

# Ungleich verteilte Corona-Infektionen zwischen den Bundesländern

Olaf Hübler

**Juni 2021**

## **Zusammenfassung**

In diesem Beitrag wird die regionale Ausbreitung von COVID-19-Infektionen untersucht. Unter Verwendung der Hauptkomponentenanalyse, des LARS- und RLASSO-Auswahlverfahrens erfolgt eine Variablenreduktion. Geprüft wird die Bedeutung von Zustandsabhängigkeit, unbeobachteter Heterogenität und Strukturbrüchen. Die empirische Analyse zeigt, dass sowohl regionale Strukturvariablen als auch regional aggregierte Persönlichkeitsmerkmale bedeutsam sind für die unterschiedliche Corona-Ausbreitung. Die nord-östlichen Bundesländer weisen einen geringeren Grad an Betroffenheit auf. Regionen mit hohem Migrantenanteil zeigen eine höhere Inzidenz als andere. Werden Persönlichkeitsmerkmale vernachlässigt, so wird die Bedeutung des Migrationseinflusses überschätzt. Mit der Schulbildung, der Armutgefährdung und der Haushaltsgröße wurden drei weitere wichtige Merkmale identifiziert. In Bundesländern mit überproportional vielen Menschen ohne Schulabschluss werden tendenziell weniger COVID-19-Fälle ausgewiesen. Je mehr Kooperationsbereitschaft und emotionale Labilität ausgeprägt sind, umso höher ist die Gefahr der Ansteckung. Ein positiv signifikanter Zusammenhang zwischen Infektionen und Tests wird durch die Schätzungen abgebildet. Eine weniger klare Verknüpfung zeigt sich zwischen Impfungen und der Zahl der Infektionen. Über die drei Corona-Wellen hinweg, offenbaren sich deutliche Veränderungen. Dies betrifft die Bedeutung des Migrantenanteil, der armutsgefährdete Familien und die geographische Lage der Bundesländer.

**JEL-Klassifikation:** C21, C23, I12, R12

**Schlagworte:** COVID-19, Bundesländer, regionale Merkmale, Persönlichkeitsmerkmale, Impfungen, PCA-Tests, Hauptkomponentenanalyse, Machine Learning, cluster-robuste Schätzung, Zustandsabhängigkeit, unbeobachtete Merkmale, Heterogenität, Corona-Wellen, Strukturbruch

Wirtschaftswissenschaftliche Fakultät, Leibniz-Universität Hannover, Telefon: +49 (0)511 762 4794;  
Email: [huebler@ewifo.uni-hannover.de](mailto:huebler@ewifo.uni-hannover.de); Internet: [www.wiwi.uni-hannover.de/de/fakultaet/emeriti/](http://www.wiwi.uni-hannover.de/de/fakultaet/emeriti/)

# 1 Einleitung

Es ist offensichtlich, dass die Verbreitung von Covid-19 regional und international stark streut. Innerhalb Deutschlands weisen besonders Schleswig-Holstein und Mecklenburg-Vorpommern eine günstige Entwicklung auf, während Sachsen und vor allem Thüringen bis vor kurzem noch stärker auf einem hohen Infektionsniveau verharrten als andere Bundesländer.

Verschiedene Erklärungen wurden und werden für die Ausbreitung von COVID-19 und die regional unterschiedliche Betroffenheit angeboten. So wurde während der ersten Welle häufig ein importierter Virus ins Feld geführt. Insbesondere Urlaubsreisen ins Ausland wurden als Erklärungsgrund genannt. Das Phänomen Ischgl war in aller Munde. Hotspots innerhalb Deutschlands, ausgelöst in einzelnen Unternehmen wie beim Schlachthof Tönnies, oder die Karnevalssitzung in der Gemeinde Gangelt im Kreis Heinsberg am 15. Februar 2020, bei der sich fast 44% der Anwesenden ansteckten und die der Virologe Streeck als erstes Superspreader-Event in Deutschland bezeichnete, führten zu Schlagzeilen in den Medien.

Da sich die zu Beginn der Pandemie geäußerten Hypothesen nicht als dauerhaft tragfähig erwiesen, wurden Alternativen angeboten, die jedoch nach kurzer Zeit auch nicht mehr befriedigen konnten. Noch heute erscheint die Erklärung des grenzüberschreitenden Verkehrs als teilweise tragfähig, wenn man sich die Deutschlandkarte, eingeteilt nach Landkreisen und der Zahl der Neuinfektionsfälle der letzten 7 Tage ansieht. Insbesondere im Osten und Süden Deutschlands zeigen sich an den Grenzen besonders stark betroffene Gebiete.

Als der partielle Lockdown im November nicht zu dem gewünschten Erfolg führte, wurde und wird zunehmend mit individuellen Einstellungen und Verhaltensweisen argumentiert. Empirisch sind bisher kaum eindeutige Muster identifiziert worden. Das liegt zum Teil auch daran, dass Teile der Bevölkerung zwar grundsätzlich Maßnahmen und auch strengere Auflagen zur Eindämmung der Pandemie befürworten, sich aber selbst nicht unbedingt an Gebote und Verbote halten. Mit Fortdauer der Pandemie nimmt die individuelle Bereitschaft ab, kurzfristig auf Freiheitsrechte zu verzichten, um gesamtgesellschaftlich gesundheitliche Schäden einzudämmen. Bei den mehr personenbezogenen Erklärungen fällt es zunächst schwer, regionale Inzidenzunterschiede dadurch zu erklären. Ins Feld geführt werden könnte hier jedoch, dass politisch geprägte Einstellungen, die sich regional ungleich verteilen, die individuelle Einhaltung der Auflagen und damit die Infektionsverbreitung beeinflussen (FAZ 09.12.20). So hat das Institut für Demokratie und Zivilgesellschaft (IDZ) und die NZZ in einem Bericht über die Stadt Köln (NZZ 15.05.2021) einen statistischen Zusammenhang zwischen dem AfD-Wahlergebnis und der Corona-Inzidenz gefunden. Es bedarf keiner weiteren Erläuterung, dass es sich dabei zunächst nur um eine statistische Korrelation handelt und nicht auf einen direkt kausalen Zusammenhang geschlossen werden kann. Gleichwohl ist dieses Ergebnis bemerkenswert. Als Begründung wird angeführt, dass AfD-Wähler weniger bereit sind, sich

impfen zu lassen als andere. Genauer erforscht ist dieser Zusammenhang allerdings noch nicht. Die Frage ist, ob sich derart kleinräumige Ergebnisse wie für den Raum Köln verallgemeinern lassen.

Bisweilen heißt es, das Infektionsgeschehen sei unklar und zwar immer dann, wenn die Gründe für regional begrenzte Hotspots nicht offensichtlich sind, so z.B. beim Hotspot im Landkreis Diepholz auf dem Spargel- und Beerenhof (HAZ 04.05.2012). Das RKI hat von einem diffusen Infektionsgeschehen gesprochen (02.12.2020), das sich nach seiner Meinung kaum noch nachvollziehen lässt. Im Gegensatz zur ersten Welle, als beispielsweise Ostdeutschland nur schwach betroffen war, hat sich das Virus flächendeckend, aber ungleichmäßig ausgebreitet. Die Zahl der Infektionen hat im November in Norddeutschland zwar auch erkennbar zugenommen, jedoch ist die Anstieg dort relativ moderat. Seit April-Mai 2021 zeichnet sich besonders für Norddeutschland ein stärkerer Rückgang bei der 7-Tageinzidenz ab und auch bei der Verdopplungszeit bestätigter Infektionen werden Vorteile des Nordens gegenüber anderen Regionen offenbar. Die Stadtstaaten Berlin und Hamburg weisen hier deutliche Vorteile auf, aber auch in allen anderen Bundesländern wird eine Entschärfung des Pandemieproblems sichtbar. Insgesamt erfährt die Diskussion um Hotspots heute weniger Aufmerksamkeit als zu Beginn der Pandemie, obwohl sich immer wieder solche herausbilden und soziale Brennpunkte zu einem neuen Schwerpunkt der Impfpolitik geworden sind. Mit ihnen können jedoch die interregionalen Unterschiede nicht vollständig erklärt werden.

Unklar bleibt, ob die regionale Verteilung in Deutschland einem ständigen Wandel unterliegt oder ob die meist nur unidimensional vorgebrachten Theorien unbefriedigend sind, ob nicht vielmehr ein komplexes Wirkungsgeflecht zum Tragen kommt. Das Robert-Koch-Institut (RKI) hat in diesem Zusammenhang (25.10. 2020 bzw. 23.10.2020) von einer diffusen Ausbreitung des Corona-Virus gesprochen. Es kann sich dabei um Cluster handeln, die wir aber (noch) nicht kennen.

