

Regionalisierung der Ergebnisse aus dem KfW-Kommunalpanel: Methodische Zugänge für ein Herunterbrechen auf Ebene der Länder oder Regionen

Christian Raffer

Regionalisierung der Ergebnisse aus dem KfW-Kommunalpanel: Methodische Zugänge für ein Herunterbrechen auf Ebene der Länder oder Regionen

Methodenpapier im Auftrag der KfW Bankengruppe

Autoren:

Christian Raffer, M.Sc.
unter Mitarbeit von Dr. Henrik Scheller

Deutsches Institut für Urbanistik
Forschungsbereich Infrastruktur, Wirtschaft und Finanzen
Zimmerstraße 13-15
10969 Berlin
Telefon 030 39001-223
Telefax 030 39001-116
E-Mail: kommunalpanel@difu.de
www.difu.de

Endredaktion:

Dr. Stephan Brand
Dr. Johannes Steinbrecher

KfW Bankengruppe (Hrsg.)
Abteilung Volkswirtschaft
Palmengartenstraße 5-9
60325 Frankfurt am Main
Telefon 069 7431-0
Telefax 069 7431-2944
E-Mail: research@kfw.de
www.kfw.de

Berlin und Frankfurt am Main, März 2023

Inhalt

Abstract.....	4
1. Einleitung und Problemstellung	6
2. Statistische Repräsentativität	8
2.1. Repräsentativität herstellen	8
2.2. Repräsentativität bewerten	9
3. Gewichtung und Hochrechnung zur Herstellung repräsentativer Ergebnisse	10
3.1. Gewichtung	10
3.2. Hochrechnung	11
3.3. Ergänzungsstichproben	12
3.4. Repräsentativität	12
4. Möglichkeiten zur Ermittlung repräsentativer Ergebnisse für Regionen und Länder.....	15
4.1. Regionalisierung auf Ebene der Regionen	15
4.2. Regionalisierung auf Ebene der Länder	20
5. Weitere Ansätze zur Regionalisierung	21
5.1. Erhöhung der Fallzahl durch Daten-Pooling.....	21
5.2. Alternative Methoden zum Imputieren fehlender Daten	23
6. Fazit und Ausblick.....	25
Literaturverzeichnis	27

Abstract

Für die wirtschaftspolitische Debatte in Deutschland werden wissenschaftlich belastbare Datenquellen und Analyseergebnisse benötigt. Für den öffentlichen Sektor gilt dies insbesondere für die Kommunen, da die zentrale Bereitstellung amtlicher Statistiken häufig nur auf aggregierter Ebene erfolgt und zudem Lücken im Umfang der verfügbaren Informationen bestehen, die durch ergänzende Erhebungsformen wie Befragungen geschlossen werden müssen (Brand und Steinbrecher 2019). Weil kommunalpolitische Entscheidungen auf lokaler Ebene getroffen werden oder Teil der Landespolitik sind, ist das Interesse an kleinräumigen, regionalisierten Ergebnissen besonders hoch.

Das KfW-Kommunalpanel ist eine etablierte Quelle für die wirtschaftspolitische Debatte kommunaler Fragestellungen. Befragungen wie das KfW-Kommunalpanel liefern wertvolle Einblicke in die zeitliche Entwicklung der beobachteten Einheiten. Um die Repräsentativität der gewonnenen Ergebnisse zu gewährleisten, müssen jedoch die immer wiederkehrenden Herausforderungen in der Zusammensetzung der Nettostichprobe mit geeigneten Methoden adäquat adressiert werden, beispielsweise durch Gewichtung und Ziehung von Ergänzungsstichproben.

Die seit Jahren angewandte Methodik des KfW-Kommunalpanels zur Berechnung von Kenngrößen wie Investitionstätigkeit und Finanzierungssituation der Kommunen zielt darauf ab, möglichen Verzerrungen zu begegnen und repräsentative Ergebnisse für Gesamtdeutschland zu ermitteln (Krone und Scheller 2020). Eine Übertragung der Ergebnisse auf kleinere geografische Einheiten, die sogenannte Regionalisierung, war bei der Konzeption des KfW-Kommunalpanels nicht vorgesehen und ist deshalb bisher nur partiell möglich. Aufgrund der großen Relevanz regionalisierter Ergebnisse für die kommunale Ebene sollen deshalb die Möglichkeiten einer Regionalisierung des KfW-Kommunalpanels im Folgenden grundlegend diskutiert werden.

Für die Hochrechnung für Gesamtdeutschland werden die Investitionsvariablen u.a. nach Regionen betrachtet, die in der Regel aus mehreren Bundesländern bestehen. Dieser Schritt setzt bereits voraus, dass die für eine Region gezogenen Kommunen repräsentativ für diese sind. Für die Ebene der Regionen kann im vorliegenden Methodenpapier aufgezeigt werden, dass die Teilstichproben hinreichend groß sind, um repräsentative Ergebnisse zu erhalten. Basierend darauf wird eine Regionalisierung auf Ebene größerer Regionen als möglich beurteilt.

Eine Regionalisierung auf Länderebene ist hingegen nur bedingt möglich, da dort – der Konzeption des KfW-Kommunalpanels als deutschlandweit repräsentative Befragung geschuldet – die Teilstichproben teilweise zu klein sind, um statistisch gesicherte Aussagen für alle Bundesländer zu treffen.

Eine Vielzahl von Merkmalen, die im Rahmen des KfW-Kommunalpanels abgefragt werden, unterscheiden sich von den kardinalen Investitionsvariablen dahingehend, dass sie auf mehrere Ausprägungen skaliert sind. Der Rückgriff auf die Stichprobentheorie zeigt, dass repräsentative Aussagen für Gesamtdeutschland und auch die gebildeten Regionen möglich sind. Das Vorgehen wird durch Gewichtung der Antworten gestützt. Auch hier ist das Ziel der Methodik, eine bundesweit repräsentative Kennzahl zu erhalten. Eine Regionalisierung auf Ebene der Bundesländer ist auch hier aufgrund zu geringer Antwortzahlen nicht für jedes Land ratsam.

Um eine Regionalisierung trotz kleiner Fallzahlen zu erreichen, gibt es verschiedene statistische Methoden, die genutzt werden können. Vor allem Small-Area-Estimation und Nearest-Neighbour-Methoden sind dabei vorstellbar. Im Detail wird auch die Nutzung gepoolter Daten aus mehreren Befragungsjahren analysiert.

1. Einleitung und Problemstellung

Das vom Deutschen Institut für Urbanistik jährlich im Auftrag der KfW durchgeführte KfW-Kommunalpanel ist derzeit eine der größten Befragungen kommunaler Kämmerereien und Finanzdezernate in Deutschland. Seit 2009 werden dabei regelmäßig Kennzahlen zur Finanzlage, der Investitionstätigkeit und den Finanzierungsbedingungen der Kommunen abgefragt (Scheller und Schneider 2016). Die für deutsche Kommunen mit mehr als 2.000 Einwohnern bundesweit repräsentativen Ergebnisse sind eine wichtige Grundlage für die politische Debatte über die Finanzsituation der kommunalen Landschaft. Mit Blick auf finanzielle Problemlagen oder den Zustand der kommunalen Infrastruktur werden die bundesweiten Zahlen auch medial regelmäßig aufgegriffen.

Lösungen für kommunale Probleme setzen aber häufig auf der Landesebene an. Insofern wird regelmäßig der Wunsch an das KfW-Kommunalpanel herangetragen, stärker regionalisierte Ergebnisse zu veröffentlichen. Vereinzelt wird die Methodik darum auch unabhängig vom Kommunalpanel auf Bundeslandebene herangezogen, um eigene Erhebungen durchzuführen.¹ Bundeslandscharfe Ergebnisse über ganz Deutschland wären auch aus Sicht des KfW-Kommunalpanels zu begrüßen, da sich daraus besser begründete Handlungsempfehlungen ableiten ließen. Dem sind allerdings methodische Grenzen gesetzt, weil die Teilnahme der Kommunen freiwillig ist und die Rücklaufquote regionale und zeitliche Unterschiede aufweist. Während sich beispielsweise im Jahr 2015 lediglich 425 Kommunen rückmeldeten (Scheller und Schneider 2016), taten dies 2018 insgesamt 770 (Krone und Scheller 2020). Als Konsequenz schwankt auch die Zusammensetzung der Nettostichprobe.

