselwirkungen zwischen Wohnraumnachfrage, Mietpreisentwicklungen und deren Auswirkungen auf die städtische Bevölkerung zu analysieren. Im Folgenden wollen wir kurz auf die Problemstellung und die sich daraus ableitenden Fragen eingehen sowie das weitere Vorgehen skizzieren.

1.1. Problemstellung

Die Nachfrage nach Wohnraum in den größten Städten Deutschlands spitzt sich immer weiter zu. Dies ist einerseits auf die starke Wirtschaftsdynamik zurückzuführen, die vermehrt Menschen aus dem In- und Ausland auf der Suche nach Beschäftigung anzieht. Andererseits wird dies durch das ausgeprägte Kultur- und Freizeitangebot begünstigt, das insbesondere Menschen aus ländlichen Regionen in die Großstädte zieht. Als Paradebeispiel des strukturellen Wandels wird häufig Berlin genannt. Anfang der 2000er Jahre hatte die Stadt zu weiten Teilen den Ruf einer armen Stadt, heute ist sie als Start-up-Zentrum Deutschlands bekannt. Als bevölkerungsreichste Großstadt Deutschlands verzeichnete Berlin in dem Zeitraum von 2012 bis 2022 ein Bevölkerungswachstum von 11,25% (Statista 2023), während die Bevölkerung im gesamten Bundesgebiet um 4,76% (Statistisches Bundesamt 2023) gestiegen ist. Der starke Zuzug führt zu einer Verknappung des Wohnraums und lässt im Umkehrschluss die Preise für Mieten und Eigentum weiter ansteigen. Dies wird insbesondere bei der Entwicklung der durchschnittlichen Angebotsmieten für Wohnungen in Berlin ersichtlich, die im Zeitraum von 2012 bis 2022 von 7,08 €/m² um rd. 70% auf 12,05 €/m² gestiegen sind (empirica 2023). Angesichts dieses rasanten Anstiegs stellt sich die Frage, wie sich derartige Veränderungen auf die Leistbarkeit von Wohnraum der Berliner Bevölkerung auswirken und welche weiteren strukturellen Folgen damit einhergehen.

Um zu überprüfen, wie sich eine steigende Mietentwicklung auf die Leistbarkeit der Bevölkerung auswirkt, ist sie im Verhältnis zur Einkommensentwicklung zu betrachten. Für diese Betrachtung wird die Kennzahl der Mietbelastung herangezogen, die sich auf den Anteil des Einkommens einer Haushaltseinheit bezieht, der für Mietkosten aufgebracht wird. Sie lag in Berlin im Jahr 2022 bei durchschnittlich 27,2%, was somit leicht unter dem bundesweiten Durchschnitt von 27,8% liegt. Auf den ersten Blick vermittelt diese Kennzahl zwar einen positiven Eindruck, jedoch erfordert sie eine differenzierte Betrachtung. Zum einen fließen alte und neue Mietverträge in die Statistik mit ein. Alte Mietverträge wurden auf einem deutlich geringeren Niveau abgeschlossen und können aufgrund der gesetzlichen Regelungen zu Mietpreiserhöhungen nur in einem geregelten Maß angehoben werden. Somit wird aus der Kennzahl nicht ersichtlich, wie es frisch Zugezogenen oder innerhalb Berlins umgezogenen Haushalten ergeht. Weiterhin weisen verschiedene Bevölkerungsgruppen unterschiedliche Mietbelastungsquoten auf. Während vor allem Alleinerziehende und armutsgefährdete Haushalte von überdurchschnittlichen Mietbelastungsquoten betroffen sind, fal-

len die Werte für Haushalte mit hohen Nettoeinkommen signifikant geringer aus (Lebuhn u. a. 2017). Dieser Unterschied macht sich auf dem Berliner Wohnungsmarkt vor allem durch die Gentrifizierung der zentralen Bezirke wie Kreuzberg und Friedrichshain bemerkbar. Die Aufwertung eines Stadtteils und der damit verbundene Anstieg des Mietniveaus setzt einen Segregationsprozess in Gang, wodurch einkommensschwächere Haushalte verdrängt und durch wohlhabendere Haushalte ersetzt werden. Bei diesem Fall handelt es sich um erzwungene Segregation, die dann vorliegt, wenn sich einkommensschwächere Gruppen in immer weniger Quartieren der Stadt mit Wohnraum versorgen können. Dem gegenüber steht die freiwillige Segregation, die auf individuellen Präferenzen beruht und sich durch eine Ansiedlung in der Nähe zu sozialen oder kulturell ähnlichen Bevölkerungsgruppen auszeichnet.

Als mögliche Folge der fehlenden sozialen Durchmischung kommt es zu einer Konzentration von Bevölkerungsgruppen mit ähnlichen sozioökonomischen Eigenschaften, wie beispielsweise dem Einkommensniveau oder dem Beschäftigungsgrad, in bestimmten Stadtgebieten. Dies steht im Widerspruch zu dem Gedanken, dass eine soziale Mischung als Garant für die Stabilität von Quartieren erachtet wird. Die dadurch begünstigte Konzentration von sozialen Problemlagen kann durch die individuell oder kollektiv als benachteiligend wahrgenommenen Lebensumstände weiter verstärkt werden und einen Nährboden für Konfliktpotenzial bilden. Auf der anderen Seite könnten Konfliktsignale von der wohlhabenderen Bevölkerung als Gefahr wahrgenommen werden und in einer weiteren Abschottung dieser Bevölkerungsgruppe resultieren, wie es beispielsweise schon in den USA mit den „gated communities“ praktiziert wird (Dohnke und Seidel-Schulze 2013).

Aus diesem Grund rückt die Herstellung gleichwertiger Lebensverhältnisse zunehmend in das Zentrum der Politik, insbesondere vor dem Hintergrund der aktuellen Debatte um die Integration von Flüchtlingen. Um ein weiteres soziales Auseinanderdriften zu verhindern, besteht somit das Erfordernis, Konzentrationen von Bevölkerungsgruppen frühzeitig zu erkennen und die Treiber der räumlichen Segregation zu untersuchen.

1.2. Zielsetzung

Deshalb verfolgt die vorliegende Forschungsarbeit das Ziel, am Beispiel des Berliner Wohnungsmarktes die Auswirkungen von sozioökonomischen Faktoren und Mietentwicklungen auf die räumliche Segregation verschiedener Bevölkerungsgruppen zu erforschen. Dem Ziel wurde die folgende Leitfrage zugrunde gelegt:

Wie beeinflussen sozioökonomische Faktoren und Mietentwicklungen die räumliche Segregation verschiedener Bevölkerungsgruppen in Berlin?

Um das Ziel zu erreichen und die Leitfrage zu beantworten, wurden die folgenden Unterfragen definiert:

- Wie verteilen sich Bevölkerungsgruppen in Berlin? Können Entwicklungen in der Verteilung festgestellt werden?
- Können sinnvolle Cluster anhand von (wenigen) sozioökonomischen Variablen gebildet werden? Sind diese Cluster über die Zeit stabil?
- Wie hat sich das Verhältnis von Einkommen zu Miete in Berlin während des Zeitraums von 2009 bis 2019 entwickelt und auf welche spezifischen Bevölkerungsgruppen hat sich diese Entwicklung positiv respektive negativ ausgewirkt?
- Welche Einflussfaktoren sind die wesentlichen Treiber der räumlichen Segregation?

Die gewonnenen Erkenntnisse der Forschungsarbeit könnten Stadtpolitikern, Stadtplanern und sozialen Organisationen dienlich sein, um fundierte Entscheidungen zu treffen und gezielte Maßnahmen für bestehende Viertel als auch für Neubaugebiete zu entwickeln.

1.3. Aufbau der Arbeit

Um eine fundierte Analyse der räumlichen Segregation in Berlin zu ermöglichen, ist es zunächst essenziell, sich eingehend mit dem Thema des Immobilienmarktes und der sozioökonomischen Einflussfaktoren, die diesen beeinflussen, vertraut zu machen. Dieser Schritt beinhaltet nicht nur das Verschaffen eines Überblicks über die genannten Grundlagen, sondern ermöglicht es ebenfalls, die zu analysierenden Parameter zu bestimmen, auf welchen sich die Analysen stützen.

Der Kern der Arbeit konzentriert sich auf die quantitative Analyse der Daten, wobei ein besonderes Augenmerk auf die Herausarbeitung der für die Forschungsfragen relevanten Informationen gelegt wird. Nach der Beschreibung der Daten und dessen Aufbereitung werden diese mittels spezifischer statistischer Methoden analysiert, um Zusammenhänge hinsichtlich der Forschungsfrage klar herauszuarbeiten.

Die Anwendung von Cluster- und Regressionsanalysen spielt eine zentrale Rolle bei der Schaffung einer aussagekräftigen Basis für Beobachtungen und das Ziehen von Schlussfolgerungen. Durch diese statistischen Methoden wird es möglich, tiefergehende Einblicke in die Strukturen und Dynamiken des Berliner Wohnungsmarktes zu gewinnen und zu verstehen, wie verschiedene Faktoren interagieren und welche Auswirkungen sie auf die Wohnsituation verschiedener Bevölkerungsgruppen - und somit auf die räumliche Segregation - haben.

2. Grundlagen

In diesem Kapitel wird der Wohnungsmarkt in Berlin untersucht, wie er sich im Laufe der Zeit entwickelt hat und welche Rolle sozioökonomische Faktoren und demografische Veränderungen dabei spielen. Betrachtet werden dabei, wie diese Faktoren zusammenwirken und den Wohnraum in der Stadt beeinflussen, insbesondere im Hinblick auf die Segregation verschiedener Bevölkerungsgruppen.

2.1. Wohnungsmarkt allgemein

In Deutschland reflektiert der Wohnungsmarkt signifikante Entwicklungen, die wichtige Indikatoren für die aktuellen wirtschaftlichen und sozialen Dynamiken darstellen. Ein Schlüsselmaß für das Verständnis der Trends auf dem deutschen Immobilienmarkt ist die Analyse der erteilten Baugenehmigungen für Wohnungen. Diese sind nicht nur ein Spiegelbild der gegenwärtigen Marktbedingungen, sondern geben auch Auskunft über zukünftige Bauvorhaben und somit über die prospektive Verfügbarkeit von Wohnraum. Die vorliegende Abbildung 1 bietet einen umfassenden Überblick über die monatlichen Baugenehmigungen für Wohn- und Nichtwohngebäude von 2010 bis 2024, dargestellt in Tausend (Statistisches Bundesamt 2024c). Die Originalwerte der Genehmigungen werden durch die blaue Linie visualisiert, während die rote Linie die Trend-Konjunktur-Komponente nach dem Berliner Verfahren abbildet, was eine Glättung der monatlichen Schwankungen und Hervorhebung langfristiger Tendenzen ermöglicht.

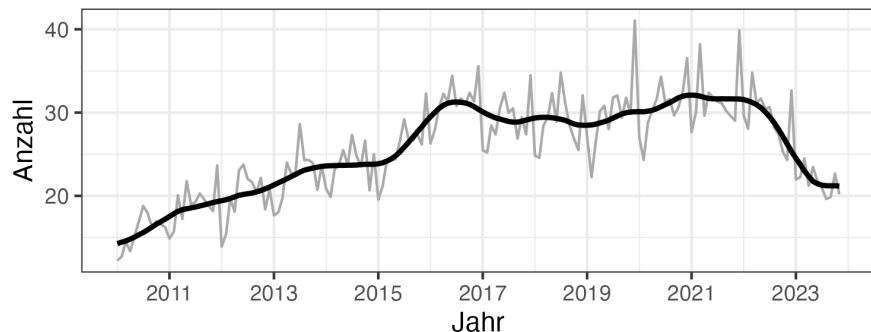


Abbildung 1: Monatlich genehmigte Wohnungen in Deutschland

Die Grafik veranschaulicht, dass die Anzahl der Baugenehmigungen im Laufe der Zeit Schwankungen unterliegt, mit einem erkennbaren Trend des Anstiegs bis etwa zum Jahr 2016. Nach diesem Höhepunkt zeigten die Zahlen Schwankungen, blieben aber im Mittel konstant, was auf eine gewisse Stabilität des Marktes auch während der Herausforderungen der Corona-Pandemie hindeutet. Ab dem Jahr 2022 ist jedoch ein deutlicher Abwärtstrend zu erkennen. Dieser Rückgang lässt sich insbesondere auf stark gestiegene Baukosten und verschlechterte Finanzierungskonditionen zurückführen (Statistisches Bundesamt 2024c).

Es ist wichtig zu beachten, dass diese Daten teilweise auf Schätzungen beruhen, insbesondere für Niedersachsen, für das im Berichtsmonat November 2023 keine Ergebnisse vorlagen. Für die übrigen Bundesländer wurden die Daten übermittelt. Diese methodischen Einschränkungen sollten bei der Interpretation der Daten und der Entwicklung von Prognosen berücksichtigt werden. Dennoch kann ein deutlicher Trend, besonders in den letzten Quartalen, erkannt werden.

2.2. Wohnungsmarkt Berlin

Diese gesunkene Baukapazität wird besonders in Metropolregionen wie Berlin spürbar, wo die Binnenwanderung und die sinkende Haushaltsgröße eine stetig wachsende Nachfrage nach neuem Wohnraum schaffen (Arbeitskreis der Oberen Gutachterausschüsse, Zentralen Geschäftsstellen und Gutachterausschüsse in der Bundesrepublik Deutschland (AK OGA) 2023). Ein Phänomen, das sich unweigerlich in den Mietpreisen widerspiegelt, die in der Hauptstadt zunehmend ansteigen. Dieser Trend zur Urbanisierung, gekoppelt mit dem Zuzug von außerhalb und dem Wunsch nach mehr Wohnfläche pro Person, setzt den Wohnungsmarkt in Berlin unter Druck (Statistik Berlin Brandenburg 2022). Der gegenwärtige Abschwung in der Baukonjunktur und die dadurch bedingte niedrigere Baukapazität könnten die Situation verschärfen, da sie potenziell nicht ausreichen, um den hohen Wohnungsbedarf zu decken.

Diese Sorge um eine mittelfristig niedrigere Baukapazität wiegt schwer vor dem Hintergrund des anhaltend hohen Wohnungsbedarfs in Berlin. Das Bild wird noch komplexer, wenn man bedenkt, dass die genutzte Wohnfläche pro Person aus verschiedenen Gründen gestiegen ist und wahrscheinlich weiter steigen wird. Zusätzlich beeinflusst die Nettozuwanderung nach Deutschland und der Trend zur urbanen Verdichtung den Bedarf an Wohnraum in der Stadt (Jonas, Martin und Theobald 2023). All diese Faktoren führen zu einer erhöhten Nachfrage, die sich direkt auf die Mieten auswirkt und diese in die Höhe treibt.

Die Entwicklung der Mieten in einer dynamischen Stadt wie Berlin ist nicht nur ein ökonomischer Indikator, sondern spiegelt auch tiefgreifende soziale Prozesse wider. Bei Berlin handelt es sich um eine Großstadt die zu 85% aus Mietwohnungen besteht, wohingegen etwa ein Sechstel der Wohnungen im Bestand der Stadt sind (Marcinczak und Bernt 2021). Demnach können steigende Mieten zu einer Verstärkung der sozialen Segregation führen, indem sie einkommensschwächere Bevölkerungsgruppen aus innerstädtischen Gebieten verdrängen und somit die soziale Durchmischung verringern. Diese Marktreaktion auf Angebot und Nachfrage beeinflusst unmittelbar die Lebensqualität und soziale Struktur der Stadt und kann langfristige Auswirkungen auf das soziale Gefüge haben.

In Berlin zeigt sich, dass Quartiere mit einer zuvor hohen Durchmischung zunehmend homogener werden, was sowohl Chancen als auch Herausforderungen mit sich bringt. Während einige Gebiete aufgewertet werden und die Lebensqualität dort steigt, können andere unter einer Vernachlässigung leiden. Dieses Phänomen der „Gentrifizierung“ ist besonders in Städten mit einem

angespannten Wohnungsmarkt wie Berlin zu beobachten, wo die Nachfrage das Angebot übersteigt und die Preise in die Höhe treibt.

Um die sozialen Auswirkungen der Mietpreisentwicklung in Berlin zu verstehen, ist es daher essenziell, über die bloße Betrachtung der baulichen Aktivität und Mietpreise hinauszugehen. Es bedarf einer tiefgehenden Analyse der Mietdynamik in Verbindung mit sozialen Indikatoren wie Einkommen, Demografie und Zugang zu sozialer Infrastruktur. Nur so kann ein umfassendes Bild der Wohnsituation in Berlin gezeichnet werden, das die Risiken der sozialen Segregation offenlegt und Ansätze zur Förderung einer ausgewogenen sozialen Entwicklung bietet.

2.3. Demografie Berlin

In der Auseinandersetzung mit dem demografischen Wandel innerhalb einer städtischen Bevölkerung zeigt sich, dass die Entwicklung der einheimischen Bevölkerung stark den nationalen Trends ähnelt. Jedoch offenbart eine detailliertere Betrachtung signifikante Unterschiede, insbesondere im Hinblick auf das Durchschnittsalter und die Geburtenrate. Während das Durchschnittsalter der deutschen Bevölkerung bei 43 Jahren liegt, weist die Gruppe der Einwanderer ein jüngeres Durchschnittsalter von 35 Jahren auf. Diese jüngere Altersstruktur unter den Einwanderern spiegelt sich auch in einer höheren Geburtenrate wider (Ulrich Raiser 2005).

Interessanterweise zeigt sich, dass der Anteil der Bevölkerung im erwerbsfähigen Alter (21 bis 45 Jahre) unter den Nicht-Deutschen bei 53,8 Prozent liegt, was die jugendliche Demografie dieser Gruppe unterstreicht. Im Gegensatz dazu bilden Einwanderer lediglich 13,2 Prozent der Gesamtbevölkerung Berlins. Diese Zahlen verdeutlichen die demografische Dynamik innerhalb der Stadt, in der ein relativ geringer Anteil von Einwanderern eine deutlich jüngere Altersstruktur aufweist als die einheimische Bevölkerung (Ulrich Raiser 2005).

Des Weiteren ist der Anteil der über 65-Jährigen unter der deutschen Bevölkerung mit 16,2 Prozent signifikant höher als der entsprechende Anteil unter den Einwanderern, der nur bei 5,3 Prozent liegt. Diese Statistiken beleuchten die demografischen Unterschiede zwischen einheimischen Bürgern und Einwanderern, die für soziale, wirtschaftliche und politische Planungen innerhalb der Stadt von großer Bedeutung sind. Sie unterstreichen die Notwendigkeit einer differenzierten Betrachtung der Bevölkerungsstruktur bei der Entwicklung von Maßnahmen zur Integration und Förderung der gesellschaftlichen Teilhabe aller Bevölkerungsgruppen (Budlinger 2021).

Die demografische Entwicklung und Geschichte der Einwanderung in Berlin ist eng mit den politischen und wirtschaftlichen Bedingungen der geteilten Stadt verknüpft. In der Zeit, als Berlin noch in Ost und West geteilt war, beeinflusste die jeweilige Wirtschaftspolitik maßgeblich die Einwanderungspolitik. Die westberliner Regierung sah sich aufgrund wirtschaftlicher Notwendigkeiten gezwungen, Arbeitskräfte aus Ländern wie der Türkei oder Jugoslawien zu rekrutieren. Diese Entscheidung führte dazu, dass die Verwaltungspolitik West-Berlins zu einem Drehkreuz für die Ansiedlung von Einwanderern wurde und somit eine wesentliche Rolle für die Bildung der

noch heute erkennbaren ethnischen Konzentrationsgebiete spielte. Insbesondere Stadtteile wie Kreuzberg, die unter Gebäudeleerstand litten, mussten bei der Auswahl von Mietern Kompromisse eingehen, was die Vielfalt und Multikulturalität der Bevölkerung weiter förderte (Marcinczak und Bernt 2021).

Im Gegensatz dazu war die Einwanderungspolitik in Ost-Berlin eher vernachlässigbar. Die wenigen, die ins Land kamen, wurden in der Regel nach Beendigung ihrer Tätigkeit wieder in ihre Heimatländer zurückgeschickt. Diese Politik führte dazu, dass Ost-Berlin auch nach dem Fall der Berliner Mauer ethnisch weniger durchmischt blieb (Marcinczak und Bernt 2021).

Ein weiterer Aspekt der Einwanderungsgeschichte Berlins betrifft die langfristigen Auswirkungen der Einwanderung auf die soziale Struktur der Stadt. Einwanderer, die längere Zeit in ihrem Gastland leben, sind tendenziell weniger von Segregation betroffen und erfahren häufiger Assimilation, was eine soziale und wirtschaftliche Aufwärtsmobilität impliziert. Diese Mobilität hängt von verschiedenen Faktoren ab, darunter Einkommen, Bildungsniveau und Dauer des Aufenthalts. Es gibt jedoch auch eine Gegenbewegung zu diesem Trend. Einige Einwanderer bevorzugen es, innerhalb ihrer ethnischen Gemeinschaften zu bleiben, selbst wenn sie sozial aufsteigen. Sie sehen in der ethnischen Infrastruktur Vorteile wie besseren Zugang zu Wohnraum und Beschäftigungsmöglichkeiten, was die Komplexität der Einwanderungsdynamik in einer Stadt wie Berlin unterstreicht (Marcinczak und Bernt 2021).

2.4. Charakter Berlins

Berlin präsentiert sich als eine Stadt mit einzigartigen Charakteristika, nicht zuletzt aufgrund seiner polyzentrischen Struktur, die maßgeblich durch die zuvor erwähnte historische Teilung in Ost- und West-Berlin geprägt wurde. Diese Teilung hat tiefgreifende und langanhaltende Spuren in der städtischen Landschaft hinterlassen, welche auch heute noch in Form von Segregation sichtbar sind (Marcinczak und Bernt 2021).

Die historische Trennung Berlins hat zu einer Stadtentwicklung geführt, in der sich sozioökonomische und kulturelle Unterschiede zwischen den verschiedenen Bezirken manifestiert haben. Die Bezirke im Osten und Westen der Stadt weisen jeweils eigene, deutlich unterscheidbare Charakteristika auf, was sich in der Architektur, Infrastruktur und im sozialen Gefüge widerspiegelt. Während einige Bezirke eine dynamische Entwicklung erfahren und von Modernisierung und Gentrifizierung geprägt sind, kämpfen andere mit sozialen Herausforderungen und einer stagnierenden Wirtschaft (Einem 2015).