Es fehlt an empirischen Analysen, die systematisch aufzeigen, warum es zu regionalen Disparitäten kommt, welche Kräfte entscheidend dafür sind, in welchem Maße die Ursachen einem Wandel seit Beginn der Pandemie unterliegen. Anliegen dieses Beitrags ist, empirisch unter Einbeziehung eines möglichst breiten Feldes an potentiellen direkten und indirekten Einflussgrößen herauszufinden, welche Determinanten statistisch einen wesentlichen Erklärungsbeitrag liefern und welche vernachlässigbar sind, warum die Pandemie im Raum ungleich verteilt ist, ob das Vorzeichen des ermittelten Effekts spezifikations- und methodenabhängig ist oder nicht, ob sich im Verlauf der Pandemie unterschiedliche Einflüsse identifizieren lassen. Aufgrund unzureichender Datenlage kann es nicht vorrangig darum gehen, kausale Zusammenhänge aufzudecken, obwohl dies wünschenswert wäre. Vielmehr können nur regionale Strukturmuster identifiziert werden, die in einem engen statistischen Zusammenhang mit der Zahl der Infektionen stehen. Der deskriptive Charakter des Beitrags soll damit

hervorgehoben werden. Besser wäre eine empirische Analyse auf Kreisebene oder noch besser für eine große Personen- und/oder Betriebsstichprobe. Selbst auf Bundesländerebene sind nicht alle benötigten Informationen verfügbar. Und ein Problem auf dieser Ebene ist ganz offensichtlich. Nämlich die Heterogenität innerhalb der Bundesländer lässt sich damit nicht abbilden. Unterschiedliche Infektionsmuster in Teilregionen, die von denen in der Gesamtregion abweichen, bleiben unentdeckt.

Gegenüber bisherigen empirischen Analysen zur regional unterschiedlichen Ausbreitung des Corona-Virus wird hier darüber hinausgehend folgenden Problemen nachgegangen. Erstens wird untersucht, ob Zustandsabhängigkeit zu Beginn einzelner Phasen der Pandemie und Heterogenität aus der Vor-Pandemie-Zeit die regionale Ausbreitung der Infektionen beeinflussen. Zweitens werden clusterrobuste statt klassische gaußsche Standardfehler bei den Schätzungen benutzt, um dem Moulton-Problem zu begegnen, das bei Verwendung wiederholter gleichförmiger oder konstanter Regionaldaten auftaucht (Moulton 1990, Hübler 2014). Drittens werden nicht alle verfügbaren Informationen verwendet, da zwischen ihnen starke statistische Zusammenhänge bestehen, die sich nicht isolieren lassen. Vielmehr werden Informationen verdichtet und so nur wenige beobachtbare oder künstliche, nicht beobachtbare Variablen ausgewählt. Viertens wird untersucht, wie die übrig gebliebenen Bestimmungsgrößen zueinander stehen und welche Bedeutung in der verdichteten Form unbeobachtete im Vergleich zu den beobachteten Einflüssen haben. Fünftens wird der Frage nachgegangen, ob sich ein zeitlich weitgehend konstantes Muster der regionalen Infektionsverteilung abbilden lässt oder ob sich systematische Wellenbewegungen und Strukturbrüche identifizieren lassen.

## **2 Ursachen für eine regional unterschiedliche Ausbreitung des Corona-Virus**

Bei der Suche nach Ursachen für regionale Unterschiede fehlt es an einfachen Erklärungen. Stets spielen verschiedene Faktoren eine Rolle, die je nach Region betrachtet werden müssen. Das RKI teilte auf Anfrage unter anderem mit, dass die Bevölkerungsdichte nicht der einzige Grund sein kann, sondern dass auch die Mobilität und möglicherweise das Schutzverhalten von Bedeutung sind. Man könne die Verteilung aber nicht genauer erklären. Generell seien "solche Erkrankungen nicht zum gleichen Zeitpunkt überall gleich, beispielsweise auch bei der Influenza nicht" (02.12.2020).

Neben den Wirkungen spezieller Merkmale sollte verstärkt untersucht werden, welchen Einfluss Maßnahmen, die der Bekämpfung von COVID-19 dienen, auf die regionale Ausbreitung der Corona-Infektionen haben. Lockdowns, Reisebeschränkungen, Maskenpflicht und Abstandsregeln sind hier zu nennen, d.h. Maßnahmen, die überwiegend für die gesamte Volkswirtschaft gelten. Regional differenzierte Informationen stehen kaum zur Verfügung.

Etwas anders sieht dies bei der Zahl und dem Umfang durchgeführter Impfungen aus. Verwendet wird hier die Bundesländerrangzahl für die Erstimpfungsquote. Für die Beurteilung der Wirksamkeit von Maßnahmen ist es wesentlich, ob die Dynamik der Infektionen in regionaler Betrachtung von den ersten Erfahrungen mit COVID-19 getrieben wird oder ob unbeobachtete Heterogenität den Ausschlag gibt. In der Fachliteratur wird hier von „state dependence“ und „unobserved heterogeneity“ gesprochen (Heckman 1978, 1981).

Ansonsten finden sich in verschiedenen Studien theoretische Überlegungen und empirische Ergebnisse zu möglichen Determinanten für die regionale Ungleichverteilung der Infektionen, ohne dass sich eine klare Theorie herausgebildet hat. Zu nennen sind hier die Beiträge von Akbarpour et al. (2020), Benitez et al. (2020), Brown/Ravallion (2020), Desmet/Wacziarg (2020), Galasso et al. (2020), Goldstein/Lee (2020), Knittel/Ozaltun (2020), Krekel et al. (2020), McLaren (2020), Papageorge et al. (2020), Qui et al. (2020) und Sa (2020). Dabei wird jeweils nur der Einfluss weniger Bestimmungsgründe untersucht oder vage Vermutungen werden geäußert. Bei empirischen Untersuchungen sind die Beschränkungen häufig in Abhängigkeit von den verfügbaren Daten vorgegeben und hängen von der Betrachtungsebene (Individuen, Haushalte, Unternehmen, Regionen, Volkswirtschaften) ab. In den genannten Studien werden folgende Merkmale herangezogen: die individuelle Mobilität, Risikoaversion, prosoziale Motive, Vorerkrankungen, Alter, Geschlecht, Bevölkerungsdichte, Gesundheitszustand, Gini-Koeffizient, Migrantenstatus, Armutsgefährdung. Brown und Ravallion (2020) heben hervor, dass Einkommensarmut und Einkommensungleichheit die Infektionsraten ansteigen lassen. Sie finden, dass starke Effekte von der Rassenzugehörigkeit ausgehen, dass arme Menschen weniger imstande sind, sich gegen Infektionen zu schützen, dass die ältere Bevölkerung und junge Familien eine größere soziale Distanz zu anderen Menschen haben, dass mit zunehmender Bevölkerungsdichte die Infektionsgefahr signifikant steigt. Toya/Skidmore (2021) kommen zu dem Ergebnis, dass Länder mit höherem Einkommen, geringerer Bevölkerungsdichte, älterer Bevölkerung, weniger Krankenhausbetten, mehr Freiheiten und mehr PCR-Tests höhere Corona-Infektionsraten als andere aufweisen. Verkehrsbeschränkungen sind danach mit einer schwächeren Ausbreitung der Pandemie verknüpft. Keine klaren Hinweise finden sich zu den Auswirkungen eines Lockdowns. Desmet/Wacziarg (2020) heben ebenfalls die Bedeutung der Bevölkerungsdichte hervor, zeigen, dass die Infektionen mit der Familiengröße zunehmen, dass die Altersstruktur bedeutsam ist, dass Minderheiten stärker betroffen sind, dass Menschen in Pflegeeinrichtungen höhere Infektionsraten aufweisen als andere. Frauen sehen nach Galasso et al. (2020) COVID-19 eher als ernstes Gesundheitsproblem an als Männer, plädieren für restriktivere Maßnahmen zur Bekämpfung der Pandemie und schützen sich auch selbst mehr gegen Ansteckung. Die Bedeutung des Wetters für die Ausbreitung von Corona-Infektionen untersuchen Qui et al. (2020). In diesem Zusammenhang betonen Burdett et al. (2021), dass Temperatur, Sonne und Regen während des ersten Lockdowns in Großbritannien Einfluss auf Outdoor-Aktivitäten hatten und insofern

können daraus Auswirkungen auf die Corona-Ansteckung folgen. Klima- und Umweltvariablen sowie Gesundheitsindikatoren beziehen Knittel/Ozaltun (2020) in ihre Analyse über Todesrisiken bei Corona-Infektionen ein. Sie finden, dass Personen, die nicht arbeiten und daher nicht pendeln, stärker von ernstesten COVID-19-Erkrankungen und Todesfällen betroffen sind. Papageorge et al. (2020) arbeiten regionale Unterschiede heraus. Menschen in Regionen mit hohem Durchschnittseinkommen, mit hohen Sommer- und niedrigen Wintertemperaturen weisen ein höheres COVID-19-Todesrisiko auf. Die Autoren können keinen Zusammenhang mit Übergewicht, der Anzahl der ICU-Betten oder den Armutsraten feststellen. Regional werden aber ganz unterschiedliche Todesraten ausgewiesen. McLaren (2020) findet, dass Minderheiten stärker von COVID-19 betroffen sind. Für Hispanics und Menschen asiatischer Herkunft sind die sonst gefundenen Zusammenhänge fragil und verschwinden, wenn der Bildungsgrad, der Beruf und das Pendlerverhalten als Kontrollvariable berücksichtigt werden. Für Weiße und Schwarze scheint der robuste Zusammenhang nicht tangiert zu werden, ob das Einkommen, die Armutsrate, die Schulbildung und der Beruf in die empirische Untersuchung eingehen. Auch der Zugang zu einer Krankenversicherung ist für diese Gruppe unerheblich in Bezug auf die Infektionsgefahr wohl aber, ob der öffentliche Verkehr genutzt wird.