Es stellt sich damit die grundlegende Frage, ob eine stärkere Regionalisierung methodisch vertretbar ist und nicht zu unpräzisen Schätzungen und methodisch bedingt starken jährlichen Schwankungen regionalisierter Ergebnisse führen würde. Das vorliegende Methodenpapier soll zur Klärung dieser Fragestellung beitragen. Dabei wird wie folgt vorgegangen:

- In einem ersten Schritt wird eine kurze, generelle Übersicht über Befragungsdaten und ihre Repräsentativität gegeben.

¹ Der Landesrechnungshof Niedersachsen hat beispielsweise im Jahr 2021 eine landesweite Erhebung von Investitionsrückständen bei niedersächsischen Kommunen durchgeführt (LRH Niedersachsen 2021). Ebenso erhebt das Kompetenzzentrum für kommunale Infrastruktur Sachsen nach vergleichbarer Methodik Zahlen für die sächsischen Kommunen (Sydow et al. 2020).

- Danach wird die Methodik des KfW-Kommunalpanels zur Gewichtung und Hochrechnung von regionalen Teilergebnissen zu einer gesamtdeutschen Aggregation vorgestellt und kritisch diskutiert.
- In einem dritten Schritt wird die Frage erörtert, ob eine Regionalisierung auf Ebene einzelner Regionen und/oder Länder möglich ist.
- Abschließend werden ein kurzer Überblick über weitere Ansätze zur Regionalisierung der Ergebnisse gegeben, insbesondere das Stichprobenpooling, Nearest-Neighbour-Methoden sowie Small-Area-Estimation, und deren Eignung für die Daten des KfW-Kommunalpanels kurz diskutiert.

2. Statistische Repräsentativität

Befragungsdaten sind dann repräsentativ, wenn man mit ihnen zuverlässig Eigenschaften der Grundgesamtheit schätzen kann (Rendtel 1993, S. 2). Dazu ist in der Regel eine ausreichend große Stichprobe notwendig. Damit auch für kleinere Teilgruppen hinreichend große Fallzahlen vorliegen, werden oft geschichtete Stichproben gezogen. Da Individuen aus den kleinen Teilgruppen dann überproportional vertreten sind, ist die gesamte Stichprobe nicht ohne weiteres repräsentativ für die Grundgesamtheit. Durch eine Gewichtung mit sogenannten Designgewichten können allerdings auch bei einer geschichteten Stichprobe repräsentative Ergebnisse erzeugt werden (Horvitz und Tompson 1952).

Probleme entstehen immer dann, wenn nicht alle gezogenen Teilnehmenden antworten. Das erschwert nicht nur die Repräsentativität aufgrund gesunkener Fallzahlen. Darüber hinaus ist es zudem problematisch, wenn dabei das Nicht-Antworten selektiv ist und ein sogenannter Bias auftritt, d. h., wenn sich die nicht-antwortenden Teilnehmenden systematisch von denen unterscheiden, die antworten (Rendtel 1993, S. 1).² Dabei kann es sein, dass manche der gezogenen Teilnehmenden von Anfang an nicht antworten oder andere erst nach einigen Befragungswellen nicht mehr (Panelmortalität). Ist das Nicht-Antworten selektiv, dann können die relevanten Merkmale der Teilnehmenden mit dem Ausfall korrelieren. Die statistischen (Panel)Schätzer sind damit verzerrt und inkonsistent, was man Attrition-Bias nennt (Fitzgerald et al. 1996). Grundsätzlich kann auch das Problem eines sogenannten Non-Coverage-Bias auftreten. Dies ist dann der Fall, wenn die Gesamtheit, aus der die Stichprobe gezogen wird, nicht der Grundgesamtheit entspricht. Dies kann z. B. durch ein ungeeignetes Umfragedesign entstehen (Vollnhals 2017, S. 21f.).

2.1. Repräsentativität herstellen

Damit innerhalb einer Befragungswelle die Stichprobe repräsentativ für die Grundgesamtheit ist, benötigt man Querschnittsgewichte (Heady und Holst 2008, S. 129). Dabei kann man entweder die erwähnten Designgewichte mit sogenannten Non-response-Gewichten kombinieren oder man gewichtet direkt die Nettostrichprobe im Vergleich zur Grundgesamtheit.

² Ein Beispiel dafür ist der sogenannte Mittelschichts-Bias: In sozioökonomischen Panelbefragungen haben Angehörige der Mittelschicht meist eine höhere Antwortbereitschaft und sind daher oft übervertreten, auch wenn sie keine überproportionale Auswahlwahrscheinlichkeit im Design der Befragung haben.

Das geläufigste Gewichtungungsverfahren ist dabei die Gewichtung durch den klassischen Populationsschätzer. Die Gewichtung der Stichprobenergebnisse erfolgt dabei durch den Kehrwert der Auswahlwahrscheinlichkeit einer Einheit (Rendtel 1993, S. 3). Geringen Fallzahlen kann man mit Ergänzungsstichproben begegnen.

2.2. Repräsentativität bewerten

Im Prinzip kann man die Repräsentativität gar nicht bewerten. Da der wahre Wert der Grundgesamtheit i. d. R. nicht bekannt ist, kann man auch nicht sagen, ob der geschätzte Wert diesen gut vorhersagt (von der Lippe und Kladroba 2002). Schätzer konvergieren üblicherweise allerdings in einer ausreichend großen Stichprobe zum wahren Populationswert. D. h. zu kleine Stichproben eignen sich demnach nicht, um mit hinreichender Sicherheit repräsentative Ergebnisse zu schätzen, und bei einer Einschätzung der Repräsentativität sollte man die Fallzahlen miteinbeziehen (Prein et al. 1994). Dabei lässt sich nicht ein allgemeiner Schwellenwert festlegen, ab dem eine Stichprobe als ausreichend groß gilt. Dieser hängt vielmehr von einer Vielzahl von Faktoren ab und ist von Befragung zu Befragung unterschiedlich. Dazu gehören u. a. die Größe der Grundgesamtheit, die Varianz der Variable von Interesse, das Konfidenzlevel sowie der Umfang des Schätzfehlers, den man in Kauf nimmt (Bartlett et al. 2001). Die Notwendigkeit der Varianz zur Berechnung von Mindeststichproben zeigt bereits, dass die Ex-Ante-Bewertung bei einmalig durchgeführten Befragungen schwer zu realisieren ist. Das KfW-Kommunalpanel hat hier den Vorteil, dass bei der Berechnung auf Vorjahreswerte zurückgegriffen werden kann.

Wichtig ist außerdem, dass verwendete Schätzer nicht systematisch verzerrt sind und nicht die bereits erläuterten Verzerrungen durch Non-Coverage vorliegen. Probleme der Non-Coverage kann man allerdings durch das Untersuchen des Umfragedesigns ermitteln. Als Indiz kann man die Verteilung von bekannten Merkmalen (z. B. aus externen Quellen wie amtlichen Statistiken) der Nicht-Antwortenden untersuchen, von denen man erwarten kann, dass sie mit den unbekanntem Antworten korrelieren. Unterscheiden sich diese Merkmale signifikant von denen der Grundgesamtheit, so könnte eine Verzerrung vorliegen (Koch und Blohm 2015). Des Weiteren kann man die Teilnahmerate untersuchen, d. h. welcher Anteil der Befragten geantwortet hat, denn eine hohe Teilnahmerate senkt die Wahrscheinlichkeit für einen Bias (Vollnhals 2017).

3. Gewichtung und Hochrechnung zur Herstellung repräsentativer Ergebnisse

Die Grundgesamtheit des KfW-Kommunalpanels bilden alle deutschen Kommunen mit mehr als 2.000 Einwohnern sowie alle Landkreise.³ Die Bruttostichprobe⁴ besteht aus einer Vollerhebung der Kommunen und Kreise mit über 20.000 Einwohnern sowie einer geschichteten Zufallsstichprobe der Kommunen mit 2.000 bis 20.000 Einwohnern. Diese disproportionale Schichtung erfordert per se eine Gewichtung der Ergebnisse, um repräsentative Aussagen über die Grundgesamtheit der Kommunen ab 2.000 Einwohnern treffen zu können (Krone und Scheller 2020, S. 15). Mithilfe der Gewichtung kann zudem dafür kontrolliert und korrigiert werden, dass die Rücklaufquote nicht in jeder Sub-Population (siehe unten) gleich hoch ist.