Zusammenfassend lässt sich festhalten, dass diese anhaltende Segregation in Berlin nicht nur ein räumliches, sondern auch ein soziales Phänomen ist. Die Trennlinien zwischen verschiedenen sozialen Gruppen werden demnach verstärkt durch Unterschiede in Einkommen und Bildungsniveau, Zugang zu sozialen Ressourcen und politischen Entscheidungen. Die räumliche Nähe zu sozialer Infrastruktur wie Schulen, Gesundheitsdiensten und öffentlichen Verkehrsmitteln variiert

stark zwischen den Bezirken und verstärkt dabei bestehende soziale Ungleichheiten (Krajewski 2015).

2.5. Auswahl der Variablen

Um die räumliche Segregation präzise erfassen und analysieren zu können, ist die sorgfältige Auswahl adäquater Variablen, die als Grundlage für die Untersuchung dienen, von entscheidender Bedeutung. Die Auswahl der Variablen für die Analyse räumlicher Segregation in Berlin orientiert sich an der Notwendigkeit, die beiden zentralen Aspekte der Forschungsfrage – sozioökonomische Faktoren und Mietentwicklungen – zu erfassen und deren Einfluss auf die Verteilung verschiedener Bevölkerungsgruppen zu messen.

Für den Bereich der Mietentwicklungen wurden die Variablen Kaltmiete in Euro und Kaltmiete pro Quadratmeter als direkte finanzielle Indikatoren gewählt. Sie bieten einen quantitativen Maßstab für das Wohnkostenniveau und erlauben es, sowohl absolute als auch flächenbezogene Mietpreise zu berücksichtigen. Die zusätzliche Aufnahme der Wohnfläche als Variable ermöglicht es, den Effekt von Wohnungszuschnitten und -größen auf die Segregation zu analysieren, da größere oder kleinere Wohnungen unterschiedliche Bevölkerungsgruppen anziehen können.

Im Hinblick auf sozioökonomische Faktoren wurde eine Reihe von Variablen ausgewählt, die signifikante Einflüsse auf das Mietpreisniveau und die Wohnpräferenzen erwarten lassen. Die Ausländerquote, der Anteil der Bevölkerung über 60 Jahre und die durchschnittliche Kaufkraft wurden aufgrund ihrer vermuteten starken Korrelation mit Mietpreisen und Wohnraumnachfrage ausgewählt. Diese Variablen bieten Einblicke in demografische Strukturen und wirtschaftliche Kapazitäten, die für die Segregation von Belang sind. Des Weiteren sind Zusammenhänge dieser Variablen im Hinblick auf Segregation bereits erforscht und in ihrer Relevanz bestätigt (Einem 2015).

Die Arbeitslosenquote ist besonders in Berlin von Interesse, wo sie über dem Bundesdurchschnitt liegt. Eine hohe Arbeitslosenquote kann sozioökonomische Disparitäten verstärken und hat potenziell einen starken Einfluss auf die Segregation innerhalb der Stadt.

Die Variable Oberklassewagen wurde als Proxy für finanzielle Stärke und damit zusammenhängende hohe Mietpreise in bestimmten Vierteln aufgenommen. Es wird angenommen, dass ein höherer Anteil an Oberklassewagen in einem Gebiet auf eine gehobene sozioökonomische Schichtung hinweist, die wiederum Einfluss auf das Mietpreisgefüge haben kann.

Die Variable Kreditrisiko wurde in die Analyse aufgenommen, um finanzielle Stabilität und das potenzielle Ausfallrisiko der Bewohner in den verschiedenen Stadtteilen abzubilden. Diese Größe ermöglicht es, die wirtschaftliche Vulnerabilität in Beziehung zur räumlichen Segregation zu setzen und kann Hinweise auf sozioökonomisch bedingte Wohnortpräferenzen liefern.

Um die räumliche Komponente der Segregation zu adressieren, wurden die Variablen für Wohngebäudetypen – Einfamilienhäuser (EFH), Mehrfamilienhäuser (MFH) und Wohnblöcke – hinzugefügt. Diese Kategorien sollen dazu beitragen, die Segregation im Hinblick auf unterschiedliche Wohnformen zu verstehen und deren Beziehung zu den Mietpreisen zu klären.

Die Variable Haushalte wurde aus einem eher technischen Hintergrund in die Untersuchung integriert, da sie als normative Größe für die Berechnung weiterer Variablen, wie beispielsweise der Kaufkraft (pro Haushalt), dient.

Insgesamt soll diese sorgfältige Auswahl und Begründung der Variablen sicherstellen, dass die Forschungsarbeit ein umfassendes und differenziertes Bild der räumlichen Segregation und ihrer Determinanten in Berlin zeichnet.

3. Daten

Bevor in das Datenkapitel und die Methodik unserer Untersuchung eingestiegen wird, ist es von essenzieller Bedeutung, den Prozess der Datenbereinigung und die spezifische Verwendung der Variablen zu erläutern. Die Analyse stützt sich auf zwei umfangreiche Datensätze, einen Sozialdatensatz und einen Mietdatensatz, die zusammen über 200 Variablen umfassen. Angesichts dieser Vielfalt war eine sorgfältige Vorarbeit unerlässlich, um die für unsere Forschungsfrage relevanten Daten herauszufiltern und entsprechend aufzubereiten.

3.1. Datenaufbereitung

In Anlehnung an die Erkenntnisse aus der Literaturanalyse, die im vorherigen Kapitel dargestellt wurden, haben wir bestimmte Variablen identifiziert, die im Zusammenhang mit Mietentwicklungen und sozialräumlicher Segregation von besonderer Bedeutung sind. Diese Variablen und ihre verschiedenen Ausprägungen werden im Folgenden eingehender betrachtet.

Alle statistischen Analysen führen wir mit R (R Core Team 2023) durch. Neben den Basisfunktionen verwenden wir folgende Packages (ohne spezielle Reihenfolge):

<code>Hmisc</code>	Erstellen der Korrelationsmatrix
<code>stargazer</code>	Erstellen von Regressionstabellen und Export nach L ^A T _E X
<code>kableExtra</code>	Exportieren von Data-Frames als L ^A T _E X Tabellen
<code>vtable</code>	Erstellen von Summary-Statistiken und Export als L ^A T _E X Tabelle
<code>ggpubr</code>	Zusätzliche Layout-Optionen für <code>ggplot2</code> Plots
<code>ggrepel</code>	Erweiterung für <code>ggplot2</code> zur Verbesserung der Beschriftung
<code>gridExtra</code>	Zusätzliche Layout-Optionen für <code>ggplot2</code> Plots
<code>vroom</code>	Schnelleres Einlesen von Daten im Vergleich zu <code>read.csv</code>
<code>tidyverse</code>	Sammlung verschiedener Packages für Datenmanipulation und Visualisierung, insbesondere <code>ggplot2</code> für das erstellen von Plots sowie <code>dplyr</code> und <code>tidyverse</code> für Datentransformationen
<code>sf</code>	Arbeiten mit Geodaten und räumlichen Objekten

Zum Verfahren der Datenfilterung: Zunächst galt es zu ermitteln, welche Quadrate des Inspire-Grids, einem 1x1km Raster, das gesamte Bundesgebiet abdeckt, innerhalb der administrativen Grenzen Berlins liegen. Dies war insbesondere für den Sozialdatensatz relevant, da dieser Daten für ganz Deutschland enthält, die noch nicht spezifisch für Berlin gefiltert waren. Aus dieser Überprüfung wurde eine eigene Liste mit den `r1_ID` Koordinaten erstellt, um den Sozialdatensatz entsprechend zu filtern. Die `r1_ID` stellt dabei eine Variable dar, welche die X- und Y-Koordinaten der jeweiligen Quadrate des Grids in einer Variable zusammenfassen. Die Grenzen des Berliner Gebiets basieren dabei auf den Daten von Geoportal Berlin (2017) und das Inspire-Grid auf den GeoGittern des Bundesamt für Kartographie und Geodäsie (2017). Abbildung 2 illustriert dabei

das GeoGitter innerhalb der Berliner Stadtgrenzen. In Rot ist der ehemalige Verlauf der Berliner Mauer mit den Daten von Geoportal Berlin (2018) eingezeichnet.

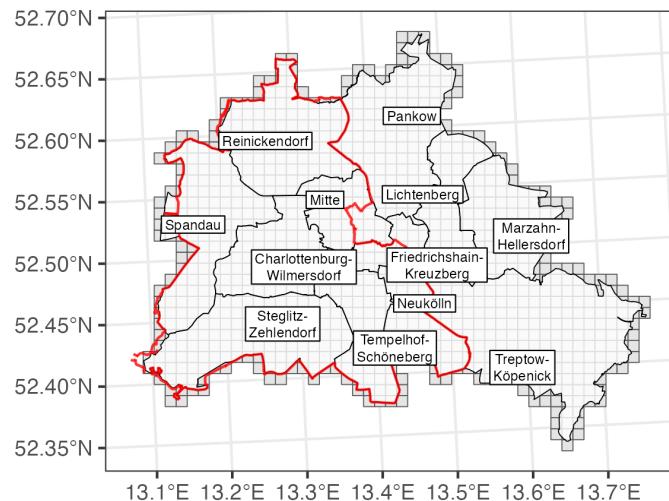


Abbildung 2: Karte Berlin

Im nächsten Schritt werden die Variablen aus den Datensätzen konstruiert, die für die zuvor erarbeiteten Merkmale für die Betrachtung der räumlichen Segregation – und somit für die Beantwortung der Forschungsfrage – essenziell sind.

Dabei wurden verschiedene bestehende Variablen aus dem Sozialdatensatz teilweise aggregiert, zusammengeführt oder neu skaliert, um eine bessere und verständlichere Vorgehensweise der Analyse zu gewährleisten. Bei dem Mietdatensatz Datensatz wurde, bis auf die Umbenennung, um eine klare und eindeutige Identifikation der relevanten Daten zu gewährleisten, keine Modifikation der Variablen vorgenommen. Tabelle 1 gibt einen Überblick, wie die verwendeten Variablen im Vergleich zum Rohdatensatz modifiziert wurden.¹

¹Ein höherer Wert beim Kreditrisiko weist dabei auf eine geringere Bonität bzw. ein höheres Ausfallrisiko hin

Neue Variable	Modifikation / Zusammensetzung
60+ (%)	Aggregierung von 6 Einzelvariablen von m/w-Personen über 60 zu einer Sammelvariablen (%)
Ausländer (%)	100 – Deutsche (%)
EFH (%)	Summe: Haustyp 1 und 2 (%)
MFH (%)	Summe: Haustyp 3 und 4 (%)
Oberklassewagen (%)	Summe: Ober- und Obermittelklassewagen (%)
Wohnblock (%)	Summe: Haustyp 5 und 6 (%)
Haushalte (1000)	Anzahl Haushalte/1000
Arbeitslose (%)	Keine Anpassung
Kaufkraft (1000€)	Kaufkraft gesamt/Anzahl Haushalte
Kreditrisiko	Aggregierung von 9 Einzelvariablen zu einer Variablen von 0 bis 1

Tabelle 1: Modifikationen der Rohvariablen

3.2. Explorative Datenanalyse - Sozialdatensatz

	n	mean	Std.Dev.	min	Pctl.25	Pctl.50	Pctl.75	max
60+ (%)	11205	28.00	6.81	7.77	23.66	28.54	32.75	100.00
Ausländer (%)	11205	10.09	10.18	0.00	3.31	6.10	13.38	74.83
EFH (%)	11205	36.82	31.47	0.00	6.23	30.81	63.16	100.00
MFH (%)	11205	26.81	17.94	0.00	12.24	25.27	39.00	100.00
Oberklassewagen (%)	9319	5.25	2.49	0.00	3.56	4.80	6.38	26.26
Wohnblock (%)	11205	32.67	30.83	0.00	2.98	24.91	57.93	100.00
Haushalte (1000)	10487	2.24	2.57	0.00	0.41	1.30	3.16	14.69
Arbeitslose (%)	11205	8.54	5.14	0.00	4.49	8.15	11.95	26.58
Kaufkraft (1000€)	10452	40.10	8.69	18.74	33.64	38.95	45.70	79.85
Kreditrisiko	11205	0.60	0.20	0.11	0.43	0.59	0.78	1.00

Tabelle 2: Summary Statistics des Sozialdatensatzes

Tabelle 2 präsentiert eine deskriptive statistische Analyse² von verschiedenen sozioökonomischen Indikatoren des Sozialdatensatzes, die potenziell Einfluss auf die räumliche Segregation in Berlin haben könnten. Die Stichprobe umfasst Daten von 11.205 Beobachtungseinheiten, mit Ausnahme der Variable „Oberklassewagen“, welche Daten von 9.319 Einheiten beinhaltet, und „Haushalte (1000)“ bzw. „Kaufkraft (1000€)“, die auf 10.487 bzw. 10452 Einheiten basieren. Ein Grund

²In Tabellen und Abbildungen sowie in Kapitel 4.2 verwenden wir einen Dezimalpunkt statt eines Kommas. Dies ist im englischsprachigen Raum sowie dem von uns verwendeten Statistikprogramm R der Standard.

hierfür könnte sein, dass im Sozialdatensatz einige Beobachten zwischen den Jahren 2006 und 2008 fehlen.

Die Altersverteilung der Bevölkerung über 60 Jahre („60+“) zeigt einen Durchschnitt von 28% mit einer Standardabweichung von 6,81%, was auf eine moderate Streuung hinweist. Die minimale und maximale Präsenz dieser Altersgruppe in den beobachteten Einheiten variiert von 7,77% bis 100%, was auf eine erhebliche Heterogenität innerhalb der Stichprobe hindeutet.

Der Anteil der Ausländer („Ausländer“) hat einen mittleren Wert von 10,09% und eine hohe Standardabweichung von 10,18%, was auf eine sehr breite Streuung dieser Bevölkerungsgruppe hinweist. Die Spannweite zwischen dem Minimum (0%) und dem Maximum (74,83%) bestätigt dies.

Die Wohnsituation wird durch Variablen wie „EFH“ (Einfamilienhäuser), „MFH“ (Mehrfamilienhäuser) und „Wohnblock“ repräsentiert. Besonders „EFH“ zeigt mit 36,82% als Mittelwert und einer hohen Standardabweichung von 31,47% relativ hohe Werte, was eine ungleiche Verteilung von Einfamilienhäusern in der Stadt suggeriert. Im Gegensatz dazu hat „MFH“ einen niedrigeren Mittelwert von 26,81% und eine niedrigere Standardabweichung von 17,94%, was auf eine geringere Streuung hinweist.

Die Variable „Oberklassewagen“ könnte als Indikator für den Wohlstand in verschiedenen Bezirken dienen, mit einem Mittelwert von 5,25% und einer geringen Standardabweichung von 2,49%, was auf eine relativ homogene Verteilung von Luxusfahrzeugen schließen lässt.

Der ökonomische Status wird durch „Arbeitslose“, „Kaufkraft“ und „Haushalte“ gemessen. Der Arbeitslosenanteil liegt im Mittel bei 8,54%, mit einer Standardabweichung von 5,14%, was auf eine heterogene Verteilung von Arbeitslosigkeit hinweist. Die Kaufkraft, ausgedrückt in Tausend €, zeigt eine deutliche Variation mit einem Mittelwert von 40,10 und einer Standardabweichung von 8,69. Die Anzahl der Haushalte („Haushalte“) wird in Tausend angegeben und zeigt einen mittleren Wert von 2,24 mit einer Standardabweichung von 2,57, was eine beträchtliche Variabilität zwischen den beobachteten Quadranten des GeoGitters nahelegt. Da es sich um eine Großstadt handelt, ist von solchen inhomogenen Bevölkerungsverteilungen, besonders in zentrumsnahen Lagen, bereits auszugehen.

Das „Kreditrisiko“ wird als letzte Variable dargestellt, mit einem Durchschnittswert von 0,60 und einer niedrigen Standardabweichung von 0,20, was auf ein gemäßigtes Maß an Variabilität im Kreditrisiko innerhalb der Stichprobe hinweist.

Die Perzentile Pctl.25, Pctl.50 (Median) und Pctl.75 geben Aufschluss über die Verteilung der Daten. Beispielsweise liegt der Median der Altersgruppe über 60 Jahre bei 28,54%, was bedeutet, dass 50% der Beobachtungseinheiten einen Anteil von 60+ Personen unter diesem Wert aufweisen. Diese Perzentile sind besonders nützlich, um die Schiefe der Verteilung zu beurteilen.

Die in dieser Tabelle dargestellten Daten liefern einen umfassenden Überblick über die sozioöko-

nomischen Bedingungen, die zur räumlichen Segregation verschiedener Bevölkerungsgruppen in Berlin beitragen könnten. Es fallen allerdings einige ungewöhnliche Werte auf, die zusätzliche Überlegungen erfordern. Insbesondere die Existenz von Werten wie 0% als Minima bei vielen Variablen wie beispielsweise „EFH“ (Einfamilienhäuser) und „Ausländer“ deuten darauf hin, dass in einigen Rasterfeldern möglicherweise keine entsprechenden Bevölkerungsgruppen oder Wohnstrukturen gemeldet sind. Dies könnte darauf zurückzuführen sein, dass das zugrundeliegende Raster auch Flächen abdeckt, die nicht für Wohnzwecke genutzt werden, wie zum Beispiel Wälder, Seen und Parks.

Da die Datenerhebung auf Bewohnerangaben bzw. -meldungen in den jeweiligen Rasterfeldern basiert, spiegeln diese Nullwerte Gebiete wider, in denen keine Bewohner gemeldet sind oder beispielsweise keine Einfamilienhäuser vorhanden sind. Es ist wichtig, diese Besonderheit bei der Analyse und Interpretation der Daten zu berücksichtigen, da sie die Variabilität und Verteilung der gemessenen sozioökonomischen Indikatoren beeinflussen kann. In Gebieten, die für das Wohnen ungeeignet sind, würden demnach solche Nullwerte erwartungsgemäß auftreten.

Um die räumliche Verteilung dieser Variablen zu visualisieren, wurden die Verteilungen auf Basis des GeoGitters innerhalb der Berliner Stadtgrenzen dargestellt.³ Exemplarisch wurden dabei die Daten aus dem Jahr 2015 verwendet, da die räumliche Streuung über die Jahre hinweg weitestgehend unverändert bleibt.

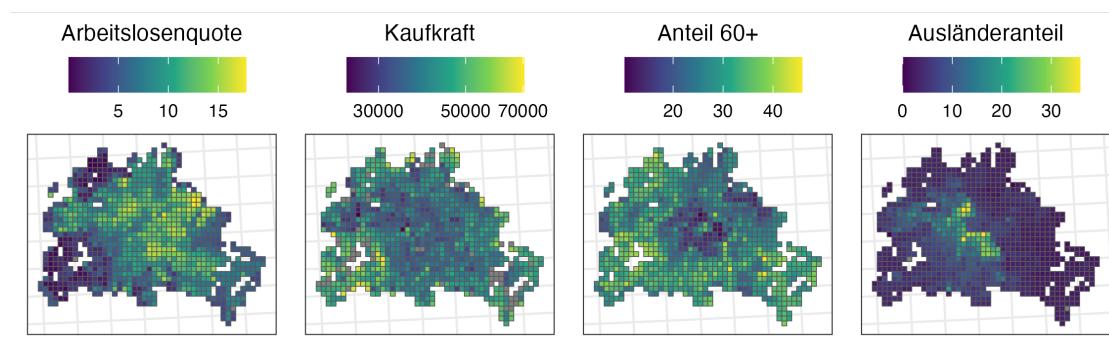


Abbildung 3: Sozialdatensatz auf Karte 2015

Bei näherer Betrachtung des Sozialdatensatzes treten bestimmte Auffälligkeiten in der Variable des Ausländeranteils auf, die einer detaillierten Untersuchung bedürfen.

Die in Tabelle 3 präsentierten Daten zeigen eine signifikante Diskrepanz im Ausländeranteil, die ein methodisches Problem im Sozialdatensatz nahelegt. Die Datenreihe zeigt für die Jahre bis einschließlich 2016 eine relativ konstante Quote des Ausländeranteils. Jedoch ist für das Jahr 2017 ein abrupter Anstieg zu verzeichnen, welcher sich in den darauffolgenden Jahren fortsetzt.

³Die Abbildungen der übrigen Variablen befinden sich im Anhang

Jahr	Ausländeranteil (%)		
	einfach	gewichtet	offiziell
2005	6.53	10.14	13.42
2009	6.51	10.09	14.00
2010	6.51	10.06	13.75
2011	6.50	10.04	13.65
2012	6.50	10.02	11.89
2013	6.45	9.96	12.63
2014	6.46	9.96	13.42
2015	6.45	9.95	14.31
2016	6.46	9.96	15.50
2017	20.91	28.79	16.74
2018	21.22	29.28	17.65
2019	22.93	31.87	18.53

Tabelle 3: Probleme im Sozialdatensatz

Dieser sprunghafte Anstieg steht im Gegensatz zu den offiziellen Daten (Statistisches Bundesamt 2024a), die zwar ebenfalls eine Erhöhung anzeigen, jedoch in einem geringeren Ausmaß.

Die räumliche Verteilung des Ausländeranteils in Berlin, wie sie durch den Datensatz abgebildet wird, bleibt trotz dieses Anstiegs jedoch weiterhin ähnlich. Dies deutet darauf hin, dass die zugrunde liegenden räumlichen Muster der Verteilung von ausländischen Bevölkerungsgruppen über die Zeit stabil geblieben sind, während das Gesamtniveau des Ausländeranteils sich verändert hat. Eine plausible Erklärung dafür könnte eine Anpassung oder Veränderung der Methodik in der Datenerfassung oder -aufbereitung seitens der datenerhebenden Stelle im Jahr 2017 sein. Solche methodischen Brüche können durch verschiedene Faktoren verursacht werden, wie etwa Änderungen in den Definitionen, demografischen Erfassungskriterien oder auch in der geografischen Abgrenzung der Daten.

Zusätzlich ist zu beachten, dass in dem betrachteten Zeitraum Daten für die Jahre 2006 bis 2008 fehlen, was die Kontinuität der Zeitreihe unterbricht und die Langzeitanalyse der Mietentwicklung und Segregation beeinträchtigt. Das Fehlen dieser Daten kann verschiedene Ursachen haben, wie etwa eine Unterbrechung in der Datenerhebung oder eine nachträgliche Korrektur, die zur Entfernung dieser Jahrgänge aus dem Datensatz führte.