Um den Analyserahmen klarer abstecken zu können, erscheint es sinnvoll, zunächst aus inhaltlicher Sicht zwischen verschiedenen Bereichen, zwischen Strukturen und Indikatoren, zu trennen, um Merkmale zusammenzufassen, die bisher in der Literatur diskutiert wurden. Anschließend werden diese gemeinsam betrachtet. Die Wirkungsrichtung ist a priori keineswegs immer eindeutig. Folgende Blöcke bieten hier Ansatzpunkte:

- (1) Geographische Strukturen, z.B. gemessen durch die Bevölkerungsdichte, durch Unterscheidung zwischen Flächen- und Stadtstaat, durch nach Himmelsrichtungen getrennte Regionen und nach Angrenzung an andere Länder;
- (2) demographische Strukturen, z.B. Frauenanteil, Anteil jüngerer und Anteil älterer Menschen, Ausländeranteil;
- (3) bildungsökonomische Indikatoren, z.B. Anteil der Bevölkerung ohne Schulabschluss, Anteil der Bevölkerung mit Universitätsabschluss, Betreuungsquote kleiner Kinder;
- (4) klimatische Einflüsse, z.B. Niederschlag, Sonnenstunden, Lufttemperatur;
- (5) ökonomische Indikatoren, z.B. Einkommen, Arbeitslosenquote, Anteil armutsgefährdeter Haushalte;
- (6) Wählerverhalten, gemessen durch den Wähleranteil für einzelne Parteien;

(7) Politikentscheidungen, z.B. gemessen durch die Zahl der Asylbewerber, durch die Zahl der Abschiebungen, durch die Aufklärungsquote von Straftaten, durch die Anzahl der Gefangenen und den Umfang der Staatsverschuldung;

(8) gesundheitsökonomische Einflüsse, z.B. gemessen durch die durchschnittliche Lebenserwartung, die Sterbewahrscheinlichkeit, den Gesundheitszustand;

(9) persönlichkeitsbedingtes Verhalten, erfasst durch die BIG5-Variablen, durch den Grad der Zufriedenheit, durch die Risikobereitschaft, das Selbstbewusstsein, die Neigung zu Optimismus, die Impulsivität, die Geduld und die Bindung an eine Region.

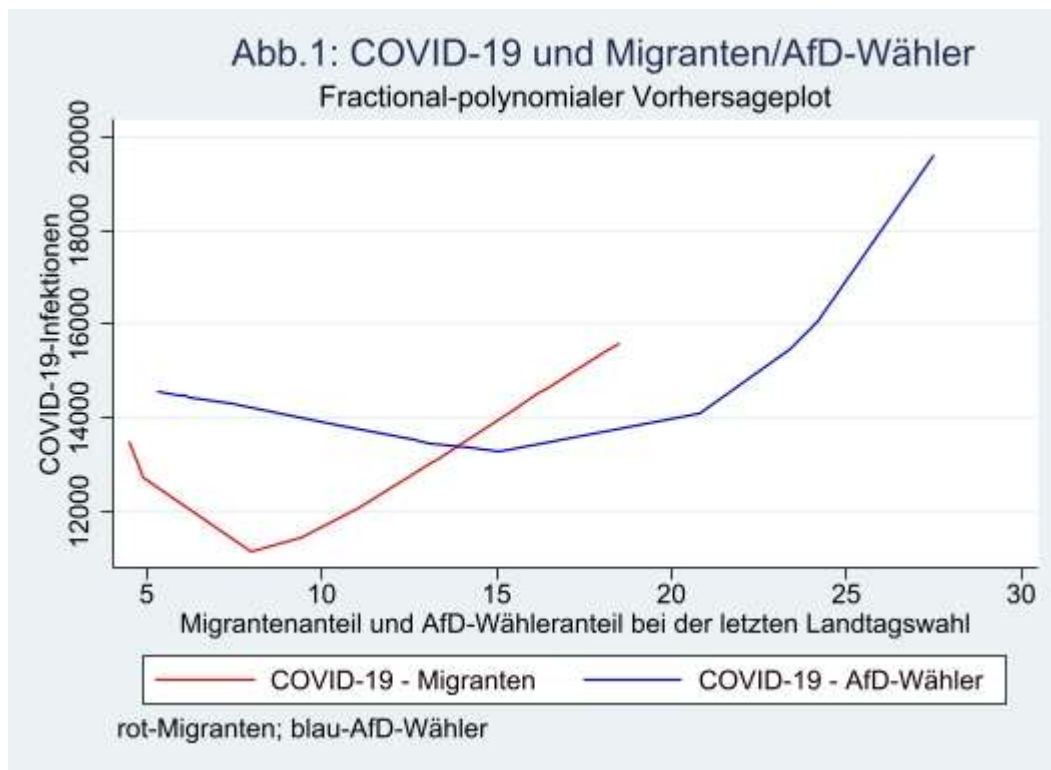
Persönlichkeitsbedingtes Verhalten wurde im Zusammenhang mit Corona nur wenig berücksichtigt. Die Zuordnung zu den verschiedenen Kategorien ist nicht immer eindeutig. So kann der Umfang der Staatsverschuldung sowohl dem Block ökonomische Indikatoren als auch dem Block Politikentscheidungen zugeordnet werden. Zwischen den genannten Variablen gibt es Überschneidungen. Interaktionen in Bezug auf die Wirkungen sind zu erwarten. Es geht zunächst darum, eine größere Anzahl möglicher Einflüsse zu erfassen, um dann mit Hilfe statistisch-ökonometrischer Verfahren, wie durch Hauptkomponentenanalysen, Panelschätzer und Machine Learning-Ansätze, wichtige Einflussgrößen von weniger wichtigen, beobachtbare von unbeobachteten Einflussgrößen zu trennen. Hierauf aufbauend kann dann analysiert werden, ob die so gewählte Spezifikation sensibel auf Veränderungen reagiert, ob es Strukturbrüche gibt, ob und gegebenenfalls welche Unterschiede sich zwischen der ersten, zweiten und dritten Welle ausmachen lassen.

### **3 Datenbasis**

Die folgende empirische Analyse greift vorrangig auf amtliche Erhebungen des statistischen Bundesamtes zurück. Um umgekehrte Kausalität möglichst auszuschließen, wird bei den Regionaldaten auf Angaben zurückgegriffen, die zeitlich vor dem Ausbruch des Corona-Virus erhoben worden sind. Analyseebene sind die Bundesländer. Einerseits werden direkt Bundesländerdaten herangezogen, insbesondere die täglichen COVID-19-Infektionsfälle, die das Robert-Koch-Institut veröffentlicht. Beobachtungszeitraum ist März 2020 bis Anfang Mai 2021. Regionale Strukturdaten des Statistischen Bundesamtes sind nur einmal jährlich erfasst. Sie bilden den wesentlichen Rahmen für die Determinanten der Corona-Infektionen. Andererseits werden Individualdaten aus dem Sozio-ökonomischen Panel (SOEP) zu Regionaldaten aggregiert – vgl. Tabelle 1 und 2.

## 4 Regionale Merkmale: Korrelationskoeffizienten und einfache Regressionsmodelle

Als erste Analyseebene dienen einfache Korrelationskoeffizienten zwischen der Gesamtinfektionszahl und den genannten Variablen bei den Blöcken (1)-(9). Nahezu alle Zusammenhänge sind signifikant – vgl. Tabelle 2, Spalte r. Das gilt gleichermaßen für die absolute Zahl der Infektionen (INFI), für die auf die Bevölkerungszahl bezogene Größe (INFI\_BEV) sowie für die logarithmische Messung (lnINFI, lnINFI\_BEV). Für einzelne Variablen ergeben sich jedoch unterschiedliche Vorzeichen für die Korrelationskoeffizienten, je nachdem ob INFI\_BEV oder lnINFI\_BEV herangezogen wird. So z.B. bei dem Merkmal AfD-Anhänger. Im Folgenden wird mit der logarithmischen Messung gearbeitet, ein Vorgehen, das sich bei sehr vielen ökonometrischen Modellen findet. Likelihood-Ratio-Tests zeigen für Box-Cox-Modelle allerdings weder eine eindeutige Überlegenheit des logarithmischen Ansatz noch des linearen Ansatzes. Fractional-polynomiale Vorhersageplots legen jedoch einen nichtlinearen Ansatz nahe, wie anhand Abb.1 für die Migranten und AfD demonstriert wird.