3.1. Gewichtung

Schon allein das Stichprobendesign erfordert wegen der in Abschnitt 1 von Kapitel 2 aufgezeigten Gründe eine Gewichtung, damit repräsentative Aussagen getroffen werden können. Die Querschnittsgewichtung erfolgt zurzeit anhand von 3 Kategorien:

- *Einwohnergrößenklasse* (2.000 - 4.999 Einwohner, 5.000 - 19.999 Einwohner, 20.000 - 49.999 Einwohner, Städte ab 50.000 Einwohner, Landkreise)
- *Region* (Süden (Bayern, Baden-Württemberg), Osten (Brandenburg, Sachsen, Sachsen-Anhalt, Thüringen, Mecklenburg-Vorpommern), Norden (Niedersachsen, Schleswig-Holstein), Westen (Nordrhein-Westfalen), Süd-West (Hessen, Rheinland-Pfalz, Saarland))
- *Gruppenspezifische Steuerkraft* (Steuereinnahmen je Einwohner).

Dabei werden in jeder Gruppe aus Einwohnergrößenklasse und Region 3 Terzile gebildet.

Der Gewichtungsfaktor w jeder der Untergruppen u entspricht dabei dem klassischen Populationsschätzer nach Rendtel (1993) und vereint Designgewichte und Non-Response-Gewicht:

³ Vgl. Methodenpapier, relevante Grundgesamtheit rd. 5.500 Kommunen (Krone und Scheller 2020).

⁴ Man unterscheidet zwischen Grundgesamtheit (alle in Frage kommenden Untersuchungseinheiten, z. B. alle Kommunen mit mehr als 2.000 Einwohnern), Bruttostichprobe (alle zur Befragung aufgeförderten Untersuchungseinheiten, z. B. alle Kommunen, die im Rahmen des Kommunalpanels eine postalische Aufforderung zur Teilnahme erhalten) sowie Nettostichprobe (alle an der Befragung teilnehmenden Untersuchungseinheiten, z. B. alle Kommunen, die in einem bestimmten Jahr am KfW-Kommunalpanel teilnehmen). Während die Nettostichprobe eine Teilmenge der Bruttostichprobe ist, ist die Bruttostichprobe eine Teilmenge der Grundgesamtheit.

$$w_u = \frac{\frac{N_{GG_u}}{N_{GG_{gesamt}}}}{\frac{N_{TN_u}}{N_{TN_{gesamt}}}} \quad (1)$$

$$w_u = \frac{\text{Anteil einer Untergruppe der Grundgesamtheit an der gesamten Grundgesamtheit}}{\text{Anteil einer Untergruppe der Teilnehmenden an der Bruttostichprobe}}$$

Insgesamt ist die Methodik der Querschnittsgewichtung bei ausreichenden Fallzahlen geeignet, um repräsentative Ergebnisse zu erhalten.

3.2. Hochrechnung

Die Hochrechnung im KfW-Kommunalpanel berechnet Summen für die Grundgesamtheit und wird ausschließlich bei Fragen zu Investitionen genutzt. Dort ist die Item-Non-Response, also die Nichtbeantwortung einzelner Fragen, recht hoch, da aufgrund von Auslagerungen in städtische Unternehmen, Eigenbetriebe, Anstalten des öffentlichen Rechts etc. viele Kommunen nicht für Investitionen in allen Infrastrukturbereichen innerhalb ihrer Kernverwaltung selbst zuständig sind. Die Verwendung des über die generelle Befragungsteilnahme bestimmten Gewichtungsfaktors würde hier zu Verzerrungen führen (siehe auch Krone und Scheller 2020, S. 12-14).

Prinzipiell erfolgt die Einteilung in Untergruppen nach Einwohnerzahl und Region wie bei der Gewichtung, lediglich die Steuerkraft wird außenvor gelassen. Allerdings wird nicht über die Anzahl der Kommunen, sondern die Anzahl der Einwohner gewichtet. Innerhalb jeder der Gruppen g wird das entsprechende Ergebnis, z. B. der Investitionsrückstand IR_g , ermittelt, indem der durchschnittliche Pro-Kopf-Rückstand IR_{PK}_g aller Kommunen, die für diese Investitionen zuständig sind, mit der Gesamteinwohnerzahl der Gruppe Bev_g multipliziert wird. Da nicht alle Kommunen einer Gruppe für die Investitionen verantwortlich sind, wird der errechnete Wert noch mit dem Anteil der verantwortlichen Kommunen multipliziert.

$$IR_g = IR_{pk_g} * Bev_g * verantwortw_g \quad (2)$$

Für die Berechnung des Gesamtinvestitionsrückstandes wird dann das Ergebnis aller Gruppen summiert. Die durchgeführte Hochrechnung entspricht dem Kenntnisstand der Wissenschaft und ist geeignet, um bei ausreichenden Fallzahlen für Deutschland repräsentative Ergebnisse zu erhalten.

3.3. Ergänzungsstichproben

Um der Panelmortalität entgegenzuwirken und die Nettostichprobe möglichst konstant zu halten, werden im Rahmen des KfW-Kommunalpanels bei Bedarf Ergänzungsstichproben in den Kommunen mit weniger als 20.000 Einwohnern gezogen. Da in Städten und Gemeinden mit über 20.000 Einwohnern und Landkreisen bereits eine Vollerhebung durchgeführt wird, erübrigen sich dort Ergänzungsstichproben. Seit 2018 werden diese disproportional geschichtet, um vor allem in den Ländern vermehrt Kommunen anschreiben zu können, in denen der Rücklauf in den Vorjahren vergleichsweise gering war. Die Ziehung von Ergänzungsstichproben und ihre disproportionale Schichtung wirkt einem sinkenden Stichprobenumfang entgegen und kann so die Repräsentativität erhöhen.

3.4. Repräsentativität

In den meisten Gruppen ist der Rücklauf (Nettostichprobe) hoch und die gezogene Bruttostichprobe bildet einen großen Teil der Grundgesamtheit ab (siehe Tabelle 1). In der kleinsten Gemeindegröße ist die Nettostichprobe zwar anteilmäßig klein, aber aufgrund der Gruppengröße der Grundgesamtheit absolut sehr hoch und erscheint generell ausreichend, um repräsentative Ergebnisse zu erhalten. Um eine möglichst konkrete Angabe zur Untergrenze für die notwendige Nettostichprobengröße zu machen, um Ergebnisse für Gesamtdeutschland zu berechnen, bedarf es allerdings einer detaillierten Analyse. Die entsprechende Berechnung hängt dabei kritisch von der Art des untersuchten Merkmals ab.

Tabelle 1: Grundgesamtheit der Nettostichprobe 2018

Gemeinden mit ... bis unter ... Einwohnern bzw. Kreise	Grundgesamtheit	Rücklauf absolut (Nettostichprobe)	Rücklauf in Prozent der Grundgesamtheit
2.000 – 4.999	2.210	119	5,4%
5.000 – 19.999	2.230	271	12,2%
20.000 – 49.999	509	163	32,0%
50.000 und mehr	191	80	41,9%
Kreise	294	139	47,3%
Summe	5.434	772	14,2%

Quelle: Krone und Scheller 2020, Aufgabenbeschreibung, Datenauszug aus dem KfW-Kommunalpanel

Im KfW-Kommunalpanel werden im Wesentlichen stetige Variablen abgefragt, die mit Hilfe einer im Rahmen des Fragebogendesigns festgelegten Metrik skaliert werden, z. B. Einschätzungen der finanziellen Gesamtsituation, skaliert in fünf Schritten von „sehr gut“ bis „mangelhaft“ (siehe auch Bartlett et al. 2001, S. 44).

Daneben spielen stetige Variablen eine Rolle, bei denen die Antwortenden konkrete Zahlenwerte angeben müssen. Dies ist vor allem bei den Investitionsvariablen der Fall.

Für stetige Variablen beschreibt Cochran (1977, S. 105) in seinem Standardwerk „Sampling Techniques“ das Vorgehen zur Berechnung von Mindeststichproben im geschichteten Sample:

$$n = \frac{\sum \frac{W_h^2 s_h^2}{w_h}}{V + \frac{1}{N} \sum W_h s_h^2} \quad (3)$$

mit $w_h = \frac{n_h}{n}$ als Anteil der antwortenden Kommunen in einer Schicht h an allen antwortenden Kommunen, $W_h = \frac{N_h}{N}$ als Anteil aller Kommunen in einer Schicht an der Grundgesamtheit, der empirischen Varianz in der Schicht der Grundgesamtheit s_h^2 sowie $V = \left(\frac{d}{t}\right)^2$ mit dem erlaubten absoluten Fehler d und dem auf einem ex-ante festgelegten α -Niveau basierenden t -Wert aus der Studentischen t -Verteilung.⁵

Die Annahme, die hinter der Berechnung steht, ist die der Zufallsstichprobe in jeder Schicht. Diese Annahme ist für das KfW-Kommunalpanel erfüllt. Darüber hinaus stellt es generell eine Schwierigkeit dar, die zur Bewertung notwendige empirische Varianz pro Schicht der Grundgesamtheit, die die Antworten einer erst noch abzufragenden Variable aufweisen wird, bereits vor der Durchführung der Befragung zu bewerten. Im Idealfall lässt sie sich zum Beispiel über vollständig vorliegende Daten der amtlichen Statistik erschließen bzw. schätzen.