	Haus-halte	Arbeits-lose	Oberkl. Wagen	Kauf-kraft	EFH	MFH	Wohn-block	Aus-länder	Kredit-risiko	60+
Haushalte	1.00									
Arbeitslose	0.40	1.00								
Oberkl.wagen	-0.02	-0.16	1.00							
Kaufkraft	-0.33	-0.71	0.19	1.00						
EFH	-0.63	-0.50	-0.01	0.42	1.00					
MFH	-0.20	-0.08	0.03	0.07	-0.31	1.00				
Wohnblock	0.74	0.56	-0.04	-0.44	-0.81	-0.22	1.00			
Ausländer	0.40	0.12	0.06	-0.10	-0.37	-0.10	0.43	1.00		
Kreditrisiko	0.56	0.59	-0.15	-0.52	-0.68	-0.04	0.70	0.45	1.00	
60+	-0.44	-0.48	0.18	0.39	0.30	0.28	-0.46	-0.23	-0.46	1.00

Tabelle 4: Korrelationsmatrix Sozialdatensatz

Die Korrelationsmatrix aus dem Sozialdatensatz enthüllt bedeutsame Beziehungen zwischen den gewählten sozioökonomischen Indikatoren, die Einblick in die strukturellen Zusammenhänge innerhalb der Bevölkerung Berlins geben. Im Folgenden fokussieren wir uns auf die stärksten und interessantesten Korrelationen, um ein tieferes Verständnis für die treibenden Kräfte der sozialräumlichen Dynamik zu gewinnen.

Die Korrelation zwischen der „Kaufkraft“ und der Variablen „Arbeitslose“ ist mit einem Wert von -0,71 stark negativ ausgeprägt. Dies deutet darauf hin, dass in Gebieten mit höherer Arbeitslosigkeit tendenziell eine niedrigere Kaufkraft vorhanden ist, was konsistent mit ökonomischen Theorien ist, die von einem direkten Zusammenhang zwischen Beschäftigung und Einkommen ausgehen.

Eine negative Korrelation ist auch zwischen „Arbeitslosen“ und „EFH“ (Einfamilienhäuser) zu erkennen (-0,50), was nahelegt, dass Regionen mit einer höheren Quote an Einfamilienhäusern tendenziell niedrigere Arbeitslosenraten aufweisen. Dies könnte auf eine höhere sozioökonomische Stellung der Bewohner in diesen Gebieten hindeuten.

Zugleich zeigt sich eine weitere deutliche Verbindung zwischen „Arbeitslosen“ und „Wohnblock“ mit einem Wert von 0,56. Dies lässt darauf schließen, dass in Stadtteilen mit einem höheren Anteil an Wohnblöcken eine höhere Arbeitslosigkeit vorherrscht, was typisch für dichter besiedelte, urbane und möglicherweise sozial schwächere Gebiete ist.

Das „Kreditrisiko“ korreliert ebenfalls positiv mit „Arbeitslosen“ (0,59), was bedeutet, dass in Gebieten mit einer höheren Arbeitslosenrate auch ein höheres Kreditrisiko besteht. Dies könnte auf eine größere finanzielle Unsicherheit in diesen Regionen hinweisen und bestärkt nochmals die starke negative Korrelation zwischen „Kaufkraft“ und der Variablen „Arbeitslose“.

Ebenso zeigt sich eine negative Korrelation zwischen „Kreditrisiko“ und „Kaufkraft“ (-0,52), was

die Annahme stützt, dass finanzielle Stabilität oft mit höherer Kaufkraft einhergeht.

Interessant ist auch die negative Korrelation von „Kreditrisiko“ mit „EFH“ (-0,68), was darauf hindeuten könnte, dass Besitzer von Einfamilienhäusern generell ein niedrigeres Kreditrisiko aufweisen, was wiederum auf stabilere wirtschaftliche Verhältnisse schließen lässt. Im Hinblick auf den aktuellen Immobilienmarkt, vor allem in Berlin, ist dies auch zu erwarten, da rapide Preissteigerungen in den letzten Jahren beobachtet werden können und somit oftmals nur finanziell starke Haushalte sich diese finanzieren können (Senatsverwaltung für Stadtentwicklung, Bauen und Wohnen 2023).

Parallel dazu ist eine starke positive Korrelation bei „Kreditrisiko“ und „Wohnblock“ zu beobachten (0,70). Dies legt nahe, dass in Gebieten mit einem höheren Anteil an Wohnblocks das Kreditrisiko tendenziell höher ist, was mit den sozioökonomischen Herausforderungen in dicht bebauten urbanen Gebieten korrespondieren könnte.

Die Altersgruppe „60+“ weist eine negative Korrelation mit „Kreditrisiko“ auf (-0,46), was darauf hindeutet, dass in Gebieten mit einem höheren Anteil älterer Menschen das Kreditrisiko geringer ist, möglicherweise aufgrund einer höheren finanziellen Sicherheit im Alter oder geringeren Verschuldung in dieser Bevölkerungsgruppe.

Betrachtet man dagegen die Korrelation der ausländischen Bevölkerungsgruppe mit dem „Kreditrisiko“ stellt man eine relativ starke positive Korrelation fest. Dies deutet an, dass in Gegenden mit einem höheren Anteil an ausländischer Bevölkerung auch ein erhöhtes Kreditrisiko besteht.

Die Korrelation zwischen „Ausländer“ und „Wohnblock“ dabei ist positiv (0,43), was auf eine stärkere Präsenz von Menschen mit Migrationshintergrund in Wohngebieten mit Wohnblöcken hinweist, was Fragen der Wohnsegregation und Integration aufwerfen könnte.

3.3. Explorative Datenanalyse - Mietdatensatz

	n	mean	Std.Dev.	min	Pctl.25	Pctl.50	Pctl.75	max
Kaltmiete (€)	1917043	627.08	516.80	0.01	346.00	467.86	706.54	15000
Kaltmiete (€/m ²)	1917043	8.50	115.04	0.00	5.59	7.00	9.45	58000
Wohnfläche (m ²)	1917043	75.81	35.59	0.01	54.99	68.89	88.00	1000

Tabelle 5: Summary Statistics Mietdatensatz

Tabelle 5 bietet eine deskriptive Statistik des Mietdatensatzes, der die Mietpreisentwicklung in Berlin darstellt, und kann somit wesentliche Einsichten in die Wohnkostenbelastung der Bevölkerung geben. Die Tabelle basiert dabei auf einer umfangreichen Stichprobe von 1.917.043 Beobachtungseinheiten.

Die durchschnittliche Kaltmiete beträgt 627,08 €, mit einer beträchtlichen Standardabweichung von 516,80 €, was auf eine weitreichende Diversität in den Mietkosten hinweist. Die Spanne der Mietpreise ist mit einem Minimum von 0,01 € und einem Maximum von 15.000 € außerordentlich groß, was auf sehr unterschiedliche Wohnverhältnisse und auch auf Ausreißer bzw. Datenfehler hinweist.

Die Kaltmiete pro Quadratmeter, ein wichtiger Indikator für die Mietpreisbewertung, zeigt einen Durchschnittswert von 8,50 € und eine noch höhere Standardabweichung von 115,04 €. Dies deutet auf eine noch stärkere Variabilität und mögliche Verzerrungen durch extrem hohe oder niedrige Werte hin. Der Maximalwert von 58.000 € pro Quadratmeter erscheint ungewöhnlich hoch und könnte auf besondere Fälle oder Erfassungsfehler hinweisen, die einer Überprüfung bedürfen.

Die durchschnittliche Wohnfläche in den erfassten Einheiten liegt bei 75,81 Quadratmetern, mit einer Standardabweichung von 35,59 Quadratmetern, was auf eine gemischte Wohnsituation mit einer Bandbreite von kleinen und großen Wohnungen schließen lässt. Auch hier sind Minimum- und Maximum Werte fraglich und es bedarf einer näheren Betrachtung, um die Plausibilität solcher Werte zu prüfen.

Die Perzentile geben weitere Einblicke in die Verteilung der Miet- und Wohnflächendaten. Das 25. Perzentil bei der Kaltmiete liegt bei 346,00 €, was bedeutet, dass 25% der Wohnungen eine Miete unter diesem Wert aufweisen. Der Medianwert von 467,86 € zeigt, dass die Hälfte der Wohnungen eine Kaltmiete unter diesem Betrag hat. Die oberen 25% der Wohnungen haben eine Miete, die über 706,54 € liegt, was die hohe Mietbelastung für einen bedeutenden Teil der Bevölkerung unterstreicht.

Folgende Abbildung visualisiert die Verteilung dieser drei Variablen auf Basis des Grids in Berlin. Dabei wurden wieder für die Visualisierung exemplarisch die Daten aus dem Jahre 2015 verwendet.

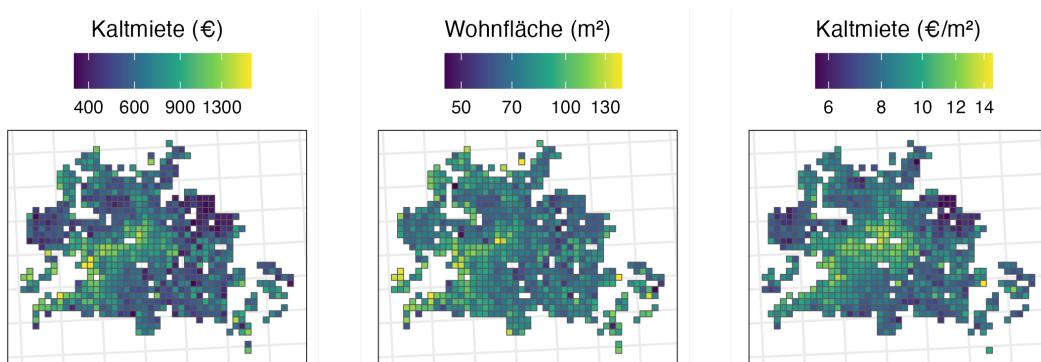


Abbildung 4: Miete auf Karte 2015

Jahr	n	Kaltmiete (€)		Wohnfläche (m ²)		Kaltmiete (€/m ²)	
		Mean	Median	Mean	Median	Mean	Median
2007	216882	439.53	362.00	74.61	69.19	6.81	5.49
2008	226760	452.74	368.56	73.89	68.67	6.22	5.60
2009	211381	471.83	382.00	74.54	69.00	6.62	5.80
2010	182552	520.49	401.38	76.37	69.30	6.77	6.11
2011	150706	562.42	432.73	76.56	69.00	7.86	6.58
2012	139518	624.59	475.00	77.71	70.00	8.02	7.30
2013	135140	700.08	525.00	79.64	70.50	8.56	8.00
2014	136700	736.57	564.00	79.32	70.51	9.58	8.50
2015	110705	758.32	595.00	79.07	70.00	9.52	8.91
2016	90353	769.64	599.00	77.03	69.00	9.83	9.10
2017	80657	847.20	662.00	76.34	67.88	10.88	10.18
2018	68290	874.30	700.08	73.55	66.20	11.93	11.00
2019	68548	918.83	729.23	71.75	65.10	13.42	11.64
2020	49753	975.44	780.00	70.88	64.03	14.47	12.66
2021	49098	1001.15	775.96	68.47	61.57	14.76	13.33

Tabelle 6: Mietdatensatz über die Jahre

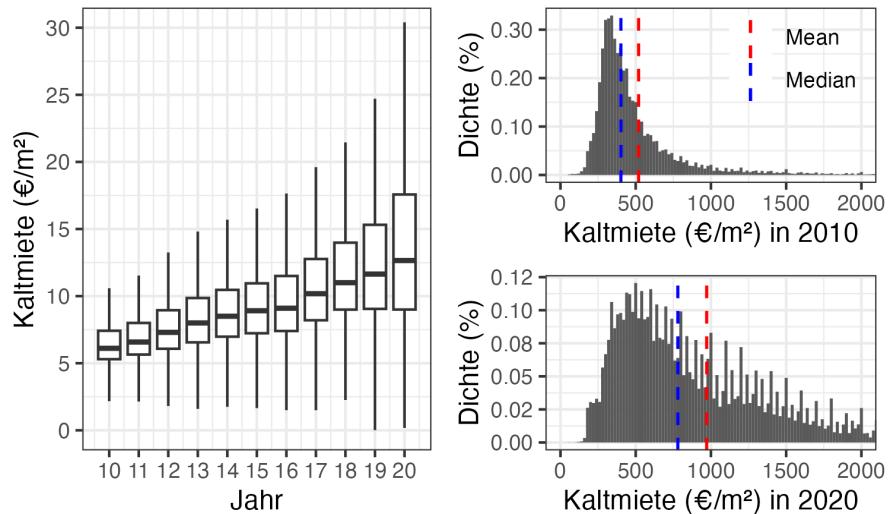


Abbildung 5: Verteilung Kaltmiete

Tabelle 6 und Abbildung 5 zeigen eine zeitliche Entwicklung der Mietpreise und Wohnfläche in Berlin, basierend auf dem Mietdatensatz über einen Zeitraum von 2007 bis 2021. Hierbei ist

wichtig zu beachten, dass die Daten des Sozialdatensatzes einen anderen Zeitraum abdecken. Für die Analysen im nächsten Abschnitt wird sich dementsprechend lediglich auf die Zeiträume beschränkt, welche in beiden Datensätzen vorhanden sind, sofern dabei Variablen aus beiden Datensätzen verwendet werden.

Die Daten veranschaulichen eine steigende Tendenz der durchschnittlichen Kaltmieten sowie der Kaltmieten pro Quadratmeter, während die Anzahl der Beobachtungen über die Jahre abnimmt. Diese Reduktion in der Stichprobengröße könnte durch verschiedene Faktoren bedingt sein, etwa durch Veränderungen in der Datenerfassung oder im Immobilienmarkt selbst.

Über den betrachteten Zeitraum hinweg steigen die durchschnittlichen Kaltmieten von 439,53 € im Jahr 2007 auf 1001,15 € im Jahr 2021, was mehr als eine Verdoppelung darstellt. Die Medianwerte folgen einem ähnlichen Trend, was auf eine breite Erhöhung der Mieten hindeutet, die sich über verschiedene Segmente des Wohnungsmarkts erstreckt. Die Wohnfläche zeigt im Vergleich dazu eine relative Konstanz mit leichten Schwankungen, was darauf hindeutet, dass die Wohnungsgrößen über die Zeit hinweg relativ stabil bleiben, während die Kosten pro Quadratmeter steigen.

Die Kaltmiete pro Quadratmeter, ein kritischer Indikator für die Bewertung der Mietbelastung, zeigt ebenfalls einen deutlichen Anstieg von 6,81 € im Jahr 2007 auf 14,76 € im Jahr 2021. Besonders auffällig ist hierbei allerdings die starke Zunahme der Streuung, welche durch Abbildung 5 besonders gut verdeutlicht wird.

Dies reflektiert eine erhebliche Steigerung der Wohnkosten in Berlin und könnte ein Indikator für zunehmende Wohnkostenbelastungen - unter anderem durch starke Kostenschwankungen verursacht - und potenziell auch für eine Verschärfung der räumlichen Segregation sein.

4. Analyse

Das nachfolgende Kapitel widmet sich der Präsentation der Analysen, die auf der fundierten Datengrundlage, wie sie zuvor beschrieben wurde, fußen. Hierbei werden die ausgewählten Variablen einer eingehenden Betrachtung unterzogen, um die Dynamiken der räumlichen Segregation in Berlin zu entschlüsseln.

4.1. Konzentrationen von Bevölkerungsgruppen

Residentielle Segregation äußert sich durch eine räumliche Konzentration von Bevölkerungsgruppen, dessen sozioökonomische oder ethnische Eigenschaften eine statistische Ähnlichkeit aufweisen(Blokland und Vief 2021). Sobald der Anteil einer Bevölkerungsgruppe in einem Teilgebiet über dem Mittelwert der gesamten Stadt liegt, wird von einer Konzentration der Gruppe gesprochen. Die intensive Ansammlung einer bestimmten Bevölkerungsgruppe kann – abhängig von den Merkmalen – zu einer positiven oder negativen Wahrnehmung führen und dazu beitragen, dass Wohnstandorte als erstrebenswert oder stigmatisiert erscheinen. Dies kann sich beispielsweise in Bezeichnungen wie "Ghettosöder" "Reichenviertel" manifestieren. Die räumliche Konzentration von Bevölkerungsgruppen kann anhand des Lokationsquotienten (LQ) gemäß der nachstehenden Formel berechnet werden:

$$LQ = \frac{\text{Anteil Bevölkerungsgruppe eines Stadtviertels}}{\text{Anteil Bevölkerungsgruppe der Gesamtstadt}}$$

Nimmt der LQ Werte über 1,0 ein, deutet dies auf eine überdurchschnittliche Präsenz der Bevölkerungsgruppe hin. Umgekehrt ist es bei Werten kleiner als 1,0, wenn die Bevölkerungsgruppe in dem Stadtviertel entsprechend unterrepräsentiert ist. Bei Werten um die 1,0 ist die Bevölkerungsgruppe in dem Stadtviertel im gleichen Maße wie im gesamten Stadtgebiet vertreten (Dohnke und Seidel-Schulze 2013).

4.1.1. Betrachtung anhand einzelner Variablen

Arbeitslosenquote Aufgrund der jahrelangen Teilung der Stadt und seiner polyzentrischen Struktur stellt Berlin im Vergleich zu anderen deutschen Großstädten ein besonders interessantes Gebiet für die Erforschung von Konzentrationen dar. Dazu wird im ersten Schritt die Verteilung der Arbeitslosenquote über das gesamte Stadtgebiet unter Verwendung des LQ untersucht. Die Arbeitslosenquote ist ein gutes Messinstrument zur Untersuchung der sozialen Struktur eines Gebiets, da sie direkte Einblicke in die Beschäftigungssituation der Bevölkerung bietet und auf wirtschaftliche Vitalität hinweisen kann. Hierfür wird die Arbeitslosenquote jeder betrachteten räumlichen Einheit mit der Arbeitslosenquote des gesamten Stadtgebiet ins Verhältnis gesetzt. Für die Interpretation der Ergebnisse ist es also zu berücksichtigen, dass ein hoher LQ auf eine

Arbeitslosenquoten in Berlin

überdurchschnittliche Arbeitslosenquote in dem jeweiligen Quadrat hinweist. Die Auswertungen werden für die Jahre 2010 und 2019 erstellt, um im Zeitverlauf mögliche Entwicklungen festzustellen. Die durchschnittliche Arbeitslosenquote in Berlin lag im Jahr 2010 bei 13,6% und im Jahr 2019 bei 7,8%.

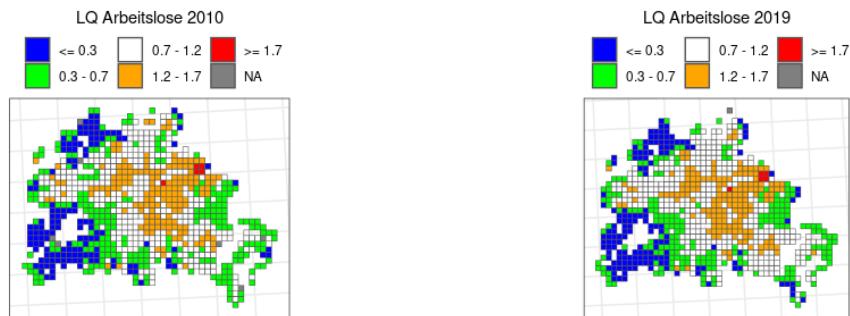


Abbildung 6: Arbeitlosenquote 2010 und 2019

Auf den ersten Blick zeichnet Abbildung 6 ähnliche Muster ab, was darauf schließen lässt, dass sich zwischen den Jahren 2010 und 2019 keine signifikanten Veränderungen in der räumlichen Verteilung von Bevölkerungsgruppen mit ähnlichen Arbeitslosenquoten ereignet haben. Dennoch lässt sich unter Betrachtung des gesamten Stadtgebietes ein heterogenes Bild erkennen. Die erste Auffälligkeit besteht in dem Unterschied der Konzentrationen zwischen dem Stadtrand und den inneren Bezirken. Die Gebiete am Stadtrand verfügen, bis auf eine Ausnahme im östlichen Teil Berlins, im Vergleich zu zentraleren Gebieten über geringere Arbeitslosenquoten. Dieser Umstand wird besonders im nördlichen und südlichen Stadtrand im westlichen Stadtgebiet bemerkbar. An keiner weiteren Stelle im restlichen Stadtgebiet lassen sich räumliche Konzentrationen von gerin- gen Arbeitslosenquoten in einer derartigen Größe erkennen. Diese Gebiete zeichnen sich überwie- gend durch eine Bebauung mit Ein- sowie Zweifamilienhäusern aus und deuten auf den Wohnsitz von wohlhabenderen Bevölkerungsgruppen hin. Es ist zu erkennen, dass die weißgefärbbten Ge- biete mit durchschnittlichen LQ-Werten als eine Art Pufferzone zwischen den Randlagen und den zentraler gelegenen Gebieten fungieren, bevor sie wieder vermehrt im Stadtkern zu finden sind. Im Gegenteil dazu zeichnet sich am Rand des östlichen Stadtgebiets eine Konzentration von Bevöl- kerungsgruppen mit besonders hoher Arbeitslosenquote ab. Es handelt sich dabei um den Bezirk Marzahn-Hellersdorf, welcher in der Öffentlichkeit für seine großflächige Plattenbau-Bebauung bekannt ist. Dies führt zur zweiten Auffälligkeit, nämlich dem unterschiedlichen Vorkommen von Gebieten mit hoher Arbeitslosenquote zwischen dem östlichen- und dem westlichen Stadtgebiet. Anhand der orangenen und roten Einfärbung der Quadrate, die repräsentativ für höhere Arbeits- losenquoten stehen, zeichnet sich im östlichen Stadtteil eine deutlich höhere Konzentration von Bevölkerungsgruppen mit überdurchschnittlichen Arbeitslosenquoten ab.

Zur Plausibilisierung der Ergebnisse soll nachfolgend eine weitere Studie herangezogen werden, die sich ebenfalls mit der Konzentration von Bevölkerungsgruppen auf Basis der Arbeitslosenquo-

Soziale Segregation und Arbeitslosigkeit

te beschäftigt. Die Studie wurde im Jahr 2021 von zwei Forschern des Instituts für Sozialwissenschaften der Humboldt-Universität zu Berlin veröffentlicht und betrachtet den Zeitraum zwischen 2007 und 2016. Aufgrund des marginalen Unterschieds zwischen den betrachteten Zeiträumen wird die Vergleichbarkeit zwischen den Studien als gut beurteilt. In Der Vergleichsstudie werden größere räumliche Einheiten betrachtet.