In Bundesländern mit einem Migrantenanteil bis 8% nehmen die Infektionen ab, je höher in dieser Gruppe der Migrantenanteil. Danach in der überwiegenden Anzahl der Bundesländer



nehmen die Infektionen, bezogen auf die Bevölkerungszahl, mit zunehmendem Migrantenanteil zu. Abnehmende Integration, verbunden mit unzureichenden Kenntnissen der deutschen Sprache und daraus folgend unzureichende Einhaltung von Schutzregeln gegen Infektionen, können ein Erklärungsgrund sein. Ein entsprechendes Muster zeigt die Abbildung für AfD-Wähler. Ist der AfD-Wähleranteil in einem Bundesland vergleichsweise gering, so halten sich deren Anhänger in der Öffentlichkeit in Äußerungen und Verhalten eher zurück, sind überangepasst. Finden sie dagegen viele Sympathisanten, so können sie offen auftreten und zu Verschwörungstheorien in Bezug auf COVID-19 stehen. Sie tun dies, weil sie davon überzeugt sind oder weil sie aus Gruppendruck mitmachen.

Aus geographischer Sicht (Block 1) zeigt sich, dass die Bevölkerung in den norddeutschen und ostdeutschen Ländern weniger anfällig ist als im Rest. Der Westen ist stärker betroffen als der Osten. Bundesländer, die eine gemeinsame Grenze mit dem Ausland haben, weisen keine signifikant höheren Infektionszahlen auf. Um es noch ein letztes Mal zu betonen. Dabei handelt es sich nicht um kausale Aussagen, sondern lediglich deskriptive Ergebnisse, die sich auf einen vorgegebenen Zeitraum beziehen. Ein ganzes Bündel an Ursachen kann dafür verantwortlich sein. Mentalitätsunterschiede wie Lebenslust, Zufriedenheit, Einstellung zur Arbeit und zur Familie, aber auch sich historisch entwickelte Unterschiede aufgrund der Religionszugehörigkeit können bedeutsam sein. Prinzipiell ändert sich an den Aussagen nichts, wenn blockweise getrennt, Regressionen geschätzt werden (Tabelle 3). Regressand ist wie bei allen hier durchgeführten Regressionen die logarithmierte Gesamtinfektionszahl pro Einwohner ( $\ln \text{INFI\_BEV}$ ).

Bei den demographischen Einflüssen (Block (2)) zeigen sich auf Basis der Korrelationskoeffizienten statistisch gesicherte Zusammenhänge zwischen dem Ausländeranteil, dem Alter der Personen und der Lebenserwartung auf der einen Seite und  $\ln \text{INFI\_BEV}$  auf der anderen Seite. Danach sind jüngere Personen (mindestens 15 Jahre alt, aber jünger als 25) stärker von COVID-19 betroffen als ältere Personen (65 Jahre oder älter). Je höher die durchschnittliche Lebenserwartung der Bevölkerung ist, umso mehr Menschen stecken sich an. Bei den Migranten besteht eine größere Infektionsgefahr als bei der einheimischen Bevölkerung. Zusammengefasst in einer Regressionsschätzung folgen zum Teil veränderte Ergebnisse. Mit zunehmender Haushaltsgröße werden statistisch signifikant mehr COVID-19-Fälle beobachtet. Der Zusammenhang bei jüngeren und älteren Personen kehrt sich genau um. Für Ältere ist das Ergebnis jedoch statistisch nicht gesichert. Das gilt auch für den Frauenanteil pro Bundesland. Je höher der Frauenanteil in einem Bundesland ist, eine umso größere Zahl an Infektionen pro Einwohner wird ausgewiesen. In den Medien wird üblicherweise berichtet (Berliner Zeitung 02.11.2020), dass sich Männer und Frauen gleichviel infizieren, dass die Sterblichkeit an COVID-19 bei Männern höher liegt (RKI 08.12.2020). Anzumerken bleibt, dass die Ergebnisse bei den Regressionen weniger signifikant sind, wenn in Bezug auf die Bundesländer cluster-robuste Standardfehler zugrunde gelegt sind ( $t_{\text{rob}}$ ). Das gilt im Übrigen auch für die meisten anderen

Regressionen. Mit diesen Standardfehlern zu arbeiten ist zu empfehlen, da die zugrunde gelegten Regressoren für den betrachteten Zeitraum im Wesentlichen zeitinvariant sind, weil nur Jahreswerte auf Bundesländerebene zur Verfügung stehen. Deshalb wird die Varianz zu niedrig ausgewiesen, ein Phänomen, auf das Moulton (1990) aufmerksam gemacht hat. Dem kann mit clusterrobusten Schätzungen begegnet werden.

Bildungsökonomisch betrachtet (Block 3), lässt sich sagen, dass bei hoher Betreuungsquote der unter Dreijährigen die Ansteckungsgefahr vergleichsweise niedrig ist, während in Bundesländern mit einem hohen Anteil an Personen mit Hochschulreife, bezogen auf ihre Altersgruppe, COVID-19 häufiger registriert wird. Zu Personen ohne Schulabschluss lässt sich auf Basis der Korrelationskoeffizienten in dieser Hinsicht nichts sagen. Wird wieder zum Vergleich die Regressionsschätzung herangezogen, so ergeben sich keine bemerkenswerten Unterschiede zu den einfachen Korrelationsergebnissen, außer dass der Zusammenhang mit Personen ohne Schulabschluss nur dann als statistisch abgesichert ausgewiesen wird, wenn übliche Standardfehler und keine clusterrobusten Standardfehler zugrunde liegen.

Klimatische Faktoren (Block 4) sind durchaus bedeutsam für die Ansteckungsgefahr auf Bundesländerebene. Auf der Ebene der Korrelationsergebnisse zeigt sich, dass niedriger Niederschlag die Infektionsgefahr verringert. Für die jährlichen Sonnenstunden und die durchschnittliche Jahrestemperatur je Bundesland lässt sich keine verlässliche Aussage treffen. Zu diesem Ergebnis kommt man auch bei Regressionsschätzungen mit allen verfügbaren Klimaindikatoren, soweit wiederum auf clusterrobuste Schätzungen Bezug genommen wird. Etwas anders sieht das nach der Regressionsschätzung mit üblichen Standardfehlern aus. Danach sind hohe Niederschlagsmengen pro Jahr, hohe Temperaturen und viel Sonnenschein mit signifikant mehr Infektionen verknüpft. Hohe Temperaturen und viel Sonnenschein bedeuten viele Außenkontakte, und daraus resultiert eine zunehmende Ansteckungsgefahr. Zu beachten ist jedoch, dass es zu weniger Kontakten auf engerem Raum kommt. Welcher Effekt sich per Saldo ergibt, bedarf einer genaueren Untersuchung. Soweit Niederschlag Schneefall bedeutet und damit auf dieser Ebene mehr Zusammenkünfte auch über Après-Ski folgen, besteht die Gefahr von mehr Infektionen, wie der Fall Ischgl gezeigt hat. Bei Regen ist ein weniger klarer Effekt zu erwarten.

Als ökonomische Indikatoren (Block 5) werden das Bruttoinlandsprodukt (BIP), die Arbeitslosenquote (ALQ), die Jugendarbeitslosenquote (J\_ALQ) und die Quote der Haushalte, die armutsgefährdet sind, herangezogen. A priori wäre einerseits zu erwarten, dass bei höherem (Durchschnitts-)Einkommen die besseren Möglichkeiten zum Schutz gegen Corona-Infektionen bestehen. Andererseits resultiert aus einem steigenden Wohlstand beruflich und privat eine zunehmende Mobilität, aus der eine steigende Gefahr der Ansteckung folgt. Umgekehrt sollte die Verbindung zwischen Arbeitslosenquote und COVID-19-Infektionen aussehen. Wer arbeitslos ist, hat weniger Geld als Beschäftigte zur Verfügung, um sich gegen

die Infektion zu schützen, wird jedoch auch weniger mobil sein, weil beruflich kaum noch Anlass dazu besteht und weil Mobilität mit Kosten verbunden ist, die man versucht zu vermeiden, außer bei der Suche nach einer neuen Beschäftigung. Möglicherweise verhalten sich Jugendliche anders als Familienväter. Beim einfachen Vergleich der Korrelationskoeffizienten zeigt sich dies jedoch nicht. Die Vorzeichen der Korrelationskoeffizienten legen nahe, dass es bei steigendem BIP, sinkender ALQ und sinkender J\_ALQ zu einer höheren COVID-19-Ausbreitung kommt. Die Wirkung einer Armutgefährdung könnte ähnlich aussehen wie die bei Arbeitslosigkeit. Die ermittelten Korrelationskoeffizienten sind im Einklang mit dieser Hypothese. Bei gleichzeitiger Berücksichtigung der vier ökonomischen Variablen in einer Regressionsschätzung verändert sich das Vorzeichen von ALQ. Es ist also positiv. Ein schwach signifikanter Zusammenhang folgt nur zwischen lnINFI\_BEV und BIP, soweit mit clusterrobuster Standardabweichung gearbeitet wird.