Panel-Befragungen haben hier grundsätzlich den Vorteil, auf Vorjahreswerte der zu bewertenden Variablen zurückgreifen zu können und die bei der letzten Welle gemessene Varianz als Schätzer zugrunde zu legen. Hier stellt sich jedoch die Frage, inwieweit diese empirische, also die aus den Antworten des Vorjahres berechnete Varianz der wahren Varianz in der Sub-Population der Grundgesamtheit entspricht. Die Validität dieser Annahme wird in Sub-Populationen der Bruttostichprobe umso schwieriger, je geringer die Zahl der Antworten ist, auf der die empirische Varianz basiert. Aus diesem Grund können die beispielhaften Berechnungen einer Mindestnettostichprobe im vorliegenden Methodenpapier nur ein vorsichtiges Indiz über die Zahl der mindestens nötigen Antworten abliefern, die zur Herstellung von Repräsentativität notwendig sind.

⁵ Für eine intuitive Erläuterung der Bedeutung von d und t als Faktoren für die Präzision siehe Bartlett et al. (2001).

Die Berechnung der Mindestnettostichprobe – also der mindestens notwendigen Zahl an Antworten – erfolgt entsprechend des angestrebten Regionalisierungsgrades jeweils für vorab spezifizierte geografische Einheiten, also pro Bundesland, pro Region oder für Gesamtdeutschland. Die Bruttostichprobe für ein Bundesland lässt sich – wie oben angemerkt – in zwei Schichten unterteilen: Kommunen in der Größe zwischen 2.000 und 19.999 Einwohnern sowie Kommunen mit mehr als 20.000 Einwohnern (inkl. Kreise).

Zur Berechnung der notwendigen Mindestnettostichprobe werden entsprechend zum Vorgehen nach Cochran (1977) die gewichtete empirische Varianz eines Befragungsjahres sowie ex-ante festgelegte Werte für den erlaubten absoluten Fehler und das zugrunde gelegte α -Niveau herangezogen. Aufgrund der Stichprobenstruktur kann eine sogenannte „finite sample correction“ notwendig werden, die bereits Bestandteil der Formel 3 ist. Würden die Berechnung der Stichprobengröße und die Korrektur für die Samplegröße separat durchgeführt, wobei zunächst die Stichprobengröße n_0 berechnet und im Anschluss die Korrektur via Formel (5) durchgeführt würde, so sähe das Vorgehen folgendermaßen aus:

$$n_0 = \frac{1}{V} \sum \frac{W_h^2 s_h^2}{w_h} \quad (4)$$

$$n = \frac{n_0}{\left(1 + \frac{1}{N} \sum W_h s_h^2\right)} \quad (5)$$

Die notwendigen Mindestnettostichproben für zwei beispielhafte Merkmale aus der Befragung werden in den Tabellen 3 und 4 dargestellt. Dabei handelt es sich einmal um das stetige Merkmal „Geplante Investitionen für das aktuelle Haushaltsjahr“ und einmal um das Merkmal „Bewertung der aktuellen Finanzsituation“ – eine Einschätzungsfrage mit einer Likert-Skala, die für die Zwecke dieses Methodenpapiers entsprechend Bartlett et al. (2001, S. 44) ebenfalls als stetig angenommen wird. Die Parameter der Berechnung wurden so festgelegt, dass als Mindestkriterium repräsentative Aussagen für alle Regionen möglich sind.

4. Möglichkeiten zur Ermittlung repräsentativer Ergebnisse für Regionen und Länder

Die zentrale Frage dieses Methodenpapiers ist es, ob sich die Daten des KfW-Kommunalpanels für eine Regionalisierung eignen, d. h. ob sich mit ihnen auch Ergebnisse für geografisch kleinere Einheiten als Gesamtdeutschland berechnen lassen. Dabei liegt das Augenmerk vor allem auf den Regionen (Süden, Osten, Norden, NRW (= Westen), und Süd-West) sowie den 13 Flächenländern. Zunächst eine Aufschlüsselung der Anzahl der Kommunen mit mehr als 2.000 Einwohnern nach Ländern und Regionen (siehe Tabelle 2).

Tabelle 2: Grundgesamtheit (Anzahl Kommunen ab 2.000 Einwohnern + Kreise) 2020, nach Ländern und Regionen

Land	Grundgesamtheit	Land	Grundgesamtheit	Region	Grundgesamtheit
Brandenburg	209	Rheinland-Pfalz	361	NRW	427
Baden-Württemberg	960	Schleswig-Holstein	225	Norden	771
Bayern	1.395	Saarland	58	Osten	1.029
Hessen	434	Sachsen	338	Süd-West	853
Mecklenburg-Vorpommern	137	Sachsen-Anhalt	148	Süden	2.355
Niedersachsen	546	Thüringen	197		
Nordrhein-Westfalen	427				

Quelle: Statistisches Bundesamt (2021a), Statistisches Bundesamt (2021b).

4.1. Regionalisierung auf Ebene der Regionen

Eine Regionalisierung auf Ebene der Regionen erscheint mit den aktuellen Daten angesichts der Anzahl der Rückläufer pro Region generell als möglich. Damit die Hochrechnung und Gewichtung funktionieren, muss grundsätzlich davon ausgegangen werden, dass die Stichproben der Untergruppen, die u. a. auf Grundlage der Regionen abgebildet werden, repräsentativ für diese Untergruppen sind. Da für das KfW-Kommunalpanel Adressdaten des Statistischen Bundesamtes genutzt werden, hat prinzipiell jede Kommune der Grundgesamtheit die Chance, Teil der Stichprobe zu werden. Insofern sollte das Risiko eines Non-Coverage-Bias gering sein. Zudem haben bisherige Analysen im Rahmen des KfW-Kommunalpanels gezeigt, dass sich für die Gruppen, die für die Hochrechnung verwendet werden, in der Regel keine signifikanten Unterschiede zwischen Teilnehmern und Nichtteilnehmern in Bezug auf verschiedene relevante Strukturvariablen (z. B. Pro-Kopf-Verschuldung, Steuerkraft, Einwohnerzahl, etc.) ergeben (Krone und Scheller 2018).

Werden auf Basis konkreter Merkmale des Befragungsjahres 2020 exemplarische Mindestnettostichproben nach dem oben beschriebenen Muster berechnet, so zeigt sich für das stetige Merkmal „Geplante Investitionen für das aktuelle Haushaltsjahr“, dass die Mindestnettostichprobe in allen fünf Regionen erreicht wird, sofern die erlaubte absolute Abweichung („acceptable margin of error“) bei einem Alpha-Niveau von 0,1 auf einen Wert von 58 Euro pro Einwohner festgelegt wird (siehe Tabelle 3).⁶ Im Gegensatz dazu sind repräsentative Aussagen für Gesamtdeutschland im betrachteten Merkmal mit einem höheren Grad an Präzision (+/- 25 Euro) möglich.

Tabelle 3: Mindeststichprobe für repräsentative Aussagen für das Merkmal “Für das aktuelle Haushaltsjahr geplante Investitionen (in Euro pro Einwohner)” des Befragungsjahres 2020⁷

Bundesland	Mindeststichprobe	Antworten	Sind Aussagen zur Verteilung repräsentativ für die Grundgesamtheit?
BB	33	15	Nein
BW	104	83	Nein
BY	83	100	Ja
HE	49	37	Nein
MV	9	11	Ja
NI	54	63	Ja
NW	45	62	Ja
RP	36	31	Nein
SH	93	18	Nein
SL	25	5	Nein
SN	67	21	Nein
ST	73	31	Nein
TH	106	19	Nein
Region			
Süden	95	183	Ja
Osten	87	97	Ja
Süd-West	42	73	Ja
NRW	45	62	Ja
Norden	79	81	Ja
Deutschland	443	496	Ja

⁶ Interpretieren lässt sich dieser Wert folgendermaßen: Liegt in einem beliebigen Bundesland der gewichtete empirische Mittelwert der für das aktuelle Haushaltsjahr geplanten Investitionen bei 500 Euro pro Einwohner, so ist davon auszugehen, dass der wahre Mittelwert innerhalb einer Bandbreite von +/- 58 Euro zu finden ist (Bartlett 2001, S. 45; von der Lippe 2011, S. 15). Für die dargestellte Variable bewegen sich die empirischen Mittelwerte zwischen 192 (Rheinland-Pfalz) und 522 (Bayern).