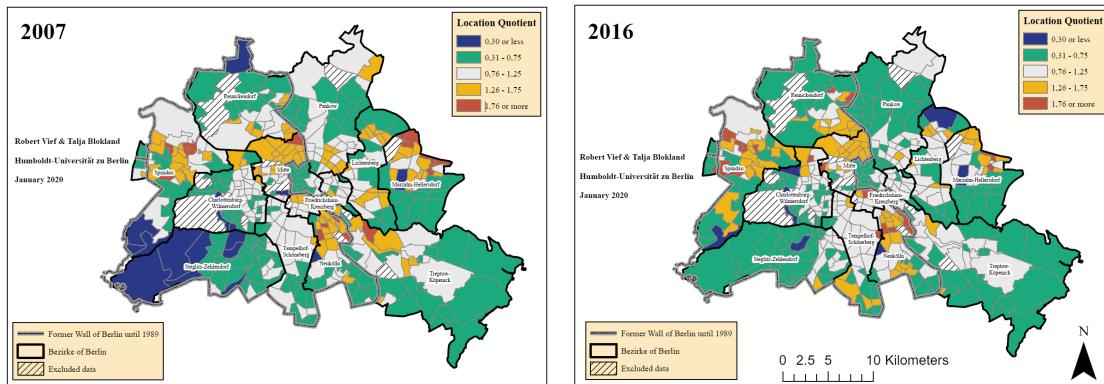


Abbildung 7: LQ Arbeitslosenquote Vergleichsstudie 2007 und 2016 (Blokland und Vief 2021)

Während sich die Konzentration in den vorher genannten auffälligen Gebieten (süd-westlicher Stadtrand und Marzahn Hellersdorf) leicht normalisierte (LQ-Werte näher an der 1), durchliefen einige Gebiete im Norden eine gegenteilige Entwicklung, wie es Abbildung 7 zu entnehmen ist. Grundsätzlich macht sich aber kein klarer Trend im gesamten Stadtgebiet bemerkbar, da die verschiedenen Gebiete in dem Betrachtungszeitraum unterschiedliche Entwicklungen vollzogen. Sollte es zu Veränderungen gekommen sein, haben die LQ in der Regel Werte ihrer benachbarten Intervalle angenommen (bspw. blau wurde zu grün), woraus sich schließen lässt, dass es zu keinen drastischen Umbrüchen gekommen ist. Interessanterweise kann der Verdacht, dass in dem östlichen Stadtgebiet mehr Gebiete mit erhöhter Arbeitslosigkeit vorhanden sind, nicht widerlegt werden. Dies macht sich vor allem an den Ergebnissen des Jahres 2016 bemerkbar, das mehr Gebiete mit überdurchschnittlicher Arbeitslosenquote im westlichen- als im östlichen Stadtgebiet aufzeigt. Hinsichtlich dem Unterschied zwischen den Randlagen und den inneren Bezirken zeichnet die Vergleichsstudie ein ähnliches Bild.

Bei der Interpretation der Ergebnisse gilt es zu beachten, dass unterdurchschnittliche respektive überdurchschnittliche Konzentrationen auf eine mangelnde soziale Durchmischung hindeuten. Beide Studien vermitteln somit den Eindruck, dass dieser Umstand auf mehr als die Hälfte des Berliner Stadtgebiets zutrifft. Hierbei sind besonders die äußersten Randlagen und die innenliegenden Gebiete, um den Stadtkern zu erwähnen. Folglich lässt sich vermuten, dass die genannten Gebiete eine höhere Segregation als die dazwischenliegende Pufferzone aufweist und diese über die Jahre annährend stabil geblieben ist.

Kaufkraft Im nächsten Schritt wird die Konzentration von Bevölkerungsgruppen anhand der Kaufkraft pro Haushalt betrachtet, um etwaige Unterschiede in der finanziellen Stärke der Haushalte zwischen den Gebieten festzustellen. Weiterhin soll die Untersuchung dazu dienen die Ergebnisse der Arbeitslosenverteilung zu plausibilisieren und potenzielle Wechselwirkungen zwischen den beiden Faktoren zu identifizieren. Die Variable Kaufkraft spiegelt das Haushaltseinkommen wider. Sie umfasst Informationen über die Arbeitsangebotsmenge, das Kapitalvermögen, Mieteinnahmen und Leasingeinnahmen abzüglich Steuern und Sozialabgaben, einschließlich sozialer Transfers wie Arbeitslosenunterstützung, Kindergeld und Renten (Breidenbach und Eilers 2018). Demnach besteht der Hinweis, dass es sich um die Kaufkraft pro Haushalt handelt und der Wert somit maßgeblich mit der Haushaltsgröße zusammenhängt. Weisen zwei Haushalte ähnliche Kaufkraftwerte, aber unterschiedliche Haushaltsgrößen auf, steht den Bewohnern der beiden Haushälter unterschiedlich viel Geld zur Verfügung. Um ein grobes Verständnis für die pro Kopf Werte zu erhalten, wird an dieser Stelle erwähnt, dass sich die durchschnittliche Haushaltsgröße in Berlin auf 1,86 beziffert und in dem Betrachtungszeitraum annähernd stabil blieb (Statistisches Bundesamt 2022). Dennoch kann sie nützliche Implikationen für die Feststellung der wirtschaftlichen Lebensverhältnisse und die Identifikation von wohlhabenden- und finanzschwachen Gebieten liefern. Ausgehend von einem Medianwert von ca. 34.000 € in 2010 und 39.000 € in 2019 wurden verschiedene Kaufkraft-Klassen abgeleitet, wobei orangene und rote Werte auf eine unterdurchschnittliche-, weiße auf eine durchschnittliche- und grüne Werte auf eine stark überdurchschnittliche Kaufkraft hinweisen. Für das Jahr 2019 werden Intervalle mit höheren Werten gewählt, um einen infaltionsbereinigten Vergleich zwischen den zwei Jahren zu ermöglichen.

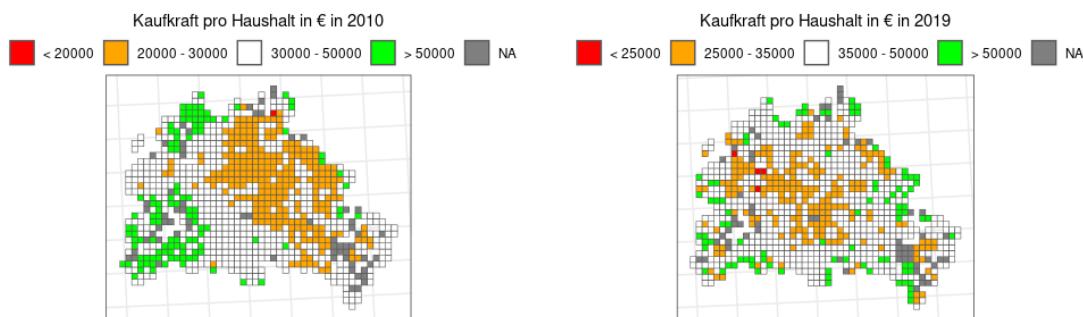


Abbildung 8: Kaufkraft 2010 und 2019

Der im Rahmen der Arbeitslosenverteilung festgestellte Unterschied zwischen Randlagen und zentralen Gebieten kann anhand der Abbildung 8 zur Verteilung der Kaufkraft-Klassen bestätigt werden. Konträr zur Arbeitslosenquote, wird bei der Kaufkraftverteilung zwischen den Jahren eine signifikante Veränderung deutlich. Im Jahr 2010 ist ähnlich zur Arbeitslosenquote eine klare Trennung zwischen Ost und West mit einer dazwischenliegenden Pufferzone zu erkennen. Während fast der komplette Osten unterdurchschnittliche Werte aufweist, befinden sich im Westen überproportional viele Haushälter mit stark überdurchschnittlicher Kaufkraft. Dieses Bild hat sich im

Jahr 2019 grundlegend verändert. Das östliche Stadtgebiet hat sich weitestgehend normalisiert und verfügt mit einem dominierenden Anteil an Haushalten mit durchschnittlicher Kaufkraft ungefähr über die gleiche Anzahl an Haushalten mit unterdurchschnittlicher respektive überdurchschnittlicher Kaufkraft. Demnach kann der mögliche Schluss gezogen werden, dass die Segregation in dem Stadtgebiet abgenommen hat. Anders verhält es sich im westlichen Stadtgebiet, welches einen deutlichen Anstieg an Konzentrationen mit unterdurchschnittlicher Kaufkraft und einen Rückgang der überdurchschnittlichen Kaufkraft-Klasse verzeichnete. Zwar ist der Anteil der durchschnittlichen Kaufkraft-Klasse immer noch am höchsten, dennoch entfällt ein nicht unbedeutlicher Teil auf die unterdurchschnittliche Kaufkraft-Klasse. Die Segregation hat sich in dem Fall nicht verringert oder vermehrt, sondern wurde eher mit einer Änderung der Kaufkraft-Klasse von überdurchschnittlich zu unterdurchschnittlich geografisch umverteilt.

Wie bereits erwähnt, müssen die Veränderungen der Kaufkraft-Klassen nicht zwingend auf eine Verbesserung oder Verschlechterung der wirtschaftlichen Lebensverhältnisse hindeuten. Sie können auch durch Veränderungen der Haushaltsstrukturen begründet sein. Beispielsweise können sich in Folge des wirtschaftlichen Aufschwungs Berlins im Westen vermehrt Einpersonenhaushalte gebildet haben, während sich im Osten größere Haushalte entstanden sind. Eine genaue Untersuchung der Ursachen würde den Rahmen der Studienarbeit sprengen und wurde somit nicht weiter fortgeführt.

Ausländerquote Um einen Schritt weiterzugehen, soll neben der Betrachtung der sozialen und wirtschaftlichen Lebensverhältnisse zusätzlich die Konzentration von ausländischen Bevölkerungsgruppen untersucht werden. Hierfür wird die Ausländerquote herangezogen und zur Bildung des LQ mit der durchschnittlichen Arbeitslosenquote der Gesamtstadt ins Verhältnis gesetzt. Die durchschnittliche Ausländerquote in Berlin lag im Jahr 2010 bei 13,7% und im Jahr 2019 bei 19,2%. Der im Daten-Kapitel beschriebene methodische bei der Messung der Ausländerquote ab 2017 stellt für die nachfolgende Untersuchung keine Einschränkung dar, da die räumlichen Muster der Verteilung von ausländischen Bevölkerungsgruppen unverändert geblieben sind. Es haben sich lediglich die Prozentwerte verändert, welche durch die Bildung des LQ bereinigt werden.

Der Abbildung 9 kann entnommen werden, dass sich das Bild der vorausgegangenen Untersuchungen auch im Fall der Ausländerquote fortsetzt. Zuerst ist wieder eine klare Trennung zwischen dem Osten und Westen erkennbar. Ausländer sind fast im gesamten östlichen Stadtgebiet stark unterrepräsentiert, was besonders durch die blaugefärbten Randgebiete im Jahr 2010 deutlich wird. Im Vergleich zu 2019 sind die Konzentrationen in den genannten Gebieten auf ein normales Niveau gestiegen, was darauf schließen lässt, dass sich dort mehr Ausländer niedergelassen haben. Dennoch hat sich an der generellen Tatsache, dass Ausländer im gesamten östlichen Gebiet deutlich unterrepräsentiert sind, nichts geändert. Die Randlagen in Westberlin weisen ähnlich unterdurchschnittliche Konzentrationen von Ausländern auf. Das genaue Gegenteil ist im westli-

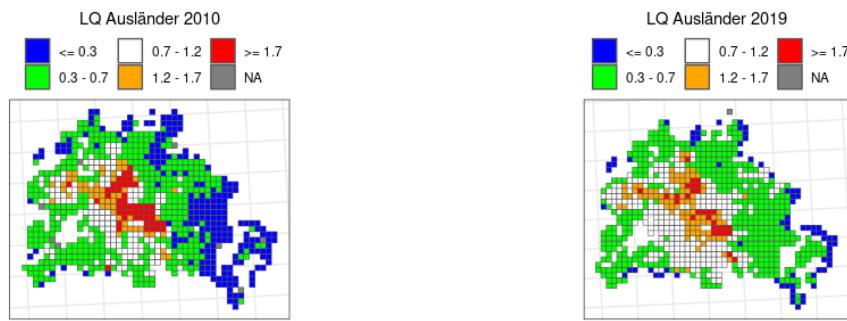


Abbildung 9: Ausländerquote 2010 und 2019

chen Stadtzentrum vorzufinden, wo sich ein zusammenhängendes Cluster aus Gebieten mit überdurchschnittlichen Ausländerquoten ausgebildet hat. Diese signifikanten Unterschiede zwischen Randlage und Stadtzentrum bekräftigen erneut das Ergebnis der grundlegend verschiedenen sozioökonomischen Strukturen der beiden Teilläume. Auch im Westen hat sich zwischen den Jahren ein leichter Trend zur verstärkten Durchmischung der Gebiete gezeigt, der unter anderem an der Abnahme der stark über durchschnittlichen Konzentrationen (rot) im Stadtkern und der Normalisierung des süd-westlich gelegenen Raums sichtbar wird. Außerdem lässt sich anhand der Entwicklung wieder die Entstehung einer gut durchmischten Pufferzone zwischen Stadtrand und Innenstadt erahnen. Um die Forschungsergebnisse zu plausibilisieren, sollen sie mit der bereits erwähnten Vergleichsstudie verglichen werden.

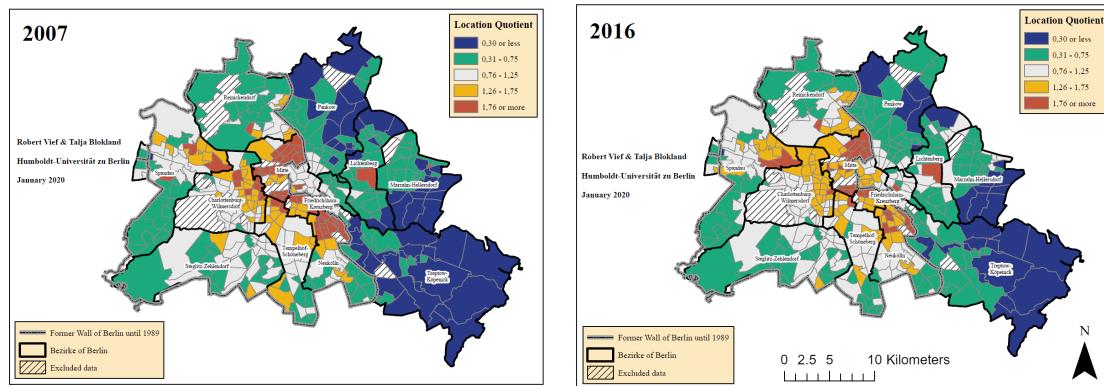


Abbildung 10: LQ Ausländerquote Vergleichsstudie 2007 und 2016 (Blokland und Vief 2021)

Die Vergleichsstudie, in Abbildung 10 zu sehen, kommt zu einem ähnlichen Ergebnis und bekräftigt die Erkenntnisse der eigenen Studie. Während vor allem in der Innenstadt Ausländer im hohen Maße vertreten sind, fällt ihre Konzentration in den außenliegenden Gebieten unterdurchschnittlich aus. Zusätzlich zeichnet sich zwischen den Jahren eine Normalisierung in den Gebieten mit stark unterdurchschnittlicher (blau) und überdurchschnittlicher (rot) Konzentration ab.

Das Resultat zeigt somit, dass die Segregation von Ausländern zu Beginn des Betrachtungszeit-

raums auf einem hohen Niveau war und trotz eines Trends zu verstärkter Durchmischung im Jahr 2019 nach wie vor präsent war. Es darf allerdings nicht vergessen werden, dass Ausländer ein sehr breit gefasster Begriff ist, da hierbei bis auf Deutsche alle restlichen Nationalitäten und Kulturen in einen Topf geschmissen werden. Ein Gebiet mit einer hohen Ausländerquote könnte demnach ebenso aus einem hohen Anteil von Bürgern aus dem EU-Ausland, wie auch einem hohen Anteil aus weiter entfernten Ländern, oder einem Mix aus beiden bestehen. Beispielsweise stammen im Zeitraum von 2011 bis 2017 im gentrifizierten Bezirk Neukölln von den 13.263 neuen Einwohnern aus dem Ausland rd. 10.000 aus der EU oder der USA und 2.600 aus muslimisch geprägten Ländern (Blokland und Vief 2021).

Im Rahmen der Untersuchungen konnten Zusammenhänge identifiziert werden, die nachfolgend näher erläutert werden sollen. Der erste Zusammenhang besteht in dem Unterschied zwischen dem östlichen und dem westlichen Stadtgebiet, der vor allem anhand der Verteilung der Ausländerquote verbildlicht wurde. Eine Begründung hierfür liefert die Geschichte Berlins. Die nach dem Zweiten Weltkrieg bis ins Jahr 1990 andauernde Teilung der Stadt sorgte für unterschiedliche Entwicklungen in West-Berlin, welches der Bundesrepublik Deutschland (BRD) angehörte, und Ost-Berlin, welches zum Staatsgebiet der Deutschen Demokratischen Republik (DDR) gehörte. Die ab Mitte der 50-er Jahre einsetzende Konjunktur der BRD lockte viele Gastarbeiter aus dem Ausland nach Westdeutschland, unter anderem auch nach Westberlin. Ein Beispiel hierfür ist das 1961 unterzeichnete Anwerbeabkommen zwischen der BRD und der Türkei. Während nach dem Anwerbestopp im Jahr 1973 ein Teil wieder in die Heimat zurückkehrte, ließ sich der andere Teil in Westberlin nieder und holte seine Familie hinterher. Bis auf wenige Werkvertragsnehmer aus mit der DDR befreundeten Ländern wie Vietnam oder Kuba, spielte sich eine Entwicklung derartiger Größe nicht in Ost-Berlin ab (Ulrich Raiser 2005). Weiterhin war Ost-Berlin von den Folgen der Deindustrialisierung Berlins, welche nach der Wende einsetzte, stärker als West-Berlin betroffen und hatte dementsprechend lange mit höheren Arbeitslosenquoten zu kämpfen (Blokland und Vief 2021). Demnach ist es nicht verwunderlich, dass in den Grafiken die Trennlinie zwischen den verschiedenen Konzentrationen an der Verlauf der Berliner Mauer erinnert und die Folgen der damaligen Geschehnisse heutzutage immer noch im Stadtbild erkennbar sind. In Abbildung 11 ist der Verlauf der ehemaligen Mauer dargestellt.

Der zweite Zusammenhang zielt auf das durchgängige Muster ab, das zwischen dem Stadtkern und den Randgebieten besteht. Ausgehend von geringen Konzentrationen in den Randgebieten werden mit zunehmender Nähe zum Stadtzentrum höhere Konzentrationen erreicht. Zwar bestehen mit Teilen der Bezirke Marzahn-Hellersdorf im Osten und Spandau im Westen Ausnahmen, im Verhältnis zu den restlichen Randgebieten fallen die Bezirke aber wenig ins Gewicht und werden deshalb nicht weitergehend untersucht. Als weiteres Merkmal konnte die Bildung einer Pufferzone mit durchschnittlichen Konzentrationen zwischen Stadtrand und Innenstadt festgestellt werden.

Demnach sind benachbarte Gebiete mit großen Wertunterschieden eher eine Seltenheit. Dieser

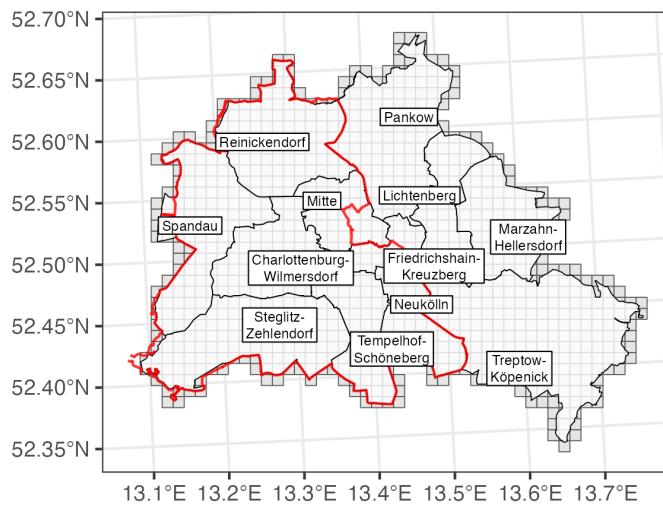


Abbildung 11: Verlauf der Berliner Mauer

Umstand äußert sich außerdem durch die Bildung von Clustern, welcher anhand der räumlichen Nähe von Gebieten mit ähnlichen Konzentrationen ersichtlich wird. Ein geeignetes Beispiel hierfür liefert das Stadtzentrum auf den Grafiken der Ausländerverteilung. Nachdem in den vorausgegangenen Untersuchungen eindimensionale Betrachtungen der Konzentrationen von Bevölkerungsgruppen vorgenommen wurden, sollen im folgenden Kapitel mehrere Variablen zu einer mehrdimensionalen Analyse gebündelt werden, um mögliche Cluster Effekte in der Bevölkerungsstruktur zu identifizieren.

4.1.2. Betrachtung anhand einer Clusteranalyse

Die Clusteranalyse verfolgt das Ziel, potenzielle Gruppierungen oder Cluster von ähnlichen Merkmalen in den Daten zu identifizieren. Sie ist ein statistisches Verfahren, das Daten in Gruppen oder Cluster unterteilt, wobei die Werte der verwendeten Variablen innerhalb desselben Clusters möglichst ähnlich sind und die Werte zwischen verschiedenen Clustern möglichst unterschiedlich sind. Der k-means Algorithmus wird verwendet, um die Variablen in Clustern zu gruppieren, wobei die Anzahl der Cluster im Voraus festgelegt wird. Der Algorithmus sucht nach Clusterzentren, die die Datenpunkte bestmöglich repräsentieren, und weist dann jeden Datenpunkt dem nächstgelegenen Clusterzentrum zu. Durch die Analyse der Cluster können Einblicke gewonnen werden, wie sich die Bevölkerung Berlins in homogene Gruppen unterteilen lässt und welche Merkmale diese Gruppen möglicherweise gemeinsam haben. Im ersten Schritt werden die Input-Variablen festgelegt. Um an die vorausgegangenen Analysen anzuknüpfen, fließen die bereits beleuchteten Variablen auch in die Clusteranalyse mit ein. Weiterhin wird durch die Auswahl sichergestellt, dass mehrere Facetten der sozioökonomischen Bevölkerungsstruktur beleuchtet werden. Anhand der Arbeitslosenquote werden Aspekte hinsichtlich des sozialen Status berücksichtigt. Die Kaufkraft

pro Haushalt erlaubt Rückschlüsse über die finanzielle Stärke der Haushälter und die Ausländerquote über die kulturelle Zusammensetzung der Bevölkerung. Zusätzlich sollen die Variablen Anteil an Einfamilienhäusern und Anteil an Senioren (60+ Jahre) in die Analyse integriert werden, um ein besseres Verständnis für ihre Verteilung innerhalb der Stadt zu erhalten und daraus entstehende Nebeneffekte zu berücksichtigen. Als Nächstes werden die Variablen um ihren Mittelwert und ihre Varianz bereinigt, um eine Gleichgewichtung zwischen den Variablen herzustellen. Zur Auswahl der Anzahl an Clustern wird der Elbow Plot (Abbildung 12) verwendet, der die Anzahl an Clustern gegen die Total Within Sum of Squares, also der Gesamtsumme der quadratischen Abweichungen innerhalb der Cluster abträgt.