Um eine mögliche Verbindung zwischen Wahlverhalten (Block 6) und COVID-19-Infektionen aufzudecken, wird der Wähleranteil von CDU, SPD, GRÜNE, LINKE, FDP und AfD bei der letzten Landtagswahl vor 2020 benutzt. Danach weisen aufgrund der Korrelationskoeffizienten die Bundesländer mit einem hohen CDU-Anteil, einem hohen Anteil für die Grünen und für die FDP eine stärkere Ausbreitung von Corona-Infektion als in anderen Bundesländern auf. Oder anders ausgedrückt, dort, wo die Bevölkerung stärker zur SPD, zur Linken oder zur AfD neigt, offenbart sich ein vergleichsweise geringes Infektionsgeschehen, wobei die Verbindung mit dem SPD-Wähleranteil statistisch nicht gesichert ist. Zu beachten ist, dass daraus nicht geschlossen werden kann, dass die Bevölkerung bei Wahlen zu den Regierungsparteien tendiert, da sie von ihnen mehr Anstrengungen zur Bekämpfung der Pandemie erwarten. In der empirischen Analyse wird das Wählerverhalten vor dem Aufkommen von Corona abgebildet. Natürlich ist auf dieser Analyseebene auch der umgekehrte Schluss unzulässig, zumal die entsprechende Regressionsschätzung mit clusterrobusten Standardabweichungen nur zu insignifikanten Ergebnissen führt. Die Vorzeichen bei der Regressionsanalyse gegenüber denen der Korrelationskoeffizienten bleiben weitgehend erhalten. Nur in Bezug auf hohe FDP-Wähleranteile sowie für die LINKE zeigt sich ein Vorzeichenwechsel. Traditionelle Parteienzuordnungen nach einem Rechts-Links-Muster sind hier nicht angebracht. Auch eine Einteilung in Parteien der Mitte und an den Rändern führt nicht weiter.

Politikentscheidungen (Block 7), die zumindest direkt nichts mit Corona zu tun haben, stehen trotz allem auf deskriptiver Ebene in einem bemerkenswerten Zusammenhang mit der gegenwärtigen Pandemie. Allein auf das Vorzeichen der Korrelation abstellend, wird der Blick darauf gelenkt, dass die Pandemie dort stärker verbreitet ist, wo die Staatsverschuldung (VSCHULD) hoch ist, wo viele Asylanträge gestellt werden (ASYL), wo Abschiebep Praxis (ABSCHIEB) verbreitet ist, wo polizeilich viele Straftaten erfasst sind (STRAF), wo viele Personen in Justizvollzugsanstalten sitzen (GEFANG) und wo eine geringe Aufklärungsquote an Straftaten (AUFKLÄR) festzustellen ist. Bei den Regressionsschätzungen ergeben sich mehrere Vorzeichenwechsel, d.h. bei ASYL, ABSCHIEB, STRAF und VSCHULD. Dies spricht für

Scheinkorrelationen. Bei Schätzungen mit clusterrobusten Standardabweichungen zeigen sich nur für GEFANG signifikante Verknüpfungen. Je mehr Personen danach in einem Bundesland in Justizvollzugsanstalten sitzen, umso stärker ist danach die Region von der Pandemie betroffen. Eine direkte Erklärung hierfür lässt sich nicht liefern. Würde nicht InINFI\_BEV, sondern nur die absolute Zahl der Infektionen zugrunde gelegt, wäre die statistische Verknüpfung klar. In einem Bundesland mit einer hohen Bevölkerungszahl sollte sowohl die Zahl der Infektionen als auch die Zahl der Strafgefangenen größer sein als in anderen Bundesländern.

Variablen, die explizit oder implizit den Gesundheitszustand zum Ausdruck bringen (Block 8), sind bei einer Zweivariablenbetrachtung wie folgt signifikant mit der Zahl der Corona-Infektionen, bezogen auf die Bevölkerungszahl, verknüpft: Je höher die durchschnittliche Lebenserwartung (LEBERW) ist und je mehr Sterbefälle pro Jahr (STERBE) in einem Bundesland auftreten, umso mehr Infektionen werden beobachtet. In Bundesländern, in denen der durchschnittliche Gesundheitszustand der Bevölkerung eher schlecht ist (KRANK), wird eine höhere Betroffenheit von Corona festgestellt. Das ist durchaus erwartungstreu. Menschen, denen es weniger gut oder sogar schlecht geht, sind einer höheren Ansteckungsgefahr ausgesetzt als andere. Etwas unerwartet ist dagegen das Ergebnis, dass Bundesländer, in denen viele Menschen ihren Gesundheitszustand als gut bis sehr gut bezeichnen, eine hohe Durchseuchung haben. Eine Erklärung könnte sein, dass Personen dieser Gruppe weniger vorsichtig sind, sich weniger in Acht nehmen. Sie glauben, sie sind robust gegenüber Ansteckung und selbst wenn sie sich anstecken, erwarten sie einen milden Verlauf. Dieses Erklärungsmuster ändert sich nicht, wenn die Ergebnisse der entsprechenden multiplen Regressionsschätzung herangezogen werden. Zu beachten ist, dass bei Hinzuziehung clusterrobuster Standardabweichungen nur noch in Bezug auf die Lebenserwartung ein signifikanter Zusammenhang folgt.

Bei den voranstehenden partiellen Erklärungsansätzen wurden schon gelegentlich mögliche unterschiedliche Verhaltensweisen als Begründungen für eine regional variierende Corona-Betroffenheit herangezogen. In dem letzten Erklärungsblock geht es nur um Einstellungen und Verhaltensweisen (Block 9), da gerade bei der zweiten und dritten Corona-Welle häufiger derartige Begründungen ins Feld geführt werden als am Anfang der Pandemie, ohne dass bisher jedoch systematisch diesem Phänomen nachgegangen wurde. Im Zentrum stehen Persönlichkeitsmerkmale, die in der Psychologie ausführlich untersucht worden sind und unter dem Kürzel Big FIVE zusammengefasst werden: Aufgeschlossenheit (offen, ideenreich, neugierig - openness), Gewissenhaftigkeit (gründlich, effizient, zielstrebig - conscientiousness), Geselligkeit (extrovertiert, kommunikativ - extraversion), Kooperationsbereitschaft (rücksichtsvoll, verträglich - agreeableness) und emotionale Labilität (angespannt, leicht nervös, verletzlich - neuroticism).

Die Korrelationskoeffizienten zwischen dem Grad der Ausbreitung und den fünf Persönlichkeitsmerkmalen sind signifikant, und zwar positiv außer in Bezug auf Gewissenhaftigkeit (CONSC) und Verträglichkeit (AGRE). Daraus lassen sich zunächst keine weiteren Schlussfolgerungen ziehen. Offenheit (openness) und emotionale Labilität (neuroticism) werden als bedeutsam für die Gefahr, von einer Corona-Infektion betroffen zu werden, ausgewiesen. Offenheit ist mit vielfältigen Kontakten zu anderen Personen und daraus ergibt sich eine hohe Ansteckungsgefahr. Emotional labile Personen lassen sich schnell aus der Ruhe bringen, machen sich häufig Sorgen, sind angstbesessen und angespannt. Sie werden übervorsichtig sein und Gebote genau einhalten, um Ansteckungen zu vermeiden. Die Empirie auf aggregierter Ebene kommt jedoch zu einem anderen Ergebnis. Neigung zum Neurotizismus ist danach mit einer stärkeren Verbreitung des Corona-Virus verbunden. Ist Neurotizismus in einem Bundesland weiter verbreitet als in anderen Bundesländern, dann wird die Politik verstärkt auf Beschwichtigung setzen, die Gefahren von Corona herunterspielen, um Ängste zu nehmen. Und daraus resultiert möglicherweise bei der Bevölkerung ein sorgloserer Umgang mit der Pandemie. Die größte emotionale Stabilität unter allen Bundesländern weisen Baden-Württemberg und Mecklenburg-Vorpommern auf, während diese Eigenschaft in Bremen, Brandenburg und Berlin am wenigsten anzutreffen ist.

Gesellige Menschen leben von Kontakten mit anderen Menschen und sind darüber einer stärkeren Infektionsgefahr ausgesetzt als andere. Bei emotional labilen Menschen auf der einen Seite und gewissenhaften, besonnenen, von Leistungsstreben geleiteten Menschen auf der anderen Seite lässt sich a priori weniger klar sagen, ob diese Persönlichkeitsmerkmale einen Hinweis auf eine Corona-Gefährdung liefern. Erstere klagen häufig über körperliche Schmerzen und könnten daher auch anfällig sein, sich mit dem Corona-Virus zu infizieren. Ihre Neigung zu erhöhter Reizbarkeit und Traurigkeit könnte jedoch auch dazu führen, dass sie sich gegenüber der Außenwelt abkapseln und somit weniger Angriffsfläche für eine Infektion bieten. Besonnene, von Selbstdisziplin geleitete Menschen werden aufgrund dieser Grundeinstellung Gefahren der Ansteckung aus dem Weg gehen. Sie werden jedoch aufgrund ihrer Kompetenz und ihres Leistungsstrebens von anderen zu Außenkontakten angehalten.

Über Big FIVE hinausgehend finden noch weitere Persönlichkeitsmerkmale Eingang in die empirische Analyse: Zufriedenheit, Optimismus, Geduld, Risikobereitschaft, Impulsivität und Heimatverbundenheit. Bis auf die Optimismus-Variable sind alle anderen positiv korreliert mit dem regionalen Grad an Corona-Infektionen. Hier zeigen sich erneut ambivalente Wirkungsmechanismen, die durch Korrelationskoeffizienten nur unzureichend abzubilden sind.