⁷ Alpha-Niveau: 0,1; erlaubter absoluter Fehler für Bundesländer und Regionen: +/-58 Euro pro Einwohner. Für Gesamtdeutschland: +/- 25 Euro pro Einwohner. Die Parameter wurden so festgelegt, dass als Mindestkriterium repräsentative Aussagen für alle Regionen möglich sind.

Bei der Berechnung der Mindestnettostichprobengröße des Merkmals „Bewertung der aktuellen Finanzsituation“, das gemäß Bartlett et al. (2001, S. 44) als stetig aufgefasst wird und eine Skalierung von fünf Antwortoptionen hat, lassen sich für die fünf Regionen repräsentative Aussagen z. B. über den Mittelwert mit relativ hoher Präzision (absoluter erlaubter Fehler: +/-0,25) treffen.⁸

Auf Ebene der Bundesländer ist bei diesem Präzisionsniveau nicht für jedes Land eine repräsentative Aussage möglich. Für die Bundesebene hingegen sind repräsentative Aussagen bereits ab einem erlaubten Fehler von +/-0,15 möglich. Ursache dafür, dass regionalisierte Aussagen in diesem Merkmal mit deutlich höherer Präzision möglich sind, sind die generell höheren Antwortzahlen sowie geringere Varianzen in den einzelnen Schichten.

Tabelle 4: Mindeststichprobe für repräsentative Aussagen in verschiedenen geografischen Zuschnitten für das Merkmal „Bewertung der gegenwärtigen Finanzsituation“ des Befragungsjahres 2020⁹

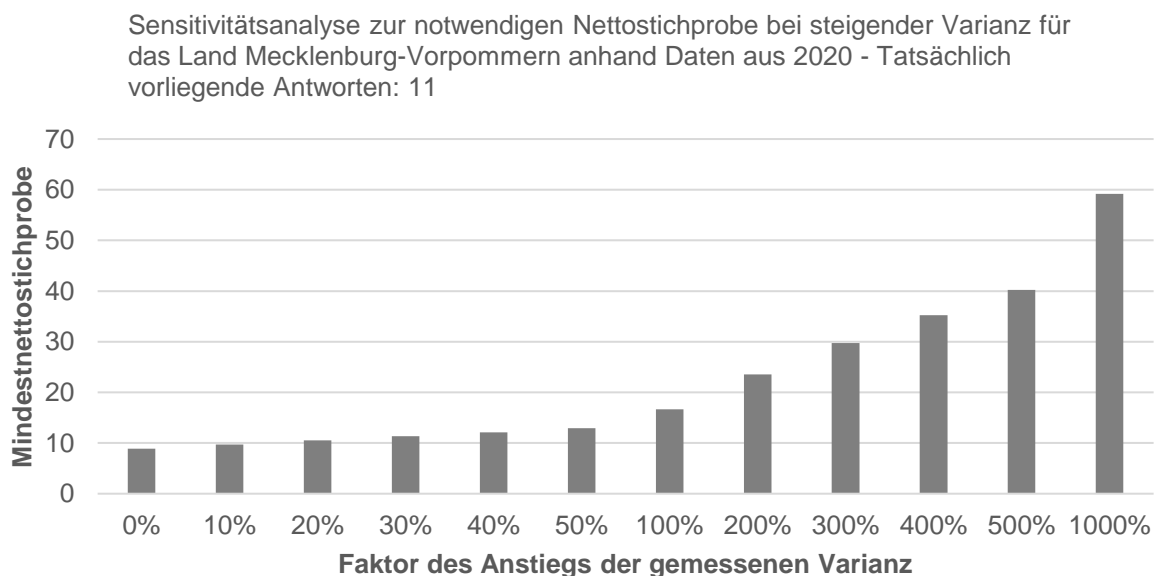
Bundesland	Mindeststichprobe	Antworten	Sind Aussagen zur Verteilung repräsentativ für die Grundgesamtheit?
BB	31	24	Nein
BW	53	124	Ja
BY	45	153	Ja
HE	54	53	Nein
MV	45	13	Nein
NI	43	98	Ja
NW	43	110	Ja
RP	98	48	Nein
SH	71	25	Nein
SL	34	10	Nein
SN	53	31	Nein
ST	36	44	Ja
TH	43	27	Nein
Region			
Süden	49	277	Ja
Osten	60	139	Ja
Süd-West	85	111	Ja
NRW	43	110	Ja
Norden	63	123	Ja
Deutschland	174	760	Ja

⁸ Interpretieren lässt sich dieser Wert folgendermaßen: Liegt in einem beliebigen Bundesland der gewichtete empirische Mittelwert der von allen antwortenden Kommunen angegebenen Finanzsituation (1 = „sehr gut“, ..., 5 = „mangelhaft“) bei 3 (= „befriedigend“), so ist davon auszugehen, dass der wahre Mittelwert innerhalb einer Bandbreite von 2,75 und 3,25 zu finden ist. Die empirischen Mittelwerte der einzelnen Bundesländer bewegen sich zwischen 2,92 (Bayern) und 4,45 (Sachsen-Anhalt).

⁹ Alpha-Niveau: 0,1; erlaubter absoluter Fehler für Bundesländer und Regionen: +/-0,25. Für Gesamtdeutschland: +/-0,15. Die Parameter wurden so festgelegt, dass als Mindestkriterium repräsentative Aussagen für alle Regionen möglich sind.

Zu den in den Tabellen 3 und 4 dargestellten Berechnungsergebnissen für die Mindestnettostichprobengröße ist anzumerken, dass sie auf der Annahme basieren, dass die empirische Varianz innerhalb einer Schicht der Nettostichprobe des Befragungsjahres 2020 der wahren Varianz derselben Schicht der Grundgesamtheit entspricht; diese Annahme ist vor allem in Schichten mit geringen absoluten Antwortzahlen leicht angreifbar. Ein Beispiel liefert dafür etwa die berechnete Mindestnettostichprobe für das Bundesland Mecklenburg-Vorpommern in Tabelle 3 (Variable: Für das aktuelle Haushaltsjahr geplante Investitionen (in Mio. Euro) im Jahr 2020). Hier legen die Antworten der elf Kommunen eine Varianz in den beiden Schichten (Kommunen von 2.000 bis 19.999 EW sowie Kommunen ab 20.000 EW) nahe, die so gering ist, dass unter den getroffenen Annahmen über den erlaubten absoluten Fehler und das Alpha-Niveau bereits die Antwort von neun Kommunen ausreichend wären, um repräsentative Aussagen zu treffen. Eine Sensitivitätsanalyse zeigt jedoch, dass die wahre, aber unbekannte Varianz in der Schicht nur um 30% über der tatsächlich gemessenen Varianz liegen müsste, damit der Umfang der notwendigen Nettostichprobe die tatsächliche Nettostichprobe des Befragungsjahres 2020 übersteigt. Angesichts der Aussage über das Erreichen der Mindestnettostichprobe für Mecklenburg-Vorpommern in Tabelle 3 ist also Vorsicht geboten. Insofern sind die dargestellten Werte lediglich als erste Annäherung über die Mindestgröße der Nettostichprobe zu verstehen.

Abbildung 1: Sensitivitätsanalyse zur angenommenen Varianz hinter der in Tabelle 3 dargestellten Mindestnettostichprobe für das Bundesland Mecklenburg-Vorpommern.



Bei der erlaubten absoluten Abweichung erreichen auch die Bundesländer Bayern (100 Antworten), Niedersachsen (63 Antworten) und Nordrhein-Westfalen (62 Antworten)

akzeptable Mindestnettostichprobengrößen. Hier jedoch basiert die empirische Varianz auf deutlich mehr Antworten. Zudem ist die Differenz zwischen der Mindestnettostichprobe und den tatsächlichen Antworten größer als in Mecklenburg-Vorpommern, weshalb die berechneten Werte generell eine höhere Validität aufweisen.