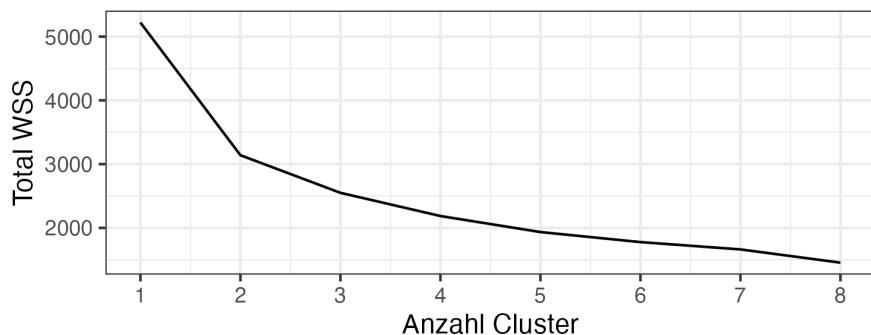


Abbildung 12: Elbow Plot Cluster

Die Abbildung 12 zeigt das Ergebnis, dass es keine perfekte Anzahl an Clustern gibt. Auf der einen Seite erlaubt eine höhere Anzahl an Clustern eine feinere Unterteilung, auf der anderen Seite steigt damit das Risiko der Unübersichtlichkeit. Um beide Aspekte zu berücksichtigen, wird die Anzahl an Clustern mit vier festgelegt. Als letzte Voraussetzung muss das Untersuchungsjahr definiert werden. Im vorliegenden Fall wird die Berechnung für das Jahr 2015 durchgeführt und das in Tabelle 7 enthaltene Ergebnis (Mittelpunkte der Cluster) erzielt:

	1	2	3	4
Anzahl Haushalte	-0.63	0.29	-0.53	2.12
Arbeitslosenquote	-0.06	0.80	-1.02	0.79
Kaufkraft	-0.28	-0.62	1.02	-0.67
Ausländeranteil	-0.52	0.13	-0.42	2.10
Anteil EFH	1.03	-0.83	0.52	-1.11
Anteil Senioren	-0.12	-0.21	0.72	-1.43

Tabelle 7: Interpretation Cluster 2015

Es ist zu erwähnen, dass die Werte nicht direkt interpretiert werden können, da sie vorher durch mittelwertbereinigt und varianzbereinigt wurden. Demnach erfolgt die Interpretation anhand der Unterschiede zwischen den Werten.

Anzahl Haushalte:

- Cluster 4 verfügt über die höchste Haushaltsdichte, woraus folgt, dass hier pro Quadratkilometer mehr Menschen als in anderen Teilen Berlins leben.
- In Cluster 1 ist die Haushaltsdichte am geringsten

Arbeitslosenquote:

- Die niedrigste Arbeitslosenquote wird in Cluster 3 verzeichnet
- Cluster 2 und 4 verfügen über die höchsten Arbeitslosenquoten

Kaufkraft:

- Cluster 3 weist die höchste Kaufkraft pro Haushalt auf, daher handelt es sich hierbei vermutlich um wohlhabende Gegenden
- Cluster 2 und 4 deuten auf finanziell schwächere Gegenden hin

Ausländeranteil:

- Der Ausländeranteil fällt in Cluster 4 am höchsten aus, was darauf hindeutet, dass es sich vermutlich um zentral gelegene Gegenden handelt
- Gebiete in Cluster 1 werden am seltensten von Ausländern bewohnt

Anteil Einfamilienhäuser:

- Cluster 1 ist am stärksten von einer Bebauung mit Einfamilienhäusern geprägt, demnach sind die Gebiete vermutlich am Stadtrand gelegen
- Cluster 4 verfügt über den niedrigsten Anteil an Einfamilienhäusern

Anteil Senioren:

- Senioren sind am stärksten in Cluster 3 repräsentiert
- Cluster 4 verzeichnet den geringsten Anteil an Senioren

Die Interpretation der Wertunterschiede ermöglicht es Rückschlüsse auf die Eigenschaften und Strukturen der Cluster zu ziehen. Deshalb soll basierend auf der Analyse eine Beschreibung mit der Cluster gewagt werden. Bei Cluster 1 handelt es sich um Gebiete, die aufgrund des hohen Anteiles an Einfamilienhäusern und der geringen Haushaltsdichte wahrscheinlich eher in den Randlagen vorzufinden sind. Eine weitere Bestätigung hierfür liefert der niedrige Anteil an Ausländern, die sich nach vorausgegangenen Analysen überwiegend im Stadtzentrum konzentrieren. Eine hohe Kaufkraft und niedrige Arbeitslosenquote deuten auf eine wirtschaftliche und soziale Stabilität der Gebiete hin. Anders verhält es sich bei Cluster 2, bei dem aufgrund der hohen Arbeitslosenquote und geringen Kaufkraft die Vermutung von wohlhabenden Gegenden ausgeschlossen werden kann und eher normale bis unterdurchschnittliche Lebensumstände anzunehmen sind. Da es sich nicht um das Cluster mit der höchsten Haushaltsdichte handelt, sind die Gebiete vermutlich vorwiegend in dem Bereich zwischen Stadtrand und Innenstadt angesiedelt. Anhand der Werte für die Anzahl der Haushalte und den Anteil an Einfamilienhäusern, die eine

ähnliche Tendenz zu Cluster 1 aufweisen, kann bei Cluster 3 ebenfalls von einer Lage in den Randgebieten ausgegangen werden. Im Vergleich zu Cluster 1, zeichnen sich die Gebiete von Cluster 3 durch eine höhere Kaufkraft und einen deutlich höheren Seniorenanteil aus. Dementsprechend ist es nicht verwunderlich, dass hier die Arbeitslosenquote am niedrigsten ausfällt. Es ist somit von wohlhabenden Gebieten auszugehen. Cluster 4 verfügt ähnlich zu Cluster 2 über eine niedrige Kaufkraft und eine hohe Arbeitslosenquote und lässt entsprechend auch normale bis unterdurchschnittliche Lebensumstände erahnen. Die klare Abgrenzung zwischen den Clustern wird anhand der hohen Haushaltsdichte und dem hohen Ausländeranteil deutlich, der darauf schließen lässt, dass es sich um den westlichen Teil des Stadtzentrums handelt. Die Senioren sind dabei stark unterrepräsentiert.

Um die Interpretation der Werte zu plausibilisieren, werden die zugrundeliegenden räumlichen Einheiten entsprechend ihrer Zuordnung zu den Clustern eingefärbt und auf einer Karte dargestellt. Dadurch soll zusätzlich überprüft werden, ob die Cluster räumlich zusammenhängen und somit größere Gebiete bilden. Es wird darauf hingewiesen, dass in den Algorithmus der Clusterbildung keine geografischen Variablen eingeflossen sind. Sollten also Gebiete des gleichen Clusters räumlich beieinander liegen, ist dies ausschließlich auf ihre Ausprägungen der betrachteten Variablen zurückzuführen und nicht auf ähnliche Koordinatenwerte. Das Ergebnis wird in Abbildung 13 dargestellt.

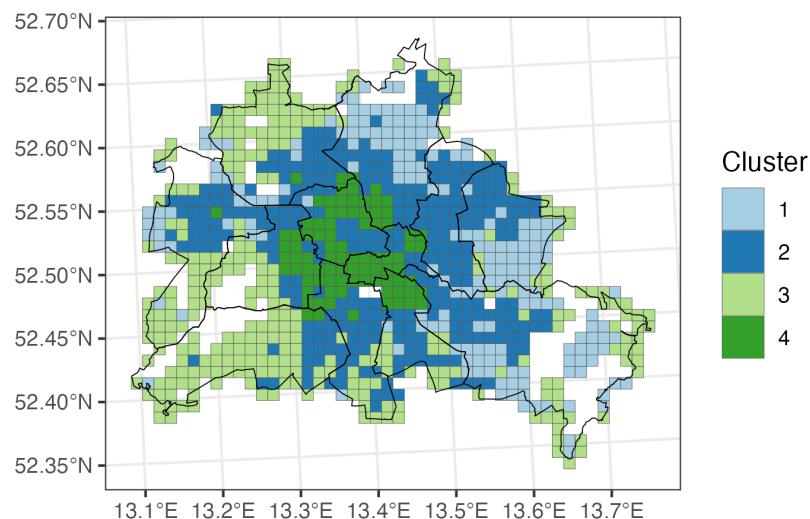


Abbildung 13: Clustering 2015

Die räumliche Lage der Cluster aus der Interpretation kann größtenteils bestätigt werden. Ebenso sind größere zusammenhängende Gebiete, die dem gleichen Cluster angehören, zu erkennen. Allerdings sind die Cluster nicht immer vollkommen homogen, wie es beispielsweise an den dunkelblauen Quadranten innerhalb des südwestlich gelegenen Cluster dargestellt wird. Weiterhin wird wieder der Unterschied zwischen dem Stadtrand und zentraler gelegenen Gebieten deutlich.

Der Unterschied zwischen Ost- und Westberlin macht sich vor allem an dem Verhältnis zwischen Cluster 1 und Cluster 3 bemerkbar. Somit liegt die Vermutung nahe, dass die meisten Senioren in Westberlin wohnhaft sind und mehr wohlhabende Gebiete vorgefunden werden können. Der zweite Unterschied ist an der stärkeren Präsenz von Gebieten des vierten Clusters in dem westlichen Zentrum zu erkennen. Cluster 2 verteilt sich gleichmäßig über die beiden Stadtgebiete und lässt somit auf keine Imbalance schließen.

Um die geografische Verteilung der Cluster zu überprüfen, erfolgt ein Abgleich mit den Wahlergebnissen des darauffolgenden Jahres. Dies beruht auf den Ergebnissen mehrerer Studien, dass sozioökonomische Faktoren einen Einfluss auf das Wahlverhalten der Bevölkerung haben können. Bürger neigen eher dazu, für Parteien zu stimmen, deren Wahlprogramme besser auf ihre individuellen Anliegen eingehen (Waiphot, Lerdtomornsakul und Homyamyen 2023). Beispielsweise unterstützen Personen mit höherem Einkommen eher konservative und rechtsgerichtete Parteien. Dies kann darauf zurückgeführt werden, dass sie glauben, dass konservative Politik, wie niedrigere Steuern und weniger staatliche Intervention, mit ihren wirtschaftlichen Interessen übereinstimmt. Auf der anderen Seite unterstützen Personen mit niedrigerem Einkommen eher linksgerichtete Parteien, die sich für Politik einsetzen, die Einkommensumverteilung, Sozialprogramme und wirtschaftliche Gleichheit fördern (Alberto und Glaeser 2004).

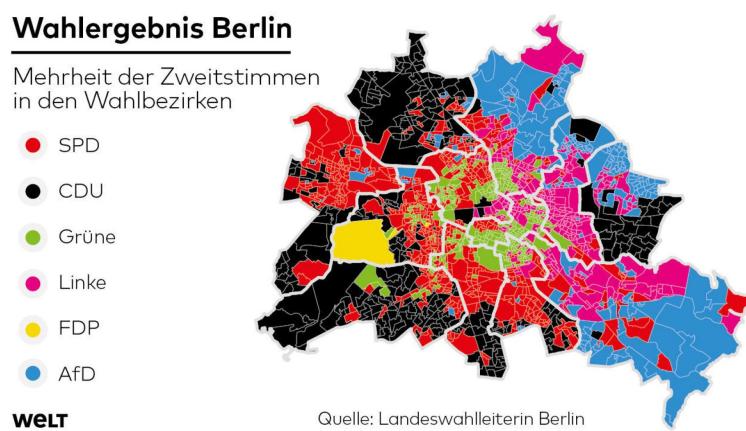


Abbildung 14: Zweitstimmen Landtagswahl 2016 (Pauly 2016)

Es lässt sich ein hoher Deckungsgrad zwischen der Clusterverteilung und den Wahlergebnissen anhand Abbildung 14 feststellen. Weiterhin können die Ergebnisse der Studie von Alesina und Rosenthal plausibilisiert werden. In den einkommensstärkeren Gegenden (Cluster 1 und 3) haben die konservativen Parteien CDU und AfD die Mehrheit der Zweitstimmen erhalten. Linksgerichtete und soziale Parteien wie die SPD und Die Linke konnten vorwiegend in den einkommenschwächeren Gegenden gewinnen. Abschließend wird überprüft, ob die Cluster im Verlauf der Zeit stabil geblieben sind. Hierfür werden Clusteranalysen für die Jahre 2005, 2010, 2015 und 2019 erstellt, um Einblicke in möglicher Veränderungen zu erhalten. Weiterhin ist die Stabilität der Cluster ein wichtiges Kriterium für die Durchführung von Regressionsanalysen, die auf den

Clustern basieren.

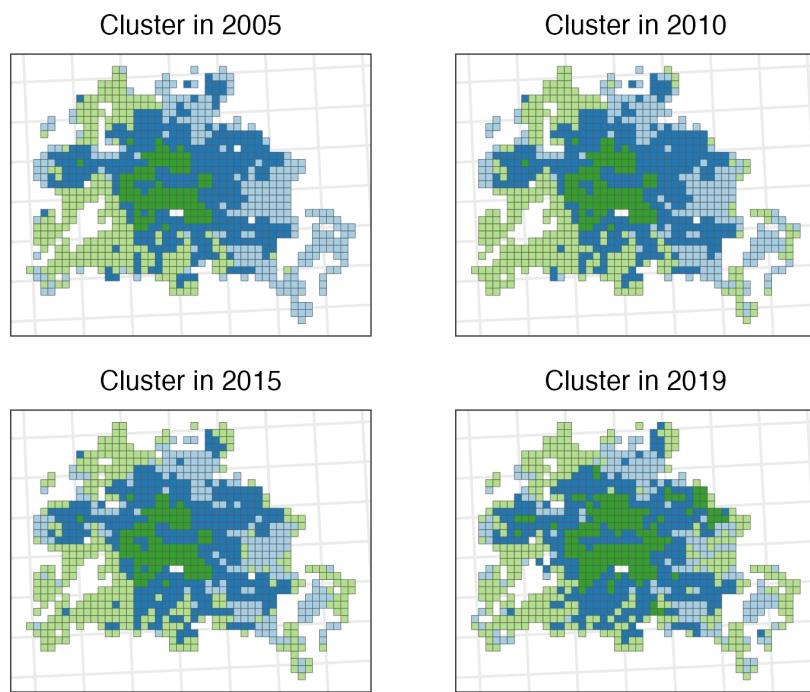


Abbildung 15: Cluster über die Zeit

Abbildung 15 zeigt, dass sich im Zeitraum von 2005 bis 2015 keine signifikanten Veränderungen ergeben haben. Erst im Jahr 2019 werden Abweichungen zu den Vorjahren ersichtlich. Bei der Interpretation der Veränderungen eröffnet sich die Limitation, dass sie nicht konkret auf einzelne Variablen zurückgeführt werden können, da sich die Cluster aus mehreren Variablen zusammensetzen.

Die erste Auffälligkeit besteht in der Expansion des vierten Clusters in östlicher Richtung. Zusätzlich hat sich am östlichen Stadtrand eine Art Enklave gebildet. Mögliche Gründe dafür könnte eine höhere Haushaltsdichte als Folge von Nachverdichtungsmaßnahmen oder ein verstärkter Zuzug von Ausländern sein. Weiterhin werden in Ostberlin vermehrt Gebiete von Cluster 1 durch Cluster 3 ersetzt. Diese Entwicklung könnte auf eine Stärkung der Kaufkraft oder den Anstieg des Seniorenteils infolge des demografischen Wandels hindeuten. Im Fall von Veränderungen wandeln sich Cluster folglich eher in ihnen ähnliche Cluster um. Dementsprechend haben sich im Betrachtungszeitraum von 2005 bis 2019 keine tiefgreifenden strukturellen Veränderungen ergeben. Lediglich im Süden macht sich eine gegenteilige Entwicklung bemerkbar. Hier sind im Zeitverlauf Gebiete von Cluster 3 zu Cluster 1 übergegangen.

4.2. Regressionsanalyse

Nun möchten wir verschiedene Regressionsanalysen durchführen, um die Entwicklung von Miete und Mietbelastungen durch verschiedene Modelle zu erklären. Wir verwenden die Methode der kleinsten Quadrate, auf Englisch auch Ordinary Least Squares oder kurz OLS genannt. Dabei handelt es sich um ein Standardwerkzeug in der Ökonometrie, das aufgrund seiner Einfachheit und Effizienz bei der Schätzung linearer Beziehungen bevorzugt wird.

Wir beginnen mit einem einfachen Modell, in dem wir die Miete bzw. Mietbelastung auf das jeweilige Jahr regressieren. Dadurch erhalten wir die durchschnittliche Mietsteigerung über gesamt Berlin. Dies gibt uns eine Baseline für die weiteren Auswertungen. Wir nutzen die (Netto)Kaltmiete in €/m² und definieren die Mietbelastung als den Quotienten aus der (Netto)Kaltmiete pro Jahr und dem mittleren Einkommen pro Gebiet. Diese Definition weicht leicht von der typischen Definition ab, welche die Bruttokaltmiete, also inklusive „kalter“ Nebenkosten wie Müllabfuhr, verwendet (Statistisches Bundesamt 2019). Wir müssen jedoch mit den uns zur Verfügung stehenden Daten arbeiten. Im nächsten Modell verwenden wir als erklärende Variable zusätzlich eine Binärvariable, welche angibt, ob sich ein Objekt in einem sozial benachteiligten „Brennpunkt“ befindet. Das Ergebnis zeigt uns, ob die Miete und Mietbelastung in solchen Gebieten stärker gestiegen ist als im Rest der Stadt. Danach verwenden wir ein Modell mit den eben gezeigten Clustern als erklärende Variablen. Wir erfahren dadurch, wie sich die Situation in den verschiedenen Clustern verändert hat und ob beispielsweise bestimmte Cluster oder Bevölkerungsgruppen stärker unter Mietpreissteigerungen gelitten haben. Zuletzt verwenden wir noch zwei Lokationskoeffizienten als erklärende Variablen, wodurch wir den Effekt von Konzentration von Arbeitslosen und Ausländern in einem Gebiet untersuchen können.

In allen Modellen verwenden wir die log-transformierte⁴ Kaltmiete und Mietbelastung als abhängige Variablen. Außerdem verwenden wir je das oberste und unterste 1% Quantil der Beobachtungen nicht, da wir bei diesen von Ausreißern ausgehen. Die Notwendigkeit einer Normalverteilung der Residuen für effiziente Durchführungen von Regressionsanalysen und insbesondere für t-Tests bildet die Grundlage dieser Überlegungen. Eine Logarithmutransformation sowie die Eliminierung von Ausreißern tragen dazu bei, dass die Residuen annähernd normalverteilt sind.

Durch diese Transformation verändert sich jedoch auch, wie die β -Werte interpretiert werden können. Das Regressionsmodell $\ln(y_i) = \beta_0 + \beta_1 \cdot x_i + \epsilon_i$ beispielsweise kann nun wie folgt interpretiert werden: Pro Einheit von x steigt y im Schnitt um $\exp(\beta_1) - 1$. Für kleine Werte von β_1 kann dieser Wert auch direkt interpretiert werden, beispielsweise gilt $\exp(0.05) - 1 = 5.13\% \approx 5\%$. Für die Konstante β_0 gilt: y nimmt den Wert $\exp(\beta_0)$ an, wenn $x = 0$. Außerdem sorgt die log-Transformation dafür, dass nicht nur rein lineare Zusammenhänge abgebildet werden können. Somit wird beispielsweise die Mietsteigerung über die Zeit nicht als linearer Zusammenhang modelliert. Dies wird in späteren Abbildungen deutlicher.

⁴Wir verwenden sowohl die Bezeichnungen log und ln. Mit beidem ist im Folgenden der natürliche Logarithmus gemeint.

Des Weiteren setzen definieren wir das Jahr 2007 als Jahr Null in den Regressionen. Wir haben dieses Jahr gewählt, da dort die Beobachtungen bei den Mietdaten beginnen. Dadurch kann die Konstante als „Wert im Jahr 2007“ statt „Wert zu Christi Geburt“ interpretiert werden, was für unsere Untersuchungen sinnvoller ist.

Für jedes der folgenden Unterkapitel haben wir eine auf den Forschungsfragen aus Kapitel 1.2 basierende Unterfrage formuliert, um intuitiv die Idee hinter den jeweils verwendeten Regressionsmodellen zu erklären.

4.2.1. Miete über Jahre

Um eine Grundlage für weitere, komplexere Regressionsmodelle zu legen, möchten wir mit einer einfachen Frage beginnen:

Wie haben sich Kaltmiete und Mietbelastung in Berlin im Laufe der Jahre durchschnittlich geändert?

Um diese zu beantworten, schätzen wir die folgenden Regressionsmodelle:

$$\begin{aligned}\ln(\text{Kaltmiete}_i) &= \beta_0 + \beta_1 \cdot \text{Jahr}_i + \epsilon_i \\ \ln(\text{Mietbelastung}_i) &= \beta_0 + \beta_1 \cdot \text{Jahr}_i + \epsilon_i\end{aligned}$$

Die Verwendung des Logarithmus zur Transformation der Kaltmiete und Mietbelastung wurde oben erläutert. Die Ergebnisse dieser Modelle sind in Tabelle 8 dargestellt und spiegelt das wider, was bereits im Boxplot in Abbildung 5 gezeigt wurde: Die Miete steigt im Durchschnitt pro Jahr. Konkret wissen wir nun, dass die Miete pro Jahr im Schnitt um $\exp(0.06) - 1 = 6.2\%$ steigt. Der Wert der Konstante β_0 lässt sich interpretieren, wenn er zurücktransformiert wird. Das Jahr Null wurde auf 2007 gesetzt, daher ergibt sich die Konstante, also die mittlere Kaltmiete im Jahr 2007, zu $\exp(1.701) = 5.48 \text{ €/m}^2$.