Bei der multiplen Regressionsschätzung kehrt sich das Vorzeichen der Regressionskoeffizienten für Extraversion (EXTRA), Gewissenhaftigkeit (CONCS), Kooperationsbereitschaft (AGRE) und Risikobereitschaft (RISIKO) gegenüber dem des Korrelationskoeffizienten um. Neigt die Bevölkerung eines Bundeslandes zu Optimismus (OPTIM), Geduld (GEDULD), oder zu

impulsivem Verhalten (IMPULS), dann ist dort eine überproportionale Verbreitung von Corona zu finden. Das scheint auch eher bei geringer Lebenszufriedenheit (LZUF) so zu sein. Bei den clusterrobusten Schätzungen zeigt sich nur für IMPULS und LZUF ein signifikanter Zusammenhang ( $\alpha \leq 0.05$ ).

Ist die Bevölkerung in einem Bundesland weitgehend geduldig, besitzt sie Langmut und Durchhaltevermögen, so sollte man meinen, dass sie sich an getroffene Maßnahmen zur Bekämpfung der Corona-Pandemie auch über einen längeren Zeitraum hält. Oder anders ausgedrückt, Geduld stärkt die Resilienz gegenüber Corona. Ungeduldige Menschen wollen dagegen auch in Corona-Zeiten nicht auf Freizeitaktivitäten verzichten. Sie befürworten zwar allgemein strenge Maßnahmen zur Bekämpfung der Pandemie, halten sich aber selbst kaum daran. Auch hier weist die Empirie in die entgegengesetzte Richtung. Bei geduldiger Bevölkerung wird eine stärkere Ausbreitung von COVID-19 beobachtet. Schätzt die Politik ihre Bevölkerung als überwiegend geduldig ein, so wird sie nicht sofort sehr strenge Maßnahmen erlassen, verbunden mit Sanktionen bei Nichteinhaltung, da sie davon ausgeht, dass auch geringfügige Auflagen werden weitgehend eingehalten. Nun besteht eine Bevölkerung nie vollständig aus geduldigen Menschen und die Ungeduldigen werden sich selbst an schwache Auflagen nicht halten.

Wird nicht nur getrennt für die neun inhaltlich unterschiedlichen Blöcke gesucht, welche der ausgewählten Variablen aus statistischer Sicht bedeutsam für COVID-19-Infektionen sind, sondern wird auch von mehr als einem Block ausgegangen, so zeigen sich zum Teil abweichende Muster. Es macht aber wenig Sinn alle Variablen aus Tabelle 3 gemeinsam zu berücksichtigen. Ein hoher Grad an Multikollinearität mit insignifikanten Einflüssen und/oder unplausiblen Regressionskoeffizienten wäre die Folge. Wird aber aus jedem Erklärungsblock nur eine Determinante herangezogen, und zwar eine die nach Tabelle 3 auf einen starken Einfluss hindeutet, so erhält man ein weitgehend stabiles Erklärungsmuster, wie die Schätzung in Tabelle 3a zeigt.

## **5 Zustandsabhängigkeit, Heterogenität, Tests und Impfung**

Zu untersuchen ist, ob die Dynamik der Infektionen in regionaler Betrachtung von den ersten Erfahrungen mit COVID-19 getrieben wird oder ob Vor-Corona-Heterogenität den Ausschlag gibt. Als Ausgangspunkt (Ausgangsbedingung) für die zweite und dritte Welle wird die Zahl der Todesfälle, jeweils bezogen auf die Bevölkerung in einem Bundesland, angenommen (state dependence). Eine erste Operationalisierung für die regionale Heterogenität soll die bundesländerspezifische Lebenszufriedenheit vor Beginn der Pandemie sein. Angenommen wird eine relative zeitliche Stabilität dieses Merkmals in coronafreien Zeiten. Lebenszufriedenheit

wird dabei als Aggregat, als zusammengefasste individuelle Bewertung verschiedener Situationen und Verhaltensweisen aufgefasst.

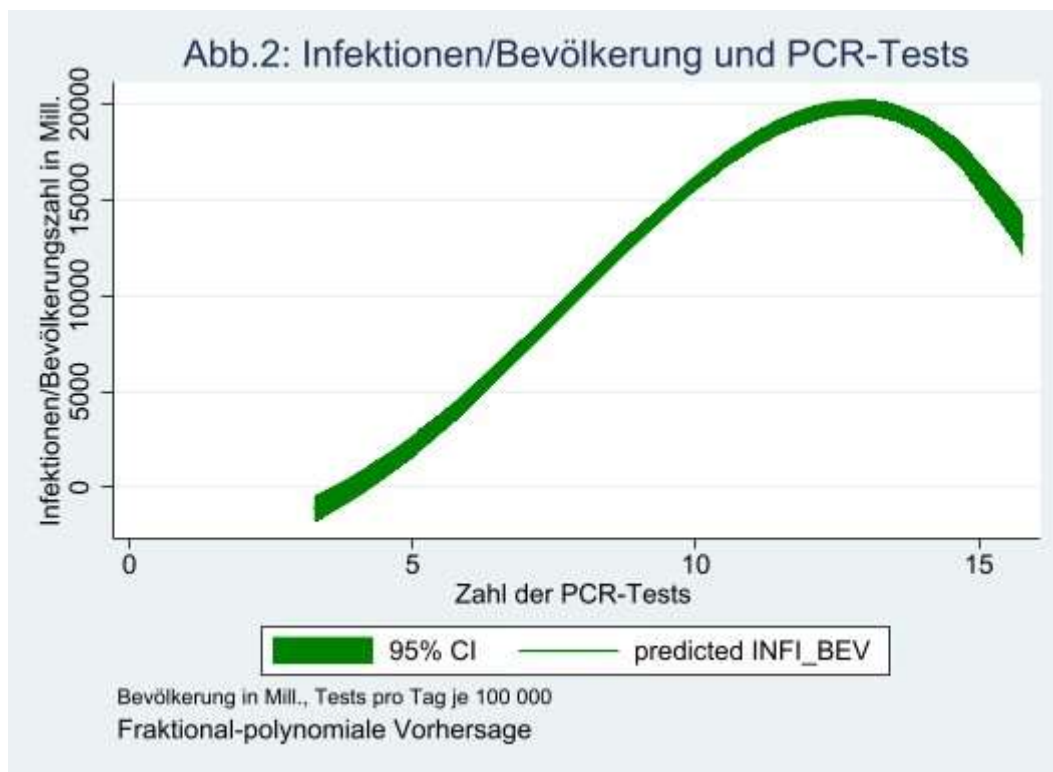
Regressionsschätzungen mit den so beschriebenen Variablen als Regressoren und bei Verwendung clusterrobuster Standardfehler finden sich in Tabelle 4. Festzuhalten bleibt, dass Zustandsabhängigkeit statistisch bedeutsam für die Zahl der Infektionen ist. Für die regionale Heterogenität, ausgedrückt durch die Lebenszufriedenheit, lässt sich dagegen kein Zusammenhang aufzeigen, egal ob die beiden Regressoren zusammen oder in getrennten Modellen in die Schätzung eingehen. Offen bleibt damit, ob tatsächlich keine Verbindung besteht oder ob die Operationalisierung unzureichend ist. In Frage kommen vor allem andere Persönlichkeitsmerkmale, die bereits in Tabelle 3 in getrennter, beschreibender Form Berücksichtigung fanden. Wird in Tabelle 4 der Regressor Lebenszufriedenheit sukzessive durch diese Variablen ersetzt, so zeigt sich für die Variablen AGRE, CONSC, EXTRA, NEURO, und SELFCON ein signifikanter Effekt – nicht ausgewiesen in den Tabellen. Das muss aber nicht heißen, dass diese die besseren Heterogenitätsindikatoren sind. Sie sind weniger umfassend als LZUF.

Neben den genannten Merkmalen sollten, wie bereits erwähnt, auch Maßnahmen, die der Bekämpfung von COVID-19 dienen, in ihren Wirkungen bei der empirischen Analyse untersucht werden. Zur Verfügung stehen einerseits die Zahl der PCR-Testungen (TESTS bzw. die normierte Zahl an Testungen) und andererseits die Bundesländerrangzahl für die Erstimpfungsquote (IMPFUNGEN\_r). Tests dienen der Identifizierung von Infektionen. Daraus folgende Quarantäne soll der Verhinderung einer weiteren Ausbreitung von COVID-19 dienen.

In Tabelle 5 sind die Regressionsschätzungen für lnINFI\_BEV mit den beiden Maßnahmen als Regressoren wiedergegeben. Zu beachten ist, dass Angaben über Tests ab der 19. Kalenderwoche 2020, d.h. ab 06.05.2020 herangezogen werden und zwar die Anzahl der Testungen in einer Woche für Deutschland insgesamt. Impfungen gehen ab Anfang Januar 2021 (01.01.2021) in die Analyse ein. Die Auswahl der weiteren Variablen in Spalte 4 und 5 basiert auf dem RLASSO-Verfahren – vgl. Abschnitt 6. Erst dann wird auf diese Ergebnisse eingegangen.