Der Blick auf die Tabelle 4 zeigt, dass gerade auf Ebene der Regionen die empirische Varianz auf deutlich mehr Antworten beruht und die Differenz zwischen der berechneten Mindestnettostichprobe und der tatsächlichen Zahl der Antworten pro Region hoch ist. Dieses Verhältnis deutet darauf hin, dass selbst bei einer größeren Unterschätzung der wahren von der empirischen Varianz die Repräsentativität von auf die Regionen bezogenen Aussagen dieser Variable nicht infrage gestellt wäre. Die Sensitivitätsanalyse am Beispiel der Region Osten zeigt, dass repräsentative Aussagen selbst dann noch mit relativ hoher Präzision (erlaubte absolute Abweichung von +/- 0,25) möglich wären, wenn die wahre Varianz in den einzelnen Schichten drei Mal so hoch läge wie es die empirische Varianz nahelegt. Der Blick auf die Nettostichproben nach Regionen (Tabelle 5) zeigt, dass die Zahl der teilnehmenden Kommunen vor allem für die Jahre seit 2017 grundsätzlich ausreichend groß erscheint, um repräsentative Aussagen zu treffen.¹⁰ In den früheren Jahren mit relativ geringeren Antwortzahlen ist der Grad an Unsicherheit entsprechend höher. Zusammenfassend erscheinen die Teilstichproben groß genug, um eine Regionalisierung auf Ebene der Regionen zu verantworten.

Tabelle 5: Nettostichproben 2009 bis 2020, nach Regionen.

Region	2009	2010	2011	2012	2013	2014	2015
NRW	85	72	77	85	84	81	65
Norden	77	68	114	93	74	69	63
Osten	111	96	115	80	83	71	60
Süd-West	57	64	76	67	57	67	51
Süden	181	250	237	200	181	187	185

Region	2016	2017	2018	2019	2020
NRW	82	100	141	122	111
Norden	73	65	127	121	124
Osten	61	63	121	144	140
Süd-West	55	78	102	96	111
Süden	206	296	283	253	279

Quelle: Datenauszug aus dem KfW-Kommunalpanel.

¹⁰ In den Jahren 2012 bis 2016 war ein deutlicher Rückgang der Stichprobengröße festzustellen. Vor allem in den letzten Jahren seit 2017 hat sich die Stichprobengröße aber wieder vergrößert und stabilisiert. Ursache dafür waren aktive Bestrebungen, um mehr Kommunen zur Teilnahme zu bewegen. Seit 2017 ist es z. B. möglich, auch online an der Befragung teilzunehmen.

4.2. Regionalisierung auf Ebene der Länder

Dass die Regionalisierung auf Ebene der Länder nur eingeschränkt bzw. mit einem entsprechend hohen Grad an Unsicherheit möglich wäre, wurde schon im vorangegangenen Abschnitt deutlich (siehe Tabellen 3 und 4). Auf Basis der Nettostichprobe in den einzelnen Bundesländern lässt sich erkennen, dass kritische Schwellenwerte selbst bei einem sehr guten Antwortverhalten bei einer Frage nur in einer Minderheit der Bundesländer möglich wäre, hier insbesondere Baden-Württemberg, Bayern und Nordrhein-Westfalen. In den restlichen Bundesländern ist die Zahl der Antwortenden teilweise deutlich zu niedrig, um anhand der Antworten Aussagen mit ausreichender Validität geben zu können. Bei solch geringen Fallzahlen ist die Repräsentativität der Ergebnisse auf Länderebene generell unwahrscheinlich, da die Schätzer wegen der geringen Fallzahlen mit zu großer Streuung um den wahren Wert der Grundgesamtheit schwanken. Insofern ist von einer Analyse auf Länderebene allein auf Basis der für das jeweilige Land vorliegenden Daten abzuraten.

Wichtig in diesem Zusammenhang ist, dass bei den vorliegenden Stichprobengrößen die Repräsentativität für Gesamtdeutschland gegeben ist, da die Größe der notwendigen Mindeststichprobe ein beschränktes Wachstum mit Blick auf die Grundgesamtheit aufweist. Das heißt, dass der Umfang der Mindeststichprobe nicht analog zur Grundgesamtheit wächst, sondern nach einem initial starken Wachstum schnell abflacht und darum nach oben hin beschränkt ist. In der Sampling-Praxis lässt sich somit schlussfolgern, dass die Mindeststichprobe für Gesamtdeutschland im Rahmen des KfW-Kommunalpanels erreicht wird, während die Fallzahlen auf Bundesländer-Ebene oftmals dahinter zurückbleiben.

5. Weitere Ansätze zur Regionalisierung

Welche weiteren Ansätze gibt es, aggregierte Ergebnisse einer Panelbefragung zu regionalisieren? Diese Fragestellung scheint vor allem im Hinblick auf die Regionalisierung auf Länderebene interessant, da dort das KfW-Kommunalpanel bislang nicht für alle Länder regionale Auswertungen zulässt. Unabhängig von den in der Folge vorgeschlagenen Herangehensweisen ist immer eine Ausweitung der Stichprobe eine Option. Denn diese liefert – ebenso wie die folgenden Methoden – eine zuverlässige Möglichkeit zur Regionalisierung und ist darüber hinaus intuitiv verständlich.

5.1. Erhöhung der Fallzahl durch Daten-Pooling

Ebenso besteht die Möglichkeit, Stichproben aus verschiedenen Jahren zusammenzufassen und so den Umstand zu nutzen, dass manche Kommunen zwar nicht in allen, aber in einigen Jahren antworten.

Tabelle 7: Nettostichproben 2018 bis 2020 und Anzahl der Kommunen, die in mindestens einem Jahr teilgenommen haben, nach Ländern

Land	2018	2019	2020	Anzahl der Kommunen mit mind. einer Teilnahme in 2018 bis 20
Brandenburg	34	20	24	49
Baden-Württemberg	114	116	125	204
Bayern	169	137	154	257
Hessen	53	41	53	85
Mecklenburg-Vorpommern	11	16	13	21
Niedersachsen	100	99	99	174
Nordrhein-Westfalen	141	122	111	193
Rheinland-Pfalz	34	41	48	73
Schleswig-Holstein	27	22	25	44
Saarland	15	14	10	21
Sachsen	23	41	32	62
Sachsen-Anhalt	34	33	44	65
Thüringen	19	34	27	49

Quelle: Datenauszug aus dem KfW-Kommunalpanel.

Beispielsweise könnte man eine zusammengefasste Auswertung von derart gepoolten Daten für die Jahre 2018 bis 2020 vornehmen. Darin wären dann alle Kommunen enthalten, die in mindestens einem der drei Jahre an der Befragung teilgenommen haben. Für die Kommunen, die an mehreren Befragungen teilgenommen haben, wird der Mittelwert der Antworten genutzt bzw. bei qualitativen Fragen etwa zum Zustand der Infrastruktur die zuletzt gegebene Antwort herangezogen.

Wie in Tabelle 7 erkennbar ist, vergrößert sich dadurch die Nettostichprobe. Entsprechend dürfte in der Regel auch die Antwortzahl in den einzelnen Merkmalen der Befragung ansteigen. In der Konsequenz wächst damit mutmaßlich auch die Zahl der Länder, für die repräsentative Ergebnisse vorgelegt werden können.

Um Anhaltspunkte darüber zu bekommen, wie sich ein solcher „Pooling-Ansatz“ auf die Mindestnettostichproben pro Bundesland auswirken kann, wurden diese gemäß oben beschriebenen Ansatz für die Variable „Qualitative Bewertung des Unterhalts, gesamt“ mit gepoolten Daten der Jahre 2018 bis 2020 berechnet. Bei der Variable handelt es sich um eine auf fünf Antwortoptionen skalierte Variable (Bartlett et al. 2001, S. 44).

Tabelle 8: Mindeststichprobe für repräsentative Aussagen in verschiedenen geografischen Zuschnitten für das Merkmal „Laufender Unterhalt in den vergangenen fünf Jahren“ des Befragungsjahres 2020¹¹

Bundesland	Mindeststichprobe	Antworten	Sind Aussagen zur Verteilung repräsentativ für die Grundgesamtheit?
BB	77	39	Nein
BW	79	157	Ja
BY	39	184	Ja
HE	41	63	Ja
MV	68	17	Nein
NI	32	140	Ja
NW	45	154	Ja
RP	144	56	Nein
SH	50	35	Nein
SL	26	16	Nein
SN	95	50	Nein
ST	28	56	Ja
TH	39	40	Ja
Region			
Süden	56	341	Ja
Osten	107	202	Ja
Süd-West	123	135	Ja
NRW	45	154	Ja
Norden	45	175	Ja
Deutschland	84	1.007	Ja

Die Ergebnisse (Tabelle 8) zeigen, dass ein Pooling der Daten dazu führt, dass bei einem angenommenen erlaubten Fehler von +/- 0,2 und der Annahme, dass die empirische Varianz der jeweiligen Schicht der wahren Varianz in der Grundgesamtheit entspricht, repräsentative Aussagen für sieben von 13 Flächenländern getroffen werden können.