Für uns interessant ist außerdem die statistische Signifikanz. Diese kann anhand der Standardfehler – dargestellt unter dem Koeffizienten in Klammern – bzw. der Anzahl an Sternchen abgelesen werden. Für die Variable Jahr beispielsweise ist der Standardfehler um ein Vielfaches kleiner als der geschätzte Koeffizient. Daher und durch die hohe Anzahl an Beobachtungen (knapp 1.9 Mio.) liegt auch der p-Wert, angegeben durch die Sternchen, nahe Null. Dieser p-Wert lässt sich im Allgemeinen interpretieren als die Wahrscheinlichkeit, dass der berechnete Koeffizient (in diesem Fall 0.060 für die Variable Jahr) in Wahrheit eigentlich Null ist. Anders ausgedrückt handelt es sich um die Wahrscheinlichkeit, dass der gefundene Zusammenhang nur durch Zufall entstanden ist. Wir können also sehr sicher sein, dass es tatsächlich eine Steigerung der Mieten über die Jahre gab. Neben dem Effekt von Jahr auf die Kaltmiete ist auch die Konstante (stark) statistisch signifikant. Interessant ist außerdem das für dieses kleine – im Sinne von wenig Parametern –

<i>Abhängige Variable</i>		
	log Kaltmiete ($\text{€}/\text{m}^2$)	log Mietbelastung (%)
	(1)	(2)
Konstante	1.701*** (0.0003)	-2.056*** (0.001)
Jahr	0.060*** (0.0001)	0.058*** (0.0001)
Beobachtungen	1,878,609	1,878,430
R ²	0.414	0.164

Anmerkung: *p<0.1; **p<0.05; ***p<0.01

Tabelle 8: Regression Miete über Zeit

Regressionsmodell hohe R² von 0.4. Das bedeutet, dass knapp 40% der Variation in der Höhe der Miete allein durch das Betrachten des Jahres erklärt werden kann.

Wichtig ist außerdem, neben der statistischen Signifikanz auch auf die ökonomische Relevanz des Zusammenhangs zu achten. Eine Steigerung von knapp 6% pro Jahr ist jedoch auch ökonomisch bedeutsam, liegt diese doch weit über dem zu dieser Zeit vorherrschenden allgemeinen Inflationsniveau bei Wohnungsmieten von unter 2% (Statistisches Bundesamt 2024b). Außerdem stellt die (Nettokalt)Miete eine der größten Kostenblöcke in den Ausgaben von Personen dar und ist im Verbraucherpreisindex 2023 mit knapp 17% gewichtet (Statistisches Bundesamt 2024d).

Betrachtet man die Regression zur Mietbelastung, so sieht man auch dort einen steigenden Trend. Die Konstante zeigt eine Mietbelastung im Jahr 2007 von knapp $\exp(-2.056) = 12.8\%$. Pro Jahr steigt dieser Wert dann im Durchschnitt um $\exp(0.05784) - 1 = 6\%$. Es ist zu beachten, dass es sich dabei um eine prozentuale Änderung der Mietbelastung handelt, nicht um eine Änderung in Prozentpunkten. Im Jahr 2008 liegt laut dem Modell die Mietbelastung also bei $12.8\% \cdot (1 + 6\%) = 13.6\%$. 2020 – also 13 Jahre später – liegt die Mietbelastung dann schon bei knapp $12.8\% \cdot (1 + 6\%)^{13} \approx 27\%$, was mehr als einer Verdopplung entspricht. Zusätzliche Steigerungen der Warmmiete durch höhere Nebenkosten – insbesondere in den letzten Jahren – sind dabei noch gar nicht enthalten. Der gezeigte Zusammenhang ist demnach ökonomisch höchst relevant.

Auch für das Modell zur Mietbelastung sind sowohl die Konstante als auch der Effekt der Variable Jahr stark statistisch signifikant. Im Gegensatz zum Modell mit der Kaltmiete ist jedoch der Wert des R² mit 0.16 geringer. Dies liegt vermutlich daran, dass uns zur Berechnung der Mietbelastung nur das durchschnittliche Einkommen pro Haushalt für je ein 1x1km Quadrat zur Verfügung steht. Dadurch entstehen teilweise auch wenig sinnvolle Werte bei der Mietbelastung, beispielsweise Werte von über 100%. Generell ist hier die Schwankung höher, was sich in einem niedrigeren R² Wert zeigt. Auf die Darstellung des korrigierten (engl.: adjusted) R² haben wir bei allen Regressionen verzichtet. Während das „normale“ R² mit einem größeren Modell auto-

matisch wächst, enthält das korrigierte R^2 – wie der Name vermuten lässt – einen Korrekturterm gegen diesen Effekt. Aufgrund der schieren Anzahl an Beobachtungen und den relativ kleinen Modellen, die wir verwenden, sind beide jedoch fast identisch und eine Darstellung beider bringt keinen Mehrwert.

Anschließend wollen wir noch die Robustheit der Regression überprüfen, um sicherzugehen, dass der t-Test überhaupt ein sinnvolles Ergebnis liefert. Dafür betrachten wir zuerst die Verteilung der Residuen, indem wir die Residuen auf die y-Achse und die gefitteten, also vom Modell vorhergesagten Werte auf die x-Achse abtragen. Abbildung 16 zeigt dies für beide Regressionsmodelle.

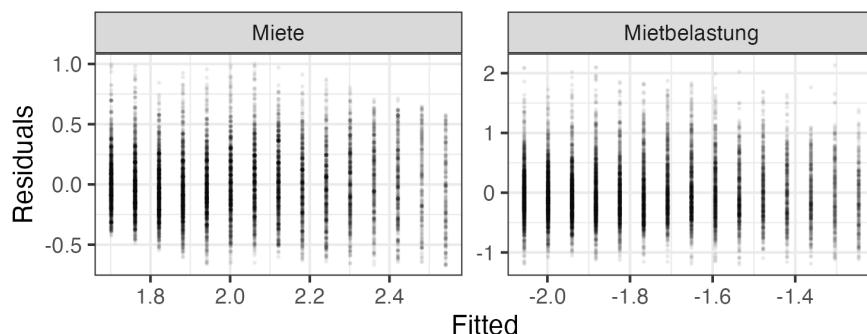


Abbildung 16: Residuen zur Regression der Miete über Zeit

In einem perfekten Regressionsmodell würden alle Punkte auf der horizontalen bei Null liegen. Das würde bedeuten, dass das Regressionsmodell alle Werte korrekt vorhersagt und es zu keinen Abweichungen kommt. Für uns ist jedoch nur wichtig, dass diese Abweichungen von Null relativ gleichmäßig verteilt sind. Dies scheint ziemlich gut der Fall zu sein. Außerdem sollte die Streuung konstant bleiben. Auch das lässt sich beobachten. Es ist wichtig zu beachten, dass es sich hier um log-transformierte Werte handelt, die so nicht direkt interpretierbar sind.

Ein auffälliger Aspekt ist die rasterartige Verteilung der Punkte entlang der Fitted-Achse. Diese resultiert aus der Spezifikation des Regressionsmodells, welches, abgesehen von der Binärvariablen, ausschließlich Jahresangaben als Eingangsgröße nutzt. Daraus ergibt sich eine Abfrage sehr regelmäßiger Werte. Würden hingegen halbe Jahre, Quartale oder Monate als Zeitangaben verwendet, blieben die Wertabstände zwar regelmäßig, wären aber feiner abgestuft.

Insgesamt sind keine Probleme mit dem Regressionsmodell erkennbar. Die Residuen sind symmetrisch um Null verteilt und es ist keine Heteroskedastizität – eine ungleichmäßige Residualvarianz – erkennbar. Somit können wir den Ergebnissen der t-Tests vertrauen.

Wir können nun die Frage vom Beginn dieses Kapitels beantworten: Sowohl die Kaltmiete als auch die Mietbelastung sind über die Jahre stark gestiegen, wobei die Kaltmiete einen etwas stärkeren Anstieg zeigt als die Mietbelastung. Außerdem können wir bereits die dritte unserer

Forschungsfragen teilweise beantworten. Zwar haben wir hier noch keine Entwicklung für spezifische Teile Berlins bzw. Bevölkerungsgruppen betrachtet, wir können aber schon festhalten, dass die Mietbelastung, also das Verhältnis von Einkommen zu Miete von 2009 bis 2019 im Mittel stark angestiegen ist. Im Folgenden wollen wir nun versuchen, diese Aussage zu präzisieren.

4.2.2. Miete in Brennpunkten

Nachdem wir im vorherigen Abschnitt gesehen haben, wie sich die Mieten in ganz Berlin über die Jahre geändert haben, wollen wir nun sozial benachteiligte Gebiete gesondert betrachten. Wir stellen also die Frage:

Hat sich die Miete und Mietbelastung in sozialen Brennpunkten über die Jahre stärker erhöht als in anderen Gebieten?

Um eine klare Basis für unsere Diskussion zu schaffen, ist es zunächst notwendig, den Begriff „sozialer Brennpunkt“ zu definieren. Ein Gebiet gilt nach unserer Definition als sozialer Brennpunkt, wenn es im Jahr 2015 eine Arbeitslosenquote von über 10% aufwies und das mittlere Einkommen zugleich im untersten 20%-Quantil im Vergleich zum Rest der Stadt lag. Wir haben diese Werte gewählt, da so eine „angemessene“ Anzahl an Brennpunkten entsteht. Zu wenige und zu viele würden die Regression bzw. den Vergleich zu nicht-Brennpunkten schwieriger machen. Wir gehen für die Regressionen davon aus, dass diese Gebiete über die Zeit stabil bleiben, was uns durch die Stabilität der Cluster (Abbildung 15) plausibel erscheint. Gemäß dieser Kriterien erfüllen 120 der 870 untersuchten Gebiete, was einem Anteil von 14%, die Kriterien eines sozialen Brennpunkts. Eine visuelle Darstellung dieser Gebiete auf einer Karte findet sich in Abbildung 17.

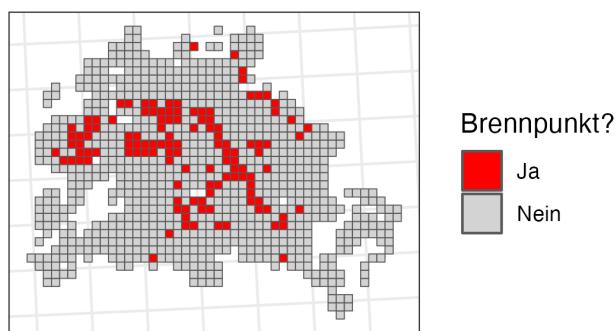


Abbildung 17: Brennpunkte 2015

Wir verwenden ein ähnliches Regressionsmodell wie im vorherigen Abschnitte, nur fügen wir noch die Dummyvariable „Brennpunkt“ sowie einen Interaktionsterm mit dem Jahr hinzu:

$$\ln(\text{Kaltmiete}_i) = \beta_0 + \beta_1 \cdot \text{Jahr}_i + \beta_2 \cdot \text{Brennpunkt}_i + \beta_3 \cdot \text{Brennpunkt}_i \cdot \text{Jahr}_i + \epsilon_i$$

$$\ln(\text{Mietbelastung}_i) = \beta_0 + \beta_1 \cdot \text{Jahr}_i + \beta_2 \cdot \text{Brennpunkt}_i + \beta_3 \cdot \text{Brennpunkt}_i \cdot \text{Jahr}_i + \epsilon_i$$

Die Konstante β_0 , transformiert mit $\exp(\cdot)$, repräsentiert die durchschnittliche Kaltmiete oder Mietbelastung im Basisjahr 2007 für nicht-Brennpunktgebiete. Der Koeffizient β_1 , angepasst durch $\exp(\cdot) - 1$, gibt die durchschnittliche jährliche prozentuale Mietsteigerung in nicht-Brennpunktgebiete an. Den Unterschied im Jahr 2007 zwischen Brennpunkten und nicht-Brennpunkten zeigt β_2 . Schließlich reflektiert β_3 , ebenfalls nach Rücktransformation mit $\exp(\cdot) - 1$, die Differenz in der jährlichen Mietpreisentwicklung zwischen Brennpunkt- und Nicht-Brennpunktgebieten ab dem Basisjahr. Die Werte sind in Tabelle 9 dargestellt.

	Abhängige Variable	
	log Kaltmiete (€/m ²)	log Mietbelastung (%)
	(1)	(2)
Konstante	1.733*** (0.0004)	-2.049*** (0.001)
Jahr	0.057*** (0.0001)	0.056*** (0.0001)
Brennpunkt	-0.101*** (0.001)	-0.025*** (0.001)
Jahr*Brennpunkt	0.008*** (0.0001)	0.008*** (0.0002)
Beobachtungen	1,878,430	1,878,430
R ²	0.421	0.165

Anmerkung:

*p<0.1; **p<0.05; ***p<0.01

Tabelle 9: Regression Miete in Brennpunkten

Für nicht-Brennpunktgebiete liegt die Kaltmiete im Jahr 2007 bei $\exp(1.733) = 5.66 \text{ €/m}^2$ und steigt pro Jahr um durchschnittlich $\exp(0.057) - 1 = 5.9\%$. Handelt es sich um ein Brennpunktgebiet, so liegt die Miete im Jahr 2007 um $\exp(-0.101) - 1 = -9.6\%$ (Achtung, nicht Prozentpunkte) niedriger bei $\exp(1.733 - 0.101) = 5.11 \text{ €/m}^2$. Von diesem niedrigeren Niveau steigt die Miete dann um $\exp(0.008) - 1 = 0.8\%$ pro Jahr stärker. Alle diese Effekte sind stark statistisch signifikant, mit p-Werten nahe null. Die Mietbelastung für nicht-Brennpunktgebiete liegt im Jahr 2007 bei $\exp(-2.049) = 12.9\%$ und steigt pro Jahr um $\exp(0.056) - 1 = 5.8\%$ (Achtung, nicht Prozentpunkte). In Brennpunktgebieten liegt die Mietbelastung zu Beginn etwas niedriger bei $\exp(-2.049 - 0.025) = 12.6\%$, steigt dann aber um $\exp(0.008) - 1 = 0.8\%$ stärker.

Der R²-Wert der beiden Modelle beträgt 42.1% bzw. 16.5%, was nur eine geringfügige Verbesserung gegenüber den Ergebnissen aus Tabelle 8 des vorherigen Abschnitts darstellt. Dies deutet darauf hin, dass die Einbeziehung der Dummyvariable „Brennpunkt“ nur einen begrenzten Beitrag zur Erklärung der Variabilität in den Daten leistet. Trotzdem liefern die Ergebnisse dieser Regression wichtige Erkenntnisse, da unser Ziel nicht die Schaffung eines möglichst präzisen Vorhersagemodells ist, sondern das Verständnis sozial relevanter Zusammenhänge in den Daten.

Um die Interpretierbarkeit der Ergebnisse aus den beiden Regressionsmodellen zu verbessern,

präsentieren wir in Abbildung 18 eine Visualisierung der durch die Modelle vorhergesagten Werte. Diese grafische Darstellung ermöglicht eine anschauliche Bewertung der Modellanpassung, indem sie die gefitteten Werte direkt mit den beobachteten Daten vergleicht und somit einen Einblick in die Aussagen der Modelle bietet.

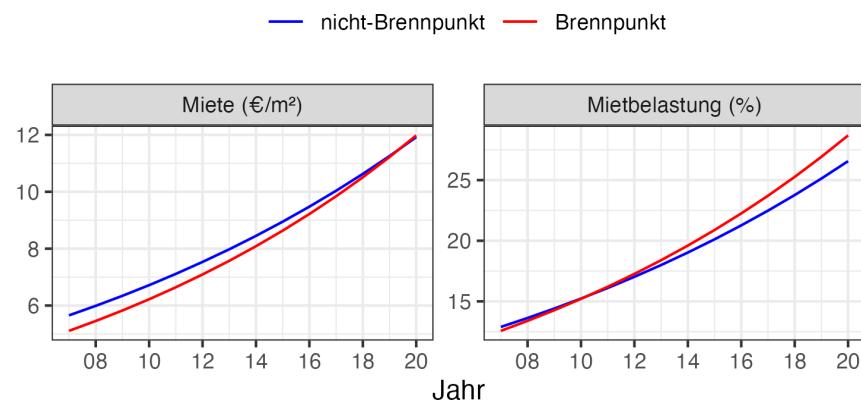


Abbildung 18: Modell Regression Brennpunkte

Es ist von Bedeutung, die Qualität der durchgeführten Regression zu evaluieren, analog zu den im vorangegangenen Abschnitt beschriebenen Verfahren. In diesem Kontext illustriert Abbildung 23 die Verteilung der Residuen (y-Achse) im Verhältnis zu den prognostizierten Werten (x-Achse). Aufgrund der Tatsache, dass der Graph keine signifikanten Unregelmäßigkeiten aufweist, ähnlich den Beobachtungen des vorherigen Abschnitts, wurde er in Anhang A verschoben.

Zum Schluss können wir jetzt die Fragen vom Beginn dieses Abschnitts beantworten. Die Ergebnisse der Regressionsanalyse deuten darauf hin, dass sich die Miete und Mietbelastung in sozialen Brennpunkten im Vergleich zu anderen Gebieten seit 2007 unterschiedlich entwickelt haben. Speziell zeigt der Koeffizient β_3 , der den Interaktionsterm zwischen dem Status als sozialer Brennpunkt und dem Jahr repräsentiert, dass die jährliche Steigerungsrate der Miete bzw. Mietbelastung in sozialen Brennpunkten signifikant von der in Nicht-Brennpunktgebieten abweicht. Dies impliziert, dass die Dynamik der Miet- und Mietbelastungsentwicklung in sozialen Brennpunkten über die Jahre hinweg eine besondere Charakteristik aufweist, die sie von anderen Gebieten unterscheidet. Diese Erkenntnis ist auch in Hinblick auf unsere dritte Forschungsfrage relevant. Wir wissen jetzt, dass der ärmere Teil der Bevölkerung unter den Mietpreisentwicklungen stärker gelitten hat.

4.2.3. Miete in den Clustern

Wir haben bereits mit den Brennpunkten eine erste Klassifizierung innerhalb der Stadt vorgenommen. Nun wollen wir einen Schritt weiter gehen und die in Kapitel 4.1.2 hergeleiteten Cluster für

eine Regressionsanalyse verwenden, um die folgenden Fragen zu beantworten:

Wie haben sich Miete und Mietbelastung in den verschiedenen Clustern entwickelt? Gibt es relevante Unterschiede zwischen den Clustern?

Dafür werden wir die folgenden Regressionsmodelle verwenden:

$$\ln(\text{Kaltmiete}_i) = \beta_0 + \beta_1 \cdot \text{Jahr}_i + \beta_2 \cdot \text{Cluster}_i + \beta_3 \cdot \text{Cluster}_i \cdot \text{Jahr}_i + \epsilon_i$$

$$\ln(\text{Mietbelastung}_i) = \beta_0 + \beta_1 \cdot \text{Jahr}_i + \beta_2 \cdot \text{Cluster}_i + \beta_3 \cdot \text{Cluster}_i \cdot \text{Jahr}_i + \epsilon_i$$

Während diese Regressionsmodelle auf den ersten Blick denen aus dem vorherigen Abschnitt sehr ähnlich sehen, gibt es doch einen relevanten Unterschied: Statt einer Binärvariable wird nun mit „Cluster“ eine kategorische Variable mit 4 Leveln verwendet. Dabei wird das erste Level, also Cluster 1, zum Basiscase, ähnlich wie es die nicht-Brennpunktgebiete in der vorherigen Regression waren. Es empfiehlt sich, die Karte der Cluster aus Abbildung 13 im Kopf zu behalten. Der Wert für Cluster 1 lässt sich also an der Konstante und dem Koeffizienten vor Jahr β_1 ablesen. Tabelle 10 zeigt die Ergebnisse der beiden Regressionen.

	Abhängige Variable	
	$\log \text{Kaltmiete} (\text{€}/\text{m}^2)$	
	(1)	(2)
Konstante	1.638*** (0.003)	-1.785*** (0.006)
Jahr	0.057*** (0.0004)	0.009*** (0.001)
Cluster2	-0.032*** (0.003)	-0.121*** (0.006)
Cluster3	0.215*** (0.004)	-0.181*** (0.007)
Cluster4	0.058*** (0.003)	-0.025*** (0.006)
Jahr*Cluster2	0.005*** (0.0004)	0.018*** (0.001)
Jahr*Cluster3	-0.011*** (0.001)	0.031*** (0.001)
Jahr*Cluster4	0.021*** (0.0004)	0.041*** (0.001)
Beobachtungen	1,357,293	1,357,293
R ²	0.403	0.089

Anmerkung:

*p<0.1; **p<0.05; ***p<0.01

Tabelle 10: Regression Cluster

Um das Verständnis der Ergebnisse aus den beiden Modellen zu vereinfachen, haben wir in Ab-

bildung 19 wieder die gefitteten Werte für die Jahre 2007 bis 2020 in Verbindung mit den vier Clustern abgetragen. Wir verzichten daher auch auf eine genauere Erläuterung des Regressionsoutputs. Einerseits verhält sich dies wie in den Kapiteln zuvor, andererseits scheint uns eine grafische Darstellung einfacherer und intuitiver zu interpretieren. Die Abbildung zeigt deutlich, wie sich Cluster 4 sowohl bei der Miete als auch bei der Mietbelastung über die Jahre von den anderen Clustern entfernt und sich bezogen auf die Mietbelastung fast verdoppelt. Die reine Miete zeigt sogar eine Verdreifachung. Dahingegen zeigt Cluster 1 bei der Mietbelastung eine Veränderung von der – nach Mietbelastung – teuersten zur günstigsten Gegend. Interessanterweise zeigt die Mietbelastung eine stärkere Auffächerung über die Zeit als die reine Miete. Das deutet darauf hin, dass in manchen Clustern (insbesondere in 1) das Einkommen ungefähr gleich stark gestiegen ist, während dies anderswo nicht der Fall war.