Es zeigt sich, dass die Zahl der erfassten Infektionen, bezogen auf die Bevölkerung, umso höher ist, je mehr getestet wird. Durch das Testen werden sonst nicht erkannten Corona-Infektionen erst entdeckt. Zu beachten, dass lnINFI\_BEV und TESTS einen gemeinsamen Trend aufweisen.

Aus Abb.2 wird deutlich, dass der Zusammenhang zwischen der Zahl der Tests und den Infektionen nicht durchgängig linear ansteigend verläuft. Wenn sehr viel getestet wird, kehrt sich der Zusammenhang um. Testpersonen nehmen sich möglicherweise bei Mehrfachtests stärker in Acht, um nicht in Quarantäne zu müssen. Die Wahrscheinlichkeit, dass ein positives Testergebnis ausgewiesen wird, steigt mit der Zahl der Tests.

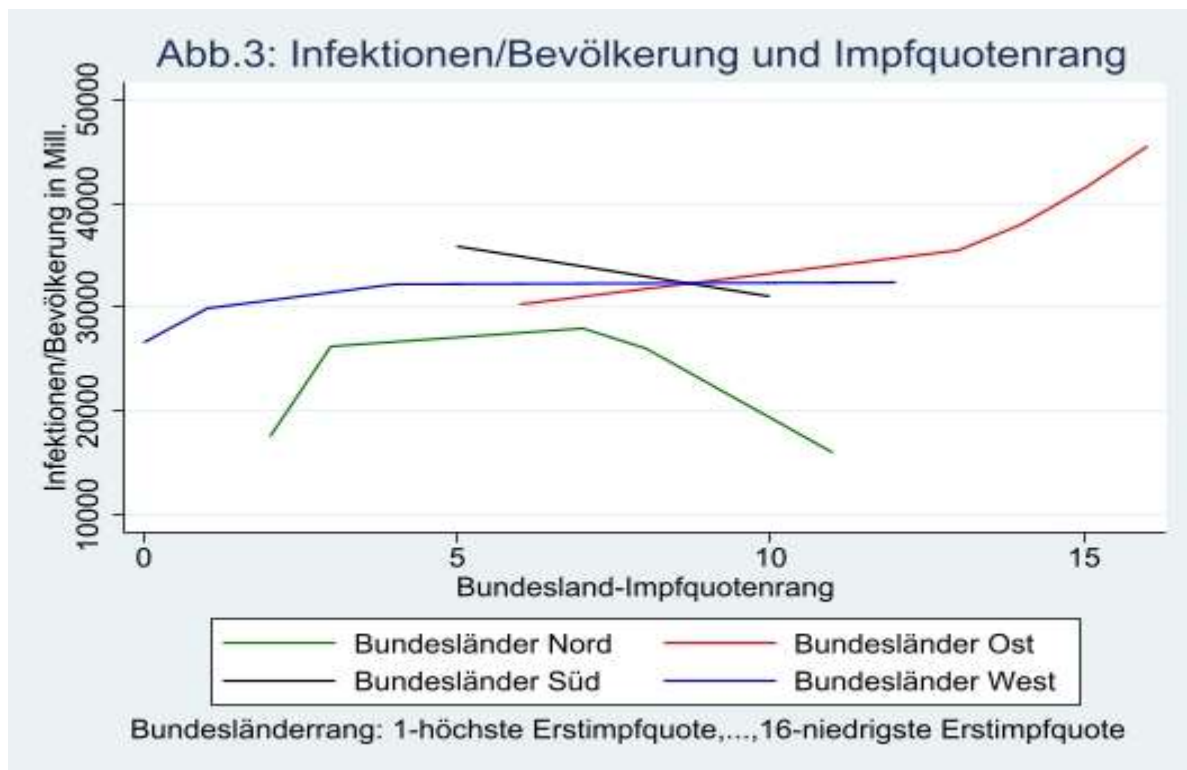


Zwischen den kumulierten COVID-19-Infektionen ( $\ln\text{INFI\_BEV}$ ) und dem regionalen Rang der Erstimpfquoten nach Bundesländern zu einem Zeitpunkt ( $\text{IMPFUNGEN\_r}$ : =1, wenn höchste Impfquote, ..., =16, wenn niedrigste Impfquote) besteht ebenfalls eine positive Beziehung. Je höher der Rang eines Bundeslandes ist, je geringer die relative Impfquote ausfällt, eine desto höhere kumulierte relative Infektionszahl folgt. Dies ist ein Hinweis dafür, dass Impfungen erfolgreich sind. Noch überzeugender ließe sich dies vermutlich demonstrieren, wenn ein längerer Zeitraum seit Beginn der Impfungen und wenn eine verzögerte Impfquote als Regressor in die Schätzung einginge. Es existieren bisher aber keine verlässlichen Untersuchungen darüber, wie lange es dauert, bis der Schutz der Impfungen wirksam wird. In Spalte 3 der Tabelle 5 gehen beide Maßnahmen gleichzeitig in die Schätzungen ein. Signifikanz und Vorzeichen ändern sich gegenüber Spalte 1 und 2 nicht. Welche der beiden Maßnahmen beim Aufdecken von Infektionen oder beim Eindämmen erfolgreicher ist, kann anhand der geschätzten Koeffizienten nicht gesagt werden, da sie skalenabhängig sind. Einen Hinweis liefern jedoch die BETA-Koeffizienten, die sich bei Standardisierung der Variablen ergeben. Die Schätzungen lauten:  $\text{BETA}(\text{TEST})=0.3172$  und  $\text{BETA}(\text{IMPFUNGEN\_r})=0.3628$ . Das Impfen ist



demnach erfolgversprechender, zumal mit dem Aufdecken von Infektionen letztlich noch nicht viel gewonnen ist, außer es erfolgt Quarantäne für die betroffenen Personen.

Abb. 3 zeigt, dass der Zusammenhang über alle Bundesländer, der in Tabelle 5 wiedergegeben ist, bei Aufspaltung in nördliche, östliche, westliche und südliche Bundesländer bisher keineswegs uniform verlaufen ist. So machen sich innerhalb der westlichen Bundesländer Unterschiede beim Impfen kaum bemerkbar bei den auf die Bevölkerungszahl bezogenen Corona-Infektionen. Zwischen den südlichen Bundesländern scheinen entgegen dem allgemeinen Trend hohe Impfquoten keineswegs die Ausbreitung von Corona zu bremsen. Die östlichen Bundesländer folgen dagegen dem allgemeinen Trend.



Zusammenfassend lässt sich bis zu dieser Stelle sagen, dass nicht nur innerhalb der Erklärungsblöcke, sondern auch zwischen den Erklärungsblöcken offensichtlich statistisch Zusammenhänge bestehen, dass die Vorzeichen spezifikationsabhängig sind und dass unbeobachtete Einflüsse den tatsächlichen Wirkungszusammenhang verzerren können. Diesen Problemen ist nachzugehen, um nicht Gefahr zu laufen, Ergebnisse fehlerhaft zu interpretieren.

## 5 Verfahren zur Verdichtung von Informationen und Variablenauswahl

Die zum Teil wechselnden Vorzeichen bei der Zwei- und Mehrvariablenbetrachtung sind ein Hinweis darauf, dass statistische Zusammenhänge innerhalb der einzelnen Blöcke bestehen. Insgesamt bleibt die blockweise Modellierung unbefriedigend. Verbindungen zwischen den Blöcken sind naheliegend. Das ergibt sich bereits aus der vollständigen Korrelationstabelle über alle erfassten Einflussgrößen. Einer einfachen Erweiterung um alle verfügbaren Regressoren sind Grenzen gesetzt. Eine Verdichtung der Information soll Abhilfe schaffen. Verschiedene statistisch-ökonomische Verfahren stehen zur Verfügung, so z.B. multivariate Statistiken oder die in neuerer Zeit an Bedeutung gewonnenen Machine Learning-Verfahren. Hier wird der Hauptkomponentenmethode, dem LARS-Ansatz und einem LASSO-Ansatz gefolgt. Sie dienen im ersten Fall der Bildung aggregierter, künstlicher, nicht direkt erhobener Variablen. Und im zweiten und dritten Fall erfolgt eine Auswahl relevanter Regressoren aus einer vorgegebenen Menge an Bestimmungsgrößen. Herangezogen werden alle Variablen, die bereits in der getrennten Blockbetrachtung Verwendung gefunden haben.

Die Hauptkomponentenanalyse (PCA) dient als statistische Technik der Reduktion der Variablen zur Beschreibung unkorrelierter Linearkombinationen aus allen verfügbaren Variablen. Ziel ist es, Linearkombinationen der Variablen mit der größten Varianz zu finden. Die erste Hauptkomponente weist die maximale Varianz auf. Die zweite hat die größte Varianz unter allen restlichen Linearkombinationen, die unkorreliert sind mit der ersten Hauptkomponente. Analog werden weitere Hauptkomponenten extrahiert. Alle Hauptkomponenten zusammen enthalten dieselbe Information wie die Originalvariablen. Zunächst ist die Anzahl der Komponenten festzulegen. Hier wurde zwischen 3 und 9 Komponenten experimentiert. Die Obergrenze ist mit der Anzahl der inhaltlich unterschiedlichen Blöcke festgelegt. Nach dem Kaiser-Dickman-Kriterium (Kaiser/Dickman 1959) könnten auch 10 Komponenten herangezogen werden, da die ersten 10 Komponenten einen Eigenwert größer 1 aufweisen. Die Untergrenze mit drei Komponenten folgt nach dem Scree-Kriterium - siehe Screeplot (Ost 1984, S.603). Faktoren bleiben danach unberücksichtigt, die nur noch langsam in den Eigenwerten abnehmen. Die ersten drei Faktoren erklären 62% der Varianz, während die restlichen 7 Faktoren nur noch 31% zusätzlich erklären. Auf die Ergebnisse im 10-Komponentenmodell wird nicht weiter eingegangen. Dies liegt an der Schwierigkeit, die 10 Faktoren auch nur ansatzweise substantiell zu interpretieren. Insgesamt lassen sich keine klar inhaltlich interpretierbaren Komponenten extrahieren. Im Dreikomponentenmodell laden auf der ersten Komponente die West-Ost-Dummy, Jugend, Bruttoinlandsprodukt, AfD, Lebenszufriedenheit und regionale Beliebtheit hoch (Tabelle 6), wobei von einer hohen Ladung gesprochen wird, wenn der Faktor-Score größer als 0.25 ist. Es handelt sich danach um ein wohlhabendes altes Bundesland mit jüngerer Bevölkerung und geringerem AfD-Wählerpotential. Das trifft am ehesten auf Hamburg, Schleswig-Holstein und gewissem Abstand auf Bremen zu. Der zweite Faktor ist geprägt von

hoher Bevölkerungsdichte, hoher Durchschnittstemperatur und einer geringen Aufklärungsquote bei Straftaten. Das trifft am besten für die Stadtstaaten Berlin, Bremen und Hamburg zu. Der dritte Faktor weist hohe Faktorladungen für die Einwohnerzahl, die Politikvariablen Asylanträge, Abschiebungen, polizeilich erfasste Straftaten, Gefangene in Justizvollzugsanstalten, Verschuldung und Sterbefälle auf. Werden die drei Komponenten der PCA (PC1, PC2 und PC3), als Regressoren herangezogen und clusterrobuste Standardabweichungen verwendet, so ergibt sich in Tabelle 7, dass lediglich der erste Faktor signifikant ist, den man noch einigermaßen gut inhaltlich interpretieren kann. Der dritte Faktor lässt sich im Grunde genommen auf die Einwohnerzahl reduzieren. Zu überlegen ist bei weiteren Analysen den ersten Faktor und vielleicht auch die Stadtstaaten als Dummy-Variablen und die Bevölkerungsdichte heranzuziehen. Zu erwähnen ist, dass bei allen drei Komponenten Persönlichkeitsmerkmale keine Rolle spielen.

Das sieht anders aus, wenn LARS und RLASSO zur Auswahl relevanter Bestimmungsgrößen für die Infektionszahlen herangezogen werden. Die Least Angle Regression (LARS) wurde von Efron et al. (2004) entwickelt und der Robust Least Absolute Shrinkage and Selection Operator (RLASSO) von Belloni et al. (2012). Letzterer erlaubt eine Schätzung unter Heteroskedastie, nicht gaußschen und geclusterten Störgrößen. Für das LARS-Auswahlverfahren werden als Basis alle Variablen verwendet, die sich bei der getrennten Schätzung der Blöcke (1)-(9) als signifikant erwiesen haben. Der LARS-Algorithmus wählt nur eine Teilmenge der insgesamt zur Verfügung stehenden Variablen aus. Ausgehend von dieser Selektion folgen die in Tabelle 8 präsentierten Schätzungen mit clusterrobusten Standardfehlern. Dass die Ergebnisse sensitiv auf Spezifikationsveränderungen reagieren, zeigt sich zum Beispiel, wenn BEVD oder SELFCON vernachlässigt wird.

Nach den Schätzungen in Tabelle 8 sind lediglich BEVD und LZUF durchgängig signifikant. Sie weisen unterschiedliche Vorzeichen auf. Sofort einleuchtend ist der positive Zusammenhang mit der Bevölkerungsdichte eines Bundeslandes. Je enger die Menschen aufeinander leben, umso mehr Kontakte untereinander sind zu erwarten, umso höher ist die Ansteckungsgefahr. Zu vermuten wäre a priori, dass zufriedene Menschen ebenfalls häufiger Kontakte mit anderen haben und sich deshalb auch häufiger anstecken. Dies bestätigt die Schätzung jedoch nicht und steht im Gegensatz zum positiven Zusammenhang zwischen  $\ln INFI\_BEV$  und LZUF im Zweivariablenmodell. Dort bleiben andere Einflüsse, insbesondere andere Verhaltensmerkmale unberücksichtigt, die mit LZUF hoch korreliert sind. Eine Erklärung für den negativen Zusammenhang könnte sein, dass sich zufriedene Personen der Ansteckungsgefahr besonders bewusst sind und mehr als andere Abwehrmaßnahmen treffen.

Nach RLASSO werden weniger Variablen als bedeutsam eingestuft als bei LARS, wie Tabelle 9 im Vergleich zu Tabelle 8 deutlich macht. Ein wesentliches Problem bei RLASSO ist, welche Variablen als Basis für dieses Verfahren zugrunde zu legen sind. Die Vorauswahl hat wie z. B.

auch bei dynamischen Paneldatenmodellen Einfluss auf das Ergebnis, d.h. hier welche Variablen letztlich in die Auswahl kommen. Dies zeigt Tabelle 9. Dort wird eine Vorauswahl nach sechs verschiedenen Kriterien getroffen:

- (1) alle Variablen aus Tabelle 3;
- (2) nur diejenigen Variablen, die in Tabelle 3 signifikant bei klassischem Standardfehler sind;
- (3) Modifikation von (2), d.h. Vernachlässigung von ARMUT und SCHULE, aber zusätzlich NEURO
- (4) wie (2), aber nur die Variablen, die bei clusterrobustem Standardfehler signifikant sind;
- (5) alle Variablen aus Tabelle 3a
- (6) alle Variablen, die nach LARS ausgewählt wurden.

Als zentrale Bestimmungsgrößen erweisen sich in Tabelle 9 die Variablen nördliche und östliche Bundesländer gegenüber dem Rest, Migrantenanteil, Familiengröße, Schulbildung, Armut, Betreuungsgrad kleiner Kinder und Lebenserwartung. Unter den Persönlichkeitsmerkmalen sind drei der Big Five Charakteristika von Bedeutung – CONSC, NEURO und AGRE. Sowohl Regionalvariablen, die aus der amtlichen Statistik stammen, als auch auf Bundesländerebene aggregierte Persönlichkeitsmerkmale werden als relevant für die ungleiche Ausbreitung von COVID-19 eingestuft.

Wird die RLASSO-Spezifikation in Spalte 2 ergänzt um den Regressor TOTE/EINWZ, so folgt in Tabelle 4, Spalte 4, dass die in den Spalten (1) und (3) als State Dependence-Variable bezeichnete Größe nicht mehr signifikant ist, wenn die als Heterogenitätsindikator bezeichnete Variable LZUF durch die Determinanten des RLASSO-Modells ersetzt werden. Die ursprüngliche Bewertung der Zustandsabhängigkeit im Vergleich zur Heterogenität in Abschnitt 5 erfährt damit eine Umkehrung. Oder anders ausgedrückt, Lebenszufriedenheit deckt Heterogenität weniger gut ab als die Regressoren der RLASSO-Spezifikation und die Corona-Toten als State-Dependence sind unzureichend.

Warum Bundesländer mit einem vergleichsweise höheren Ausländeranteil stärker als andere von der Pandemie betroffen sind, lässt sich mit verschiedenen Argumenten begründen. Vor allem in der ersten Phase der Pandemie haben Mitbürger mit Migrationshintergrund aufgrund von Sprachschwierigkeiten die erlassenen Einschränkungen nicht voll verstanden und sich daher auch nicht daran gehalten. Nach Anlaufschwierigkeiten sind von staatlicher Seite die Erlasse zur Eindämmung der Corona-Infektionen auch in verschiedenen Sprachen bekannt gemacht worden. Es wäre also zu erwarten gewesen, dass aus dem genannten Grund eine rückläufige Tendenz bei der überproportionalen Betroffenheit von Migranten hätte eintreten müssen. Das lässt sich jedoch nicht eindeutig nachweisen - vgl. hierzu Abschnitt 6. Auch während der dritten Welle werden noch Sprachschwierigkeiten verantwortlich dafür gemacht, dass Migranten im stärkeren Maße von Corona betroffen sind als Biodeutsche. Und eine noch stärkere Vermittlung der Verhaltensregeln zur Bekämpfung von Corona in den Heimatsprachen wird gefordert (vgl. Heute-Sendung am 28.04.2021). Eine generell bessere Aufklärung über die Pandemie in den

Heimatsprachen soll auch helfen, schlechter integrierte Zuwanderer davor zu bewahren, Verschwörungstheorien anzuhängen (vgl