¹¹ Alpha-Niveau: 0,1; erlaubter absoluter Fehler für Bundesländer, Regionen und Gesamtdeutschland: +/-0,2. Die Parameter wurden so festgelegt, dass als Mindestkriterium repräsentative Aussagen für alle Regionen möglich sind.

Bis auf ein Bundesland (Thüringen) ist die Differenz zwischen der notwendigen Mindestnettostichprobe und der tatsächlichen Zahl der vorliegenden Antworten so groß, dass selbst eine gravierende Unterschätzung der tatsächlichen Varianz anhand der vorliegenden Daten die Validität der Aussagen nicht beeinträchtigen würde. Ist man dazu bereit, die Annahmen zum erlaubten Fehler bzw. zum Alpha-Niveau aufzuweichen, so erhöht sich dadurch die Zahl der Bundesländer, für die repräsentative Aussagen über das Merkmal getroffen werden können. Damit geht jedoch zugleich ein höherer Grad an Unsicherheit einher.

5.2. Alternative Methoden zum Imputieren fehlender Daten

Eine alternative Möglichkeit zur direkten Schätzung mit der Nettostichprobe in den einzelnen Ländern ist – insbesondere bei fehlenden oder unzureichenden Datenpunkten –, die Daten von Kommunen aus anderen Ländern für eine Schätzung der Ergebnisse zu nutzen. Da man verschiedene Länder mit ähnlichen Merkmalen in einer Region gruppiert und bei der Schätzung nur die Regionen, nicht aber die Länder beachtet, trifft man implizit die Annahme, dass sich die Ergebnisse zwischen den Ländern in einer Region nicht systematisch unterscheiden.¹² Daher könnte man die anderen Länder derselben Region nutzen, um fehlende Werte zu schätzen, z. B. für das Saarland auf Basis der Daten von Hessen und Rheinland-Pfalz (Region Süd-West). Dafür existieren Verfahren, indem man z. B. die Werte der anderen Länder anhand von bestimmten Merkmalen gewichtet, die mit den Antworten korreliert sind. Diese Merkmale sollten von einer externen Quelle bezogen werden, z. B. über die Kassen- oder Rechnungsstatistik der statistischen Ämter.

Nearest-Neighbour-Methoden

Nearest-Neighbour-Methoden werden genutzt, um Zielwerte für eine Population zu berechnen, für die zwar diese Werte nicht vorliegen, aber Merkmale, die mit dem Zielwert korreliert sind, verfügbar sind. Dazu wird eine Population herangezogen, deren Merkmale und Zielwerte bekannt sind. Der fehlende Zielwert wird berechnet aus den vorhandenen Zielwerten von den Einheiten der Referenzgruppe, die den Merkmalen der fehlenden Einheit am ähnlichsten sind. Ähnlichkeit wird dabei über die Distanz zu den anderen Einheiten definiert, z. B. die Euklidische Distanz (Paass 1995). Im KfW-Kommunalpanel würde das bedeuten, dass für fehlende Werte einzelner Kommunen die Angaben anderer Kommunen herangezogen werden, die dieser strukturell möglichst ähnlich sind. Als

¹² Eine Grundlage der Gruppenbildung, die im Kommunalpanel für die Gewichtung bzw. Hochrechnung der Ergebnisse durchgeführt wird, ist die strukturelle Ähnlichkeit der Kommunen in den zu Regionen zusammengefassten Bundesländern (siehe auch Krone und Scheller 2020).

Strukturvariablen kämen z. B. Einwohnerzahl, die Steuerkraft oder die Arbeitslosenquote in Frage. Entsprechende Schätzungen kommen zum Beispiel für die Imputation fehlender Daten in Einkommenserhebungen in Frage, die zustande kommen, da gerade Bezieher hoher Einkommen häufiger Antworten verweigern als Normalverdiener (Rässler 2000).

Small-Area-Estimation

Methoden, die das Problem von geringen Fallzahlen in kleinen geografischen Einheiten behandeln, werden in der Literatur unter dem Begriff der Small-Area-Estimation (SAE) zusammengefasst (Ghosh und Rao 1994). Grundidee von SAE ist die Hinzunahme von zusätzlichen Statistiken (z. B. der amtlichen Statistik), um so die Stichprobe anzureichern. Diese zusätzlichen Statistiken werden dann genutzt, um den Wert von Interesse zu schätzen (Bäumer et al. 2018). Es handelt sich also um eine indirekte Schätzung und nicht um eine direkte Schätzung wie nach Horvitz und Thompson (1952), bei denen die erhobenen Stichprobendaten zur Schätzung genutzt werden. Dabei werden Modelle aufgestellt, die einen Funktionszusammenhang zwischen den Daten der Stichprobe und den zusätzlichen Statistiken herstellen und so auf Grundlage der für die geografische Region bekannten zusätzlichen Daten einen Wert schätzen (Asian Development Bank 2020). Grundsätzlich werden dabei zwei Typen von Modellen unterschieden: Einfache SAE-Modelle und regressionsbasierte Modelle (Asian Development Bank 2020). Ein Anwendungsbeispiel aus dem deutschen Kontext ist etwa die Schätzung kleinräumiger Verkehrskennzahlen mit Hilfe von Zensusdaten (Bäumer et al. 2018). Eine vergleichbare Schätzung im Rahmen des KfW-Kommunalpanels anhand amtlicher finanzstatistischer Daten wie etwa dem Realsteuervergleich, der für die vollständige Grundgesamtheit vorliegt, wäre prinzipiell vorstellbar. Damit einhergehen müsste jedoch eine Prüfung, ob die notwendige Schätzung mit hinreichender Präzision durchgeführt werden kann.

6. Fazit und Ausblick

Panelbefragungen sind eine Kombination aus Querschnitts- und Zeitreihendaten. Damit liefern sie wertvolle Einblicke in die zeitliche Entwicklung der beobachteten Einheiten. Während das KfW-Kommunalpanel als Panelbefragung angelegt ist, werden die Daten vorwiegend im jährlichen Querschnitt ausgewertet. Die Methodik zur Berechnung von Ergebnissen für Gesamtdeutschland beachtet dabei die Notwendigkeit einer Gewichtung und zieht bei Bedarf Ergänzungsstichproben. Insgesamt konnte durch die vorstehende Diskussion verdeutlicht werden, dass die für das KfW-Kommunalpanel verwandte Methodik ausreichend ist, um Repräsentativität für Gesamtdeutschland zu gewährleisten.

Die Regionalisierung der Ergebnisse für einzelne Regionen erscheint grundsätzlich machbar, wenngleich für manche Merkmale in der Befragung dies nur mit einer relativ hohen Unsicherheit möglich sein dürfte. Bereits in der Gewichtung und Hochrechnung für Gesamtdeutschland muss die Annahme getroffen werden, dass die Kommunen in den Regionen repräsentativ für diese sind. Dies wird unter Rückgriff auf flächendeckend verfügbare Daten der amtlichen Statistik sichergestellt. Außerdem sind die Fallzahlen in den einzelnen Regionen generell ausreichend groß, damit die Schätzer ihre Konvergenzeigenschaften ausspielen können.

Eine Regionalisierung auf Länderebene, die nur auf Grundlage der jeweiligen Teilstichproben der einzelnen Länder durchgeführt wird, ist hingegen nur bedingt empfehlenswert. Aufgrund der geringen Fallzahlen in einigen Ländern können für diese keine repräsentativen Ergebnisse geschätzt werden, die Schätzer unterlägen der Gefahr der Ungenauigkeit sowie der Verzerrung und hätten eine zu große Varianz. Dies schließt Auswertungen und Hochrechnungen für einzelne Bundesländer mit ausreichend hohen Fallzahlen nicht aus, würde sich aber der Kritik stellen müssen, warum Bundesländer mit geringeren Fallzahlen nicht auch ausgewertet werden.