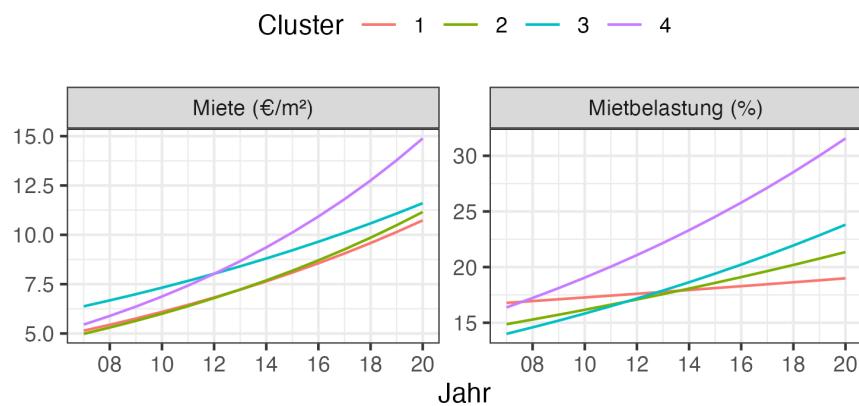


Abbildung 19: Modell Regression Cluster

Wie bereits im vorherigen Abschnitt haben wir die Abbildung 24, auf welcher wir die Residuen gegen die gefitteten Werte abtragen, um die Spezifikation des Regressionsmodells zu prüfen, in den Anhang A verschoben. Die Abbildung zeigt keine unerwarteten Unregelmäßigkeiten.

Nun können wir wieder die Fragen vom Beginn dieses Abschnitts beantworten. Dabei wollen wir auch kurz den Bezug zu der Interpretation der Cluster herstellen. Ausführlich werden wir die Verbindung im Kapitel 5 diskutieren. Die Miete und Mietbelastung entwickeln sich in den Clustern unterschiedlich. Wohlhabendere Gebiete (Cluster 3) starten mit hoher Miete, aber geringerem Anstieg, während sozial herausgeforderte Gebiete (Cluster 2) und multikulturelle Gebiete (Cluster 4) mit niedriger Miete starten und stärker steigen. Stadtrandlagen (Cluster 1) haben hohe anfängliche Mietbelastung, die am langsamsten steigt. Es gibt also signifikante Unterschiede in der Dynamik der Miet- und Mietbelastungsentwicklung zwischen den Clustern. Mit Hinblick auf unsere dritte Forschungsfrage sind diese Ergebnisse höchst relevant und werden später noch ausführlicher beleuchtet.

4.2.4. Lokationskoeffizient

Der Lokationskoeffizient gibt an, wie stark konzentriert eine Gruppe in einem Gebiet im Vergleich zu der (in unserem Fall) gesamten Stadt ist. Wenn wir diese Kennzahl in unsere Analyse einbeziehen, können wir die folgende Frage beantworten:

Hängen Miete und Mietbelastung mit der Konzentration der Arbeitslosen- und Ausländerquote zusammen? Wie hat sich dieser Zusammenhang über die Zeit verändert?

Dafür verwenden wir die folgenden beiden Regressionsmodelle, bei denen wir die stetigen Lokationskoeffizienten (LQ) für den Anteil an Arbeitslosen und Ausländern sowie je Interaktionsterme für mit dem Jahr einfügen:

$$\ln(\text{Kaltmiete}_i) = \beta_0 + \beta_1 \cdot \text{Jahr}_i + \beta_2 \cdot \text{LQ Arbeitslose}_i + \beta_3 \cdot \text{LQ Ausländer}_i \\ + \beta_4 \cdot \text{LQ Arbeitslose}_i \cdot \text{Jahr}_i + \beta_5 \cdot \text{LQ Ausländer}_i \cdot \text{Jahr}_i + \epsilon_i$$

$$\ln(\text{Mietbelastung}_i) = \beta_0 + \beta_1 \cdot \text{Jahr}_i + \beta_2 \cdot \text{LQ Arbeitslose}_i + \beta_3 \cdot \text{LQ Ausländer}_i \\ + \beta_4 \cdot \text{LQ Arbeitslose}_i \cdot \text{Jahr}_i + \beta_5 \cdot \text{LQ Ausländer}_i \cdot \text{Jahr}_i + \epsilon_i$$

Die Koeffizienten von „LQ Arbeitslose“ und „LQ Ausländer“ (β_2 und β_3) messen den Effekt einer Veränderung im jeweiligen Lokationskoeffizienten auf die (log) Miete. Die Koeffizienten der Interaktionsterme (β_4 und β_5) bildet zusätzlich den Einfluss des Jahres ab. Tabelle 11 zeigt die Ergebnisse der beiden Modelle. Insgesamt ist die Interpretation dieser Modelle rein anhand der Koeffizienten wenig intuitiv, weswegen wir nur kurz darauf eingehen werden und insbesondere die Richtung der Effekte anhand der Vorzeichen betrachten.

Die Ergebnisse unserer Regressionsanalysen zeigen statistisch signifikante Zusammenhänge zwischen der Konzentration von Arbeitslosigkeit und Ausländern und den Mietpreisen sowie Mietbelastungen in Berlin. Für das erste Modell zeigt uns die Konstante, dass im Jahr 2007 und in einem (fiktiven) Gebiet ohne Arbeitslose oder Ausländer die Kaltmiete bei $\exp(2.032) = 7.63\text{€}/\text{m}^2$ liegt. Die Mietbelastung würde bei $\exp(-1.743) = 17.5\%$ liegen. Wenig verwunderlich hat der Koeffizient von Jahr wieder ein positives Vorzeichen, die Miete und Mietbelastung steigen also pro Jahr um knapp 4 bzw. 5%.

Interessanter sind nun die Koeffizienten von „LQ Arbeitslose“ und „LQ Ausländer“. Die negativen Vorzeichen zeigen, dass die Miete und auch Mietbelastung tendenziell sinken, wenn die Konzentration von Arbeitslosen und Ausländern in einem Gebiet ansteigt. Bei der Regression zur Mietbelastung ist der (negative) Effekt der Arbeitslosen knapp doppelt so hoch wie der von Ausländern. Bei der Regression zur Miete beträgt er mehr als das 15-fache. Auch wenn überall statistische Signifikanz gegeben ist, scheint es so, als habe die Konzentration von Arbeitslosen einen weitaus größeren Effekt auf Mietpreise als die Konzentration von Ausländern dies hat. Zu einem kleine-

	<i>Abhängige Variable</i>	
	log Kaltmiete (€/m ²)	log Mietbelastung (%)
	(1)	(2)
Konstante	2.032*** (0.002)	-1.743*** (0.003)
Jahr	0.041*** (0.0002)	0.052*** (0.0005)
LQ Arbeitslose	-0.332*** (0.001)	-0.075*** (0.003)
LQ Ausländer	-0.016*** (0.001)	-0.044*** (0.001)
LQ Arbeitslose*Jahr	0.008*** (0.0002)	-0.035*** (0.0004)
LQ Ausländer*Jahr	0.016*** (0.0001)	0.021*** (0.0002)
Beobachtungen	1,357,298	1,357,298
R ²	0.428	0.085

Anmerkung:

*p<0.1; **p<0.05; ***p<0.01

Tabelle 11: Regression LQ

ren Teil stimmt dies auch für die Mietbelastung. Positive Vorzeichen bei den Interaktionstermen zeigen, dass ein höherer LQ-Wert zu einer stärkeren Mietsteigerung bzw. Mietbelastung über die Jahre führt. Die R²-Werte liegen mit knapp 40% und 9% ungefähr auf dem Niveau der Cluster-Regression (Tabelle 10).

Diese Regressionsmodelle zu visualisieren ist schwieriger als dies bei den vorherigen Modellen der Fall war, da es sich nun um drei (mehr oder weniger) stetige erklärende Variablen handelt. Eine vollständige Darstellung ist also nicht möglich. Wir haben uns daher entschieden, je zwei Werte – 0.2 und 1.2 – für „LQ Ausländer“ und „LQ Arbeitslose“ zu verwenden. Diese stellen ungefähr das 10% und 90% Quantil der Verteilung dar. Abbildung 20 zeigt diese Darstellung. Die rote durchgezogene Linie stellt also ein Gebiet dar, in dem die Konzentration von Ausländern hoch, die Konzentration von Arbeitslosen jedoch niedrig ist, und so weiter. Am Auffälligsten bei der Be trachtung der Mietbelastung ist, dass diese in Gebieten mit hoher Konzentration von Arbeitslosen nur sehr wenig gestiegen ist, während sie in Gebieten mit niedriger Arbeitslosenkonzentration gerade recht explodiert ist. Während die Mietpreise zu Beginn des Beobachtungszeitraums noch sehr eng beieinander lagen, hat sich die Spanne im Laufe der Zeit deutlich vergrößert. Bei der Miete ist keine so deutliche Divergenz sichtbar. Hier ist jedoch ein größerer Unterschied in der absoluten Höhe zu Beginn erkennbar, je nachdem wie hoch der Anteil von Arbeitslosen in einem Gebiet ist. Der absolute Unterschied zwischen einem hohen und niedrigen Ausländeranteil bleibt über die Zeit ziemlich stabil.

Wir überprüfen noch die Güte der Regression anhand der Residuen. Anders als bei den vorheri-

LQ-Ausländer — 0.2 — 1.2 LQ-Arbeitslose — 0.2 - - 1.2

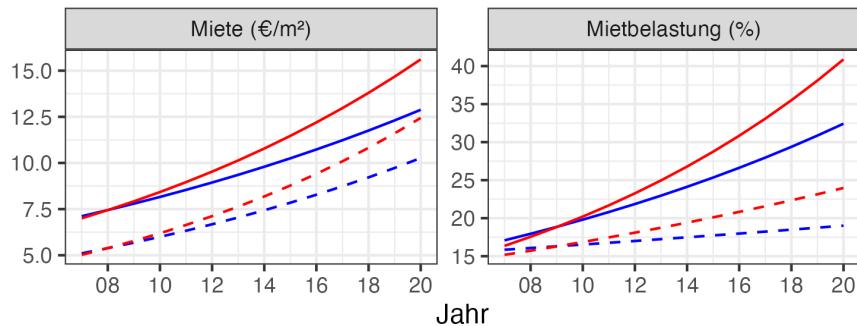


Abbildung 20: Modell Regression LQ

gen Regressionsmodellen sind die Residuenplots hier interessanter. Es zeigt sich nämlich keine so starke Regelmäßigkeit, wahrscheinlich durch die zusätzliche Betrachtung der stetigen Variablen „LQ Ausländer“ und „LQ Arbeitslose“. Die Residuen sind gleichmäßig um null verteilt, nur bei der Mietbelastung zeigen sich leichte Anzeichen von Heteroskedastizität, also einer ungleichmäßigen Streuung für verschiedene Input-Werte. Dies könnte zu verzerrten Standardfehlern und t-Statistiken führen. Allerdings ist der Effekt hier nur leicht ausgeprägt und die Standardfehler sind alle so klein, dass wir hier keine weiteren Maßnahmen (wie beispielsweise nicht-parametrische Tests) für nötig halten.

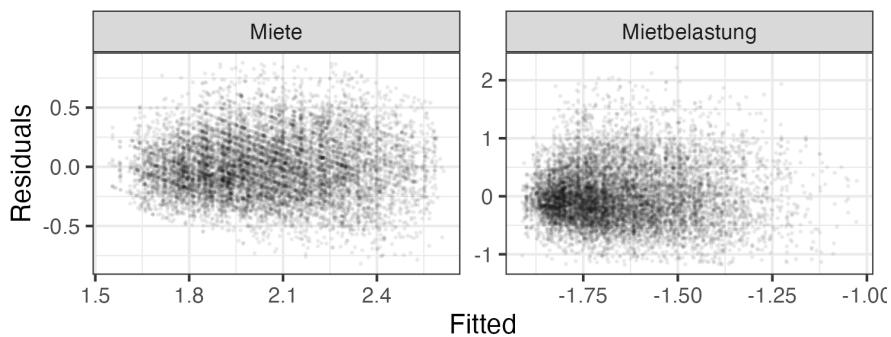


Abbildung 21: Residuen Regression Cluster

Betrachten wir wieder die Frage vom Beginn dieses Kapitels, können wir den ersten Teil mit einem klaren „Ja“ beantworten. Abbildung 20 zeigt uns, dass es definitiv einen Zusammenhang zwischen der Miete/Mietbelastung und den LQ-Werten gibt. Die Frage nach der Veränderung dieses Zusammenhangs über die Zeit lässt sich hingegen nicht so einfach beantworten. Grundsätzlich bleibt der Zusammenhang ähnlich, wenngleich er für die Miete und Mietbelastung unterschiedlich stark auseinanderdriftet. Mit Bezug auf unsere dritte Forschungsfrage könnte man sagen, dass insbe-

sondere Gebiete mit einem niedrigen Arbeitslosenanteil und gleichzeitig hohen Ausländeranteil am stärksten unter der Steigerung der Mietbelastung gelitten haben.

5. Diskussion

In diesem Kapitel erfolgt eine kritische Bewertung der erzielten Ergebnisse, um deren Aussagekraft und mögliche Interpretationen zu erörtern. Es wird untersucht, inwiefern die Ergebnisse die Forschungsfragen beantworten und welche Einsichten sie für das Studienfeld bieten. Zudem werden die Limitationen der Studie beleuchtet, die die Reichweite der Ergebnisinterpretation einschränken könnten. Diese kritische Reflexion schließt mit Überlegungen ab, wie zukünftige Forschungsarbeiten die identifizierten Lücken adressieren und die Methodik erweitern könnten, um ein umfassenderes Verständnis des Themas zu erlangen.

5.1. Interpretation der Ergebnisse

Um eine tiefgreifende und differenzierte Interpretation der Ergebnisse bezüglich der räumlichen Segregation und der Entwicklungen auf dem Berliner Wohnungsmarkt zu gewährleisten, integrieren wir die umfassenden Analysen der Konzentration von Bevölkerungsgruppen, der Clusteranalyse und der verschiedenen Regressionsanalysen. Diese kombinierte Betrachtungsweise ermöglicht es uns, die komplexen Dynamiken und deren Implikationen für die sozioökonomische Struktur Berlins zu verstehen.

Die Untersuchung der Konzentration von Bevölkerungsgruppen auf Basis des Lokationsquotienten (LQ), insbesondere in Bezug auf Arbeitslosenquote, Kaufkraft und Ausländerquote, offenbart signifikante räumliche Disparitäten, die als Indikatoren für Segregationstendenzen dienen. Hierbei ist besonders die Kaufkraft pro Haushalt hervorzuheben, deren Interpretation durch die Haushaltsgröße kompliziert wird – ein wichtiger Aspekt, der die direkte Zuordnung der Kaufkraft zu individuellen wirtschaftlichen Verhältnissen erschwert. Des Weiteren ist bei Betrachtung der Abbildung 9 ein leicht rückläufiger Trend der Segregation hinsichtlich der Ausländerquote zu beobachten.

Die darauf folgende Clusteranalyse dient der weiteren Differenzierung und ermöglicht es, Gebiete mit ähnlichen Merkmalsausprägungen zu identifizieren. Diese Analyse stützt und plausibilisiert die Ergebnisse der Konzentrationsbetrachtung, indem sie zeigt, dass Gebiete mit ähnlichen sozioökonomischen Charakteristiken räumlich zusammenhängen, was auf Segregationstendenzen hinweist. Auffällig hierbei ist die räumliche Trennung zwischen zentralen Gebieten und Randbezirken, die unterschiedliche sozioökonomische Profile aufweisen, sowie die besondere Rolle von Spandau und Marzahn, die aus dem allgemeinen Muster herausfallen. Über die Jahre hinweg verändern sich die Größen der identifizierten Cluster zwar, aber Veränderungen in der Durchmischung der Bevölkerung bleiben minimal. Es bleibt die Herausforderung, aus der Clusterbildung direkte Rückschlüsse auf einzelne Variablen zu ziehen, da die Cluster aus mehreren Variablen zusammengesetzt sind.

Die Regressionsanalyse mit Brennpunkten und Clustern verdeutlicht, dass Miete und Mietbelas-

tung insbesondere in sozialen Brennpunkten und in bestimmten Clustern (Cluster 4), stärker angestiegen sind als in anderen Bereichen der Stadt. Diese Ergebnisse unterstreichen die zunehmende finanzielle Belastung, der bestimmte Bevölkerungsschichten ausgesetzt sind. Besonders auffällig ist bei der Cluster-Regression, dass die Cluster, die durch eine höhere sozioökonomische Positionierung gekennzeichnet sind (Cluster 1 – Stadtrandlagen), tendenziell eine stabilere oder gar günstigere Entwicklung der Mietbelastung aufweisen, während Gebiete mit niedrigeren sozioökonomischen Indikatoren (Cluster 4 – Multikulturelle Gebiete) stärkeren Schwankungen und Steigerungen ausgesetzt sind. Die Analyse der Lokationskoeffizienten bietet zudem Einblicke in die Wechselwirkung zwischen der Konzentration von Arbeitslosigkeit und Ausländern und den Wohnkosten, wobei festzustellen ist, dass Gebiete mit hoher Arbeitslosigkeit im Vergleich eine günstigere Entwicklung erfahren haben.

Ein kritischer Punkt in den Ergebnissen ist die scheinbare Diskrepanz zwischen den Erkenntnissen der LQ-Regressionsanalyse und der Analyse der sozialen Brennpunkte, welche auf den ersten Blick widersprüchlich erscheinen. Während soziale Brennpunkte, definiert durch niedriges Einkommen und hohe Arbeitslosenquote, eine stärkere Steigerung der Mietbelastung erlebt haben, suggeriert die LQ-Regressionsanalyse, dass Gebiete mit hoher Arbeitslosigkeit weniger stark betroffen waren. Diese Diskrepanz bedarf einer sorgfältigen Kontextualisierung und weiterführenden Analyse, um die zugrundeliegenden Mechanismen und deren Auswirkungen auf die Wohnkostenbelastung zu verstehen.

Zusammenfassend lässt sich sagen, dass die umfassenden Analysen tiefgreifende Einblicke in die Struktur und die Dynamiken des Berliner Wohnungsmarkts sowie die damit verbundenen sozioökonomischen Segregationsprozesse liefern. Sie zeigen, dass bestimmte Gebiete und Bevölkerungsgruppen deutlich stärker von den Entwicklungen betroffen sind als andere, was die Notwendigkeit gezielter politischer Maßnahmen und Interventionen zur Förderung einer ausgeglichenen und gerechteren städtischen Entwicklung unterstreicht.

5.2. Limitation der Arbeit

Die wissenschaftliche Auseinandersetzung mit der Thematik der Segregation stößt auf diverse Limitationen, die sowohl quantitative als auch qualitative Aspekte der Forschung betreffen. Diese Einschränkungen sind vielschichtig und bedürfen einer differenzierten Betrachtung, um das Verständnis für die Komplexität des Untersuchungsgegenstands zu vertiefen.

Eine wesentliche Limitation der Arbeit ist der Umstand, dass die analysierten Daten auf dem Wohnort bzw. dem Ort basieren, an dem Menschen schlafen. Der Wohnort allein erlaubt keine genauen Rückschlüsse darauf, wie Menschen die Stadt tatsächlich nutzen. Da Personen ihre Zeit nicht ausschließlich an ihrem Wohnort verbringen, sondern sich frei innerhalb der Stadt bewegen können, entstehen sozialen Durchmischungen. In diesem Zusammenhang spielt der öffentliche Nahverkehr (ÖPNV) eine maßgebliche Rolle. Ärmere Gegenden weisen oft eine schlechtere Anbin-

dung an den ÖPNV auf (Blokland und Vief 2021). Diese Ungleichheit in der Verkehrsinfrastruktur verstärkt möglicherweise die Segregation durch Abschottung bestimmter Wohngebiete.

Ein Vergleich der Ergebnisse aus der Betrachtung anhand einzelner Variablen und der Clusteranalyse erfordert eine differenzierte Betrachtung der unterschiedlichen methodischen Ansätze. Für die Analyse der Konzentrationen wurden Gebiete anhand des Medians der Gesamtstadt in Relation gesetzt. Im Gegensatz dazu erlaubt die Clusteranalyse keine direkten Schlüsse auf durchschnittliche Gebiete, sondern basiert eher auf den extremen Ausprägungen der Variablen. Folglich entsteht der Eindruck von zwei wohlhabenden Clustern (Cluster 1 und Cluster 3) und zwei ärmeren Clustern (Cluster 2 und 4).

Eine weitere Problematik der Clusteranalyse besteht darin, dass einzelne Variablen nicht als spezifische Treiber oder Träger von Entwicklungen identifiziert werden können. Dies erschwert es, die Bedeutung einzelner Faktoren zu gewichten und zu bestimmen, inwiefern sie die beobachteten Phänomene beeinflussen.

Die Analyse stößt ebenfalls an ihre Grenzen, wenn es um die Differenzierung zwischen Angebots- und Bestandsmieten geht. Eine fehlende Unterscheidung kann die Aussagekraft der Ergebnisse beeinträchtigen und die tatsächliche Mietbelastung bestimmter Bevölkerungsgruppen verzerren. Darüber hinaus sind Variablen wie die Attraktivität eines Stadtteils oder der politische Wille zur Auflösung von Segregationsräumen unskalierbar und bleiben in der Analyse unberücksichtigt, obwohl sie einen erheblichen Einfluss auf die Entwicklung haben können.

Die größte Limitation ergibt sich jedoch aus der Tatsache, dass die Analyse zwar stichhaltige Zusammenhänge aufzeigen kann, aber keine Kausalitäten beinhaltet. Es bleibt unklar, ob die Segregation durch eine hohe Ausländerquote verursacht wird oder umgekehrt. Ebenso ist das Einkommensniveau lediglich eine Beobachtung ohne klare Aussage über dessen Rolle als Haupttreiber der Segregation.

Unbeantwortet bleibt zudem die Frage, ob die Segregation eine natürliche Entwicklung darstellt oder das Ergebnis administrativer Entscheidungen ist. Es ist unklar, ob das Ausmaß der Segregation einer integrativen gesellschaftlichen Entwicklung entspricht oder lediglich auf die räumliche Lage beschränkt bleibt. Ebenso besteht Unsicherheit darüber, ob ein Bedarf zur Auflösung der Konzentration sowohl bei einkommensstarken als auch -schwachen Gruppen vorhanden ist oder ob diese Konzentrationen trotz sozialen Aufstiegs bestehen bleiben.

Die Einbeziehung weiterer Variablen und eine umfassendere Literaturrecherche könnten einige der genannten Limitationen adressieren. Eine Analyse über einen längeren Zeitraum bis zur Gegenwart könnte zudem präzisere Aussagen über die Entwicklung und deren Treiber ermöglichen, was zu einem tieferen Verständnis der komplexen Dynamiken führen würde, die die Segregation in urbanen Räumen beeinflussen.

6. Fazit

Wie verteilen sich Bevölkerungsgruppen in Berlin? Können Entwicklungen in der Verteilung festgestellt werden? Die Untersuchung der Bevölkerungsverteilung in Berlin im Zeitraum von 2010 bis 2019 zeigt, dass sich bestimmte Muster und Trends in der räumlichen Verteilung von Bevölkerungsgruppen abzeichnen lassen. Als wesentliche Merkmale konnten Unterschiede der Bevölkerungsstruktur zwischen Ost- und Westberlin sowie zwischen den Randgebieten und zentralen Lagen identifiziert werden. Die Randgebiete sind überwiegend durch bessere wirtschaftliche Kennzahlen (niedrigere Arbeitslosenquote, höheres Einkommen) und eine geringere Arbeitslosenquote charakterisiert, während das Gegenteil auf die zentralen Gebiete zutrifft. Zwischen den beiden gegensätzlichen Zonen konnte eine Art Pufferzone mit durchschnittlichen Werten identifiziert werden, die die beiden Zonen räumlich voneinander trennt. Dennoch existieren mit Marzahn-Hellersdorf im Osten und Spandau im Westen Bezirke, die aus dem Raster fallen und am Stadtrand auf herausfordernde Lebensverhältnisse hindeuten.

Hinsichtlich potenzieller Entwicklungen kam die Forschungsarbeit zu dem Ergebnis, dass die Verteilungen in dem Zeitraum von 2010 bis 2019 weitestgehend stabil geblieben sind und sich keine drastischen Umbrüche abgezeichnet haben. Dennoch konnten sich Normalisierungseffekte feststellen lassen, welche sich besonders in Gebieten mit extremen Konzentrationen (unter- wie auch überdurchschnittlich) und im östlichen Stadtgebiet ereignet haben. Anhand der Vergleichsstudie konnten die Ergebnisse größtenteils bestätigt werden.

Obwohl einige Anzeichen für Veränderungen in der Verteilung von Bevölkerungsgruppen existieren, erweckt Berlin im Gesamten weiterhin den Eindruck einer segregierten Stadt. Dies verdeutlicht die Komplexität der städtischen Bevölkerungsstruktur und legt nahe, dass weitere Forschung erforderlich ist, um die Hintergründe der Segregation besser zu verstehen und gezielte Maßnahmen zur Förderung einer integrativen Stadtentwicklung zu entwickeln.

Können sinnvolle Cluster anhand von (wenigen) sozioökonomischen Variablen gebildet werden? Sind diese Cluster über die Zeit stabil? Die Besonderheit des Algorithmus der Clusteranalyse besteht darin, dass die Bildung der Cluster ohne die Verwendung von geografischen Variablen erfolgte. Hierfür wurden insgesamt sechs Variablen in die Analyse einbezogen, die neben sozioökonomischen Eigenschaften mit dem Anteil an Einfamilienhäusern auch bauliche Aspekte beinhaltete. Die geografische Verteilung der Cluster zeigt die Bildung von größeren räumlich zusammenhängenden Gebieten des gleichen Clusters und bestätigt damit die Forschungsfrage. Ein weiteres Indiz hierfür liefert der hohe Deckungsgrad der Clusterverteilung mit den Wahlergebnissen des darauffolgenden Jahres.

Die Stabilität der Cluster im Zeitverlauf wird durch Clusteranalysen für verschiedene Jahre bestätigt. Es zeigen sich keine signifikanten Veränderungen von 2005 bis 2015, erst im Jahr 2019

werden Abweichungen ersichtlich. Diese Abweichungen ergeben sich vorwiegend durch Wechsel zu einem Cluster mit ähnlichen Werten. Daher bleiben die strukturellen Muster größtenteils erhalten, was darauf hinweist, dass die Clusterbildung über die Zeit hinweg stabil ist.

Sollte es zu Veränderungen gekommen sein, können diese nicht eindeutig auf einzelne Variablen zurückgeführt werden, da die Cluster aus mehreren Variablen zusammengesetzt sind. Folglich ist an dieser Stelle weiterer Forschungsbedarf gegeben, um die maßgeblichen Treiber hinter den Entwicklungen zu identifizieren. Weiterhin eröffnet sich die Frage, wie sich die im Jahr 2019 abgezeichneten Veränderungen in den darauffolgenden Jahren entwickelt haben und welche Prognosen daraus für die Zukunft gezogen werden können.

Wie hat sich das Verhältnis von Einkommen zu Miete in Berlin während des Zeitraums von 2009 bis 2019 entwickelt und auf welche spezifischen Bevölkerungsgruppen hat sich diese Entwicklung positiv respektive negativ ausgewirkt? Basierend auf den durchgeföhrten Regressionsanalysen lässt sich feststellen, dass das Verhältnis von Einkommen zu Miete in Berlin im Zeitraum von 2009 bis 2019 signifikante Veränderungen erfahren hat, die sich unterschiedlich auf verschiedene Bevölkerungsgruppen ausgewirkt haben. Die Analyse der Kaltmiete und Mietbelastung zeigt einen kontinuierlichen Anstieg über die Jahre, wobei die Mietbelastung für Bevölkerungsgruppen in sozial benachteiligten Brennpunktgebieten stärker gestiegen ist als in anderen Gebieten. Dies deutet darauf hin, dass insbesondere ärmere Bevölkerungsschichten stärker von den Mietpreisseigerungen betroffen waren. Die Regression mit den Clustern als erklärende Variable zeigt, dass insbesondere das Zentrum Berlins mit seinem hohen Ausländeranteil eine hohe Steigerung in der Mietbelastung erfahren hat. Die Einbeziehung von Lokationskoeffizienten wiederum ebenso, dass die Mietbelastung in Gebieten mit einem hohen Ausländeranteil stärker gestiegen ist und dass die Entwicklung sehr unterschiedlich verlaufen ist, je nachdem wie hoch die Konzentration der Arbeitslosigkeit in einem Gebiet ist. Diese Erkenntnisse bestätigen die Annahme, dass die Entwicklung des Mietmarktes in Berlin bestimmte Bevölkerungsgruppen überproportional belastet hat, wobei sozial benachteiligte Gruppen und Gebiete mit spezifischen demografischen Merkmalen am stärksten betroffen sind.

Welche Einflussfaktoren sind die wesentlichen Treiber der räumlichen Segregation? Die vorliegende wissenschaftliche Arbeit hat sich intensiv mit den Treibern räumlicher Segregation beschäftigt und dabei besonderes Augenmerk auf die Rolle des Einkommensniveaus sowie der damit verbundenen Kaufkraft gelegt. Die Analyseergebnisse stützen die Hypothese, dass das Einkommensniveau als einer der wichtigsten Faktoren für die Segregation angenommen werden kann. Es zeigt sich eine Korrelation zwischen Einkommensniveau, Kaufkraft und der Ausländerquote, wobei interessanterweise in den untersuchten Gegenden eine leichte Abnahme der Ausländerquote festgestellt wurde. Diese Beobachtung führt zu der Annahme, dass die Ausländerquote möglicherweise nicht zu den primären Treibern der Segregation gehört.

Ein weiterer bedeutsamer Aspekt, der in dieser Arbeit hervorgehoben wird, ist der Einfluss historischer Ereignisse auf die gegenwärtige Struktur der Segregation. Sowohl die quantitative Datenanalyse als auch die qualitative Forschung deuten darauf hin, dass segregative Unterschiede entlang der ehemaligen Berliner Mauer existieren. Dies legt nahe, dass historisch bedingte Einwanderungsmuster einen nachhaltigen Einfluss auf die heutige sozioökonomische Landschaft Berlins haben und somit als signifikanter Faktor in der räumlichen Segregation betrachtet werden können.

Trotz dieser Erkenntnisse müssen jedoch Einschränkungen hinsichtlich weiterer möglicher Treiber der räumlichen Segregation gemacht werden, da diese außerhalb des Untersuchungsbereichs dieser Arbeit liegen. Die Komplexität des Phänomens räumlicher Segregation bedingt, dass zahlreiche Variablen interagieren, um die beobachteten Muster zu erzeugen. Während das Einkommensniveau - partiell auch die Ausländer- und Arbeitslosenquote - und historische Einwanderungsmuster als wichtige Einflussfaktoren identifiziert wurden, bleibt die Rolle anderer potenzieller Treiber unerforscht und öffnet somit ein Feld für zukünftige Forschungen, um ein umfassenderes Verständnis der Ursachen und Dynamiken räumlicher Segregation zu erlangen.

Literaturverzeichnis

- Alberto, Alesina und Glaeser, Edward Ludwig (2004). „Fighting Poverty in the US and Europe: A World of Difference“. In: URL: https://discovered.ed.ac.uk/permalink/44UOE_INST/iatqhp/alma9912283473502466 (besucht am 18.02.2024).
- Ulrich Raiser, Rainer Ohliger und (2005). „Integration und Migration in Berlin“. In: URL: <https://digital.zlb.de/viewer/metadata/15455495/1/> (besucht am 18.02.2024).
- Dohnke, Jan und Seidel-Schulze, Antje (2013). „Soziale Segregation und Polarisierung in deutschen Städten – Relevanz und Möglichkeiten der Messung“. In: *Informationen zur Raumentwicklung* 6, S. 529. URL: https://www.bbsr.bund.de/BBSR/DE/veroeffentlichungen/izr/2013/6/Inhalt/DL_Dohnke_SeidelSchulze.pdf?__blob=publicationFile&v=1 (besucht am 11.01.2024).
- Einem, Eberhard von (2015). *Wohnen Markt in Schieflage - Politik in Not*. URL: <https://books.google.de/books?id=6VeLCwAAQBAJ&lpg=PR5&ots=BNecDBs0hB&dq=Wohnen%20markt%20in%20schieflage&lr=de&pg=PR5#v=onepage&q=Wohnen%20markt%20in%20schieflage&f=false> (besucht am 24.02.2024).
- Krajewski, Christian (2015). *Arm, sexy und immer teurer - Wohnungsmarktentwicklung und Gentrification in Berlin*. URL: <https://link.springer.com/article/10.1007/s00548-015-0381-1> (besucht am 24.02.2024).
- Pauly, Marcel (2016). „Wen wählten Arbeitslose? Wen Gebildete? Wen ledige Frauen?“ In: *Die Welt*. URL: <https://www.welt.de/politik/deutschland/article158246437/Wen-waehlten-Arbeitslose-Wen-Gebildete-Wen-ledige-Frauen.html> (besucht am 30.01.2024).
- Bundesamt für Kartographie und Geodäsie (2017). *Geographische Gitter für Deutschland in Lambert-Projektion (GeoGitter Inspire)*. URL: <https://gdz.bkg.bund.de/index.php/default/inspire/sonstige-inspire-themen/geographische-gitter-für-deutschland-in-lambert-projektion-geogitter-inspire.html> (besucht am 22.11.2023).
- Geoportal Berlin (2017). *Bezirksgrenzen Berlin*. URL: <https://daten.odis-berlin.de/de/dataset/bezirksgrenzen/> (besucht am 22.11.2023).
- Lebuhn, Henrik u. a. (Sep. 2017). *Wohnverhältnisse in Deutschland – eine Analyse der sozialen Lage in 77 Großstädten*. URL: https://www.boeckler.de/pdf_fof/99313.pdf (besucht am 11.01.2024).
- Breidenbach, Philipp und Eilers, Lea (2018). „RWI-GEO-GRID: Socio-economic data on grid level“. In: URL: <https://www.degruyter.com/document/doi/10.1515/jbnst-2017-0171/html> (besucht am 18.02.2024).
- Geoportal Berlin (2018). *Verlauf der Berliner Mauer, 1989*. URL: <https://hub.arcgis.com/maps/ef3c99e6dfcf4f84a90ae91820b08d7a/about> (besucht am 22.02.2024).

- Statistisches Bundesamt (2019). *Pressemitteilung Nr. N 001 vom 1. Oktober 2019*. URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/2019/10/PD19_N001_129.html (besucht am 01.02.2024).
- Blokland, Talja und Vief, Robert (2021). „Making Sense of Segregation in a Well-Connected City: The Case of Berlin; in Urban Socio-Economic Segregation and Income Inequality“. In: URL: https://link.springer.com/chapter/10.1007/978-3-030-64569-4_13 (besucht am 18.02.2024).
- Budliger, Hendrik (2021). *Demografischer Wandel und Wirtschaft*. URL: <https://link.springer.com/content/pdf/10.1007/978-3-658-31521-4.pdf> (besucht am 25.02.2024).
- Marcinczak, Szymon und Bernt, Matthias (2021). „Immigration, segregation and neighborhood change in Berlin“. In: URL: https://www.sciencedirect.com/science/article/abs/pii/S0264275121003164?fr=RR-2&ref=pdf_download&rr=85892238ba779bd7 (besucht am 24.02.2024).
- Statistik Berlin Brandenburg (2022). *Zu- und Fortzüge*. URL: <https://www.statistik-berlin-brandenburg.de/bevoelkerung/demografie/zu-und-fortzuege> (besucht am 29.01.2024).
- Statistisches Bundesamt (2022). *Haushalte nach HaushaltsgröÙe und Haushaltsmitgliedern*. URL: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Haushalte-Familien/Tabellen/1-2-privathaushalte-bundeslaender.html> (besucht am 30.01.2024).
- Arbeitskreis der Oberen Gutachterausschüsse, Zentralen Geschäftsstellen und Gutachterausschüsse in der Bundesrepublik Deutschland (AK OGA) (2023). *Immobilienmarktbericht Deutschland 2023*. URL: <https://redaktion-akoga.niedersachsen.de/startseite/download/berichte/download-berichte-201849.html> (besucht am 24.01.2024).
- empirica (2023). *Entwicklung der Angebotsmieten für Wohnungen in Berlin von 2012 bis zum 2. Quartal 2023 (in Euro pro Quadratmeter) [Graph]*. In Statista. URL: <https://de.statista.com/statistik/daten/studie/535119/umfrage/mietpreise-auf-dem-wohnungsmarkt-in-berlin/> (besucht am 11.01.2024).
- Jonas, Lukas, Martin, Carolin und Theobald, Thomas (2023). „Mehr öffentlicher Wohnungsbau zum Erhalt der Kapazitäten? Eine Abschätzung möglicher Rückgänge der Bauaktivität durch gestiegene Zinsen für die Jahre 2023-24“. In: URL: <https://www.econstor.eu/handle/10419/274582> (besucht am 29.01.2024).
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Senatsverwaltung für Stadtentwicklung, Bauen und Wohnen (2023). *Immobilienmarktbericht Berlin 2022/2023*. URL: <https://www.berlin.de/gutachterausschuss/marktinformationen/marktanalyse/artikel.175633.php> (besucht am 15.01.2024).

- Statista (2023). *Einwohnerzahl von Berlin in ausgewählten Jahren von 1600 bis 2022 [Graph]*. Autoren: Zentral- und Landesbibliothek Berlin (bis 1945), Statistisches Amt der DDR (1960 bis 1989), Statistisches Bundesamt (ab 1958). URL: <https://de.statista.com/statistik/daten/studie/1079347/umfrage/bevoelkerung-in-berlin/> (besucht am 11.01.2024).
- Statistisches Bundesamt (2023). *Bevölkerung nach Gebietsstand (ab 1990)*. URL: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Tabellen/liste-gebietsstand.html#249750> (besucht am 11.01.2024).
- Waiphot, Kulachai, Lerdtomornsakul, Unisa und Homyamyen, Patipol (2023). „Factors Influencing Voting Decision: A Comprehensive Literature Review“. In: URL: https://www.mdpi.com/2076-0760/12/9/469/review_report (besucht am 18.02.2024).
- Statistisches Bundesamt (2024a). *12411-0014: Bevölkerung: Bundesländer, Stichtag, Nationalität, Geschlecht, Altersjahre*. URL: <https://www-genesis.destatis.de/> (besucht am 30.01.2024).
- (2024b). *61111-0003: Verbraucherpreisindex: Deutschland, Jahre, Klassifikation der Verwendungszwecke des Individualkonsums*. URL: <https://www-genesis.destatis.de/> (besucht am 01.02.2024).
- (2024c). *Baugenehmigungen für Wohnungen im November 2023*. URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/2024/01/PD24_028_3111.html (besucht am 29.01.2024).
- (2024d). *Pressemitteilung Nr. 020 vom 16. Januar 2024*. URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/2024/01/PD24_020_611.html (besucht am 24.02.2024).



A. Weitere Abbildungen

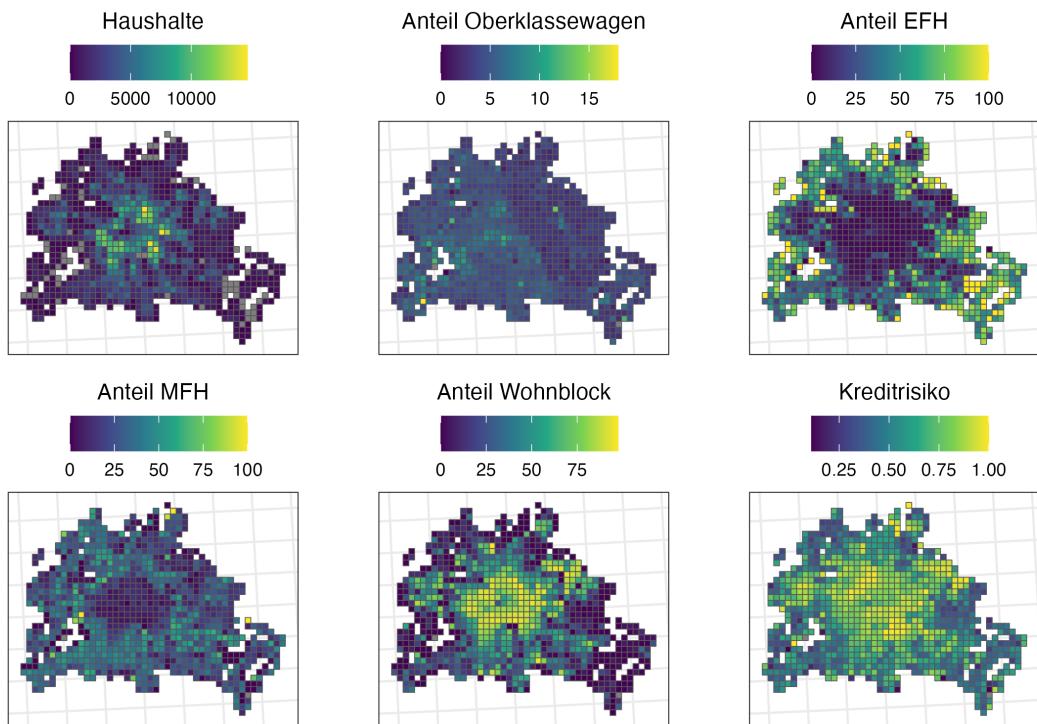


Abbildung 22: Restliche Variablen Sozialdatensatz auf Karte 2015

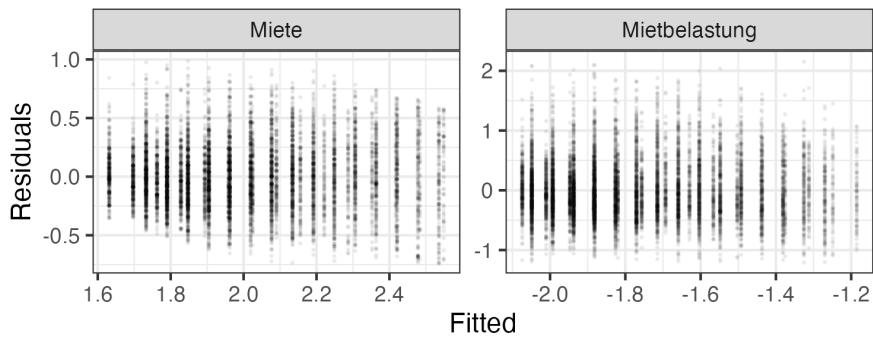


Abbildung 23: Residuen Regression Brennpunkte

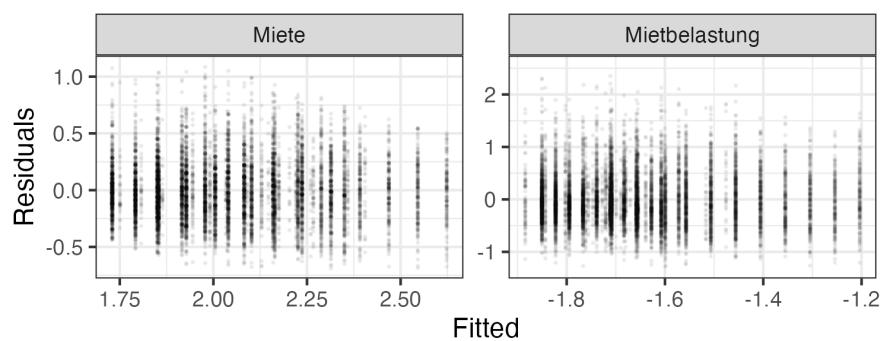


Abbildung 24: Residuen Regression Cluster

B. Beitragsübersicht der Autoren

In der folgenden Tabelle ist festgehalten, wer welche Kapitel verfasst hat.

Kapitel	Name
1.	Jamal Mengste
1.1	Pascal Gorzelany
1.2	Pascal Gorzelany
1.3	Jamal Mengste
2.	Jamal Mengste
2.1	David Kazakovic
2.2	David Kazakovic
2.3	Jamal Mengste
2.4	Jamal Mengste & David Kazakovic
2.5	Jamal Mengste & David Kazakovic
3.	David Kazakovic
3.1	David Kazakovic
3.2	David Kazakovic
3.3	David Kazakovic
4.	Pascal Gorzelany
4.1	Pascal Gorzelany
4.2	Sebastian Geis
5.	David Kazakovic & Jamal Mengste
5.1	David Kazakovic & Jamal Mengste
5.2	Jamal Mengste
6.	Pascal Gorzelany, Sebastian Geis & Jamal Mengste

Die Abbildungen in Kapitel 4.1.1 wurden von Pascal Gorzelany erstellt. Alle restlichen Abbildungen und Tabellen wurden von Sebastian Geis erstellt.