Allerdings gibt es weitere Möglichkeiten, repräsentative Ergebnisse für die Regionen und auch für die Länder zu erhalten. Für ausgewählte Fragestellungen hilfreich erscheinen dabei die Methoden der Small-Area-Estimation und der Nearest-Neighbours. Bei ersterer werden die Daten aus der Stichprobe mit externen Statistiken angereichert. Durch die Herstellung eines Zusammenhangs zwischen den externen Statistiken und den Daten der Stichprobe werden die fehlenden Daten der Stichprobe geschätzt. Bei der Methode der Nearest Neighbours werden ähnliche Observations im Datensatz genutzt, um die fehlenden Daten zu imputieren.

Darüber hinaus ist es denkbar, Durchschnittswerte mehrerer Jahre zu nutzen, um die Nettostichprobe zu vergrößern und so eine Regionalisierbarkeit auf Ebene der Länder zu erreichen. Die Bewertung einer Mindestnettostichprobe für einzelne Bundesländer anhand gepoolter Daten zeigt jedoch, dass die Limitationen auf Ebene der Bundesländer auch damit nicht vollständig aufgelöst werden können.

Mit Blick auf die Zukunft erscheint die Ausweitung der Bruttostichprobe darum als das geeignetste Mittel, um die Nettostichprobe zu erhöhen und so – bei einem entsprechend guten Rücklauf – die notwendigen Fallzahlen zu erreichen, damit den Anforderungen an Mindeststichprobengrößen auf Länderebene genüge getan wird. Gleichzeitig unterliegt dies jedoch der Schwierigkeit, dass Kommunen zur Teilnahme an immer mehr Befragungen von unterschiedlichen Seiten aufgerufen werden. In Kombination mit Personalengpässen in der Kommunalverwaltung ist darum anzunehmen, dass die Teilnahmebereitschaft über die Zeit sinkt. Dies trifft besonders in den krisenhaften Situationen der vergangenen Jahre zu, die Kämmerereien in immer schnellerer Taktung vor kurzfristig zu lösende Herausforderungen stellen.

In jedem Fall ist die Veröffentlichung von Ergebnissen, die auf Regionen oder Bundesländer herunter gebrochen werden, mit zusätzlichen Herausforderungen in der Interpretation und Vermittlung der Ergebnisse verbunden. So gilt es beispielsweise politische oder kommunalstrukturelle Besonderheiten in den Regionen zu berücksichtigen, wenn die Ergebnisse bewertet oder zur Formulierung von Handlungsempfehlungen herangezogen werden sollen. Andernfalls wäre ein Nebeneinanderstellen der regionalisierten Ergebnisse nicht ohne weiteres möglich bzw. würde leicht zu unzulässigen Vergleichen einladen. So sind ausgewiesene Euro-Beträge eigentlich nur aussagekräftig, wenn sie in Bezug gesetzt werden zu Haushaltsgrößen (wie Einnahme- bzw. Steuerkraft), örtlichen Preisniveaus oder auch normativen Anforderungen/Bedarfen, die sich von Region zu Region unterscheiden können. Dem gegenüber lassen sich qualitative Einschätzungen, wie sie im KfW-Kommunalpanel ebenfalls erhoben werden, leichter überregional vergleichen, wobei natürlich auch dies vor dem Hintergrund der jeweiligen Ausgangslage der antwortenden Befragungsteilnehmer geschehen sollte.

Literaturverzeichnis

- Asian Development Bank (2020). Introduction to Small Area Estimation Techniques. Metro Manila. DOI: <http://dx.doi.org/10.22617/TIM200160-2>.
- Bartlett, J.E.; Kotlik, J.W.; Higgins, C.C. (2001). Organizational Research: Determining Appropriate Sample Size in Survey Research. *Information Technology, Learning, and Performance Journal*, 19(1).
- Bäumer, M.; Hautzinger, H.; Pfeiffer, M. (2018). Mobilität in Deutschland 2017: Regionalisierung von MiD-Ergebnissen. Small-Area-Methoden zur Schätzung von Verkehrskennzahlen in kleinräumiger Gliederung. Online verfügbar unter http://www.mobilitaet-in-deutschland.de/pdf/MiD2017_Bericht_Regionalisierung_MiD-Ergebnisse_Small_Area-Verfahren_1218.pdf.
- Brand, S.; Steinbrecher, J. (2019). Doppelt hält besser: Umfragen als sinnvolle Ergänzung der amtlichen Statistik. Online verfügbar unter <https://www.kfw.de/PDF/Download-Center/Konzernthemen/Research/PDF-Dokumente-Fokus-Volkswirtschaft/Fokus-2019/Fokus-Nr.-253-Juni-2019-Doppelt-haelt-besser.pdf>.
- Brand, S.; Krone, E.; Scheller, H.; Steinbrecher, J. (2020). Zur Messung kommunaler Investitionsrückstände – zwischen begrifflichen Abgrenzungsschwierigkeiten, statistischen Brüchen und notwendiger Methodenvielfalt, *Jahrbuch für öffentliche Finanzen 2-2020*, S. 229-246.
- Brüderl, J. (2010). Kausalanalyse mit Paneldaten. In: Wolf C., Best H. (eds) *Handbuch der sozialwissenschaftlichen Datenanalyse*. VS Verlag für Sozialwissenschaften, Berlin.
- Cochran, W. (1977). *Sampling Techniques*, 3. Auflage. Wiley.
- Fitzgerald, J.; Gottschalk, P.; Moffitt, R. (1996). An Analysis of Sample Attrition in Panel Data: The Panel Study of Income Dynamics. mimeo, Johns Hopkins University.
- Ghosh, M.; Rao, J. N. (1994). Small area estimation: an appraisal. *Statistical science*, 9(1), 55-76.
- Headey, B. (Ed.); Holst, E. (Ed.) (2008). *A Quarter Century of Change: Results from the German Socio-Economic Panel (SOEP)*, SOEP Wave Report No. 1-2008, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin.
- Horvitz, D. G.; Thompson, D. J. (1952): A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260), 663-685.
- Koch, A.; Blohm, M. (2015). Nonresponse bias, *GESIS–Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines)*, Mannheim.
- Krone, E.; Scheller, H. (2020). 10 Jahre KfW-Kommunalpanel: Methodik der Datenerhebung und -aufbereitung. Online verfügbar unter <https://www.kfw.de/PDF/Download-Center/Konzernthemen/Research/PDF-Dokumente-Studien-und-Materialien/Methodenpapier-KfW-Kommunalpanel.pdf>.

Landesrechnungshof (LRH) Niedersachsen (2021). Kommunalbericht 2021. Online verfügbar unter <https://www.lrh.niedersachsen.de/download/174846>.

Prein, G.; Kluge, S.; Kelle, U. (1994). Strategien zur Sicherung von Repräsentativität und Stichprobenvalidität bei kleinen Samples.

Paass G. (1985). Statistical record linkage methodology: state of the art and future prospects. Bulletin of the International Statistical Society. Proceedings of the 45th Session. ISI, Voorburg NL.

Rässler, S. (2000). Ergänzung fehlender Daten in Umfragen. Jahrbücher für Nationalökonomie und Statistik, 220(1), S. 64-94.

Rendtel, U. (1993). Über die Repräsentativität von Panelstichproben: Eine Analyse der feldbedingten Ausfälle im Sozio-Oekonomischen Panel (SOEP), DIW Discussion Papers No. 70, Deutsches Institut für Wirtschaftsforschung, Berlin.

Scheller, H.; Schneider, S. (2016). KfW-Kommunalpanel 2016. Online verfügbar unter <https://www.kfw.de/PDF/Download-Center/Konzernthemen/Research/PDF-Dokumente-KfW-Kommunalpanel/KfW-Kommunalpanel-2016.pdf>.

Statistisches Bundesamt (2021a). Daten aus dem Gemeindeverzeichnis Kreisfreie Städte und Landkreise nach Fläche, Bevölkerung und Bevölkerungsdichte.

Statistisches Bundesamt (2021b). Daten aus dem Gemeindeverzeichnis Gemeinden in den Ländern nach Einwohnergrößenklassen.

Sydow, J.; Hesse, M.; Kratzmann, A. (2020). Kommunaler Investitionsbedarf im Freistaat Sachsen – Kommunalbefragung 2020. KOMKOS Analyse Nr. 17. Leipzig.

Vollnhals, S. (2017). Wahlumfragen auf der Basis von Access-Panels. Wie „Propensity Score“-Gewichtungen die Repräsentativität erhöhen können. Dissertation an der Johannes-Gutenberg-Universität, Mainz.

Von der Lippe, P.; Kadroba, A. (2002). Repräsentativität von Stichproben. Marketing, 24(1), S. 227-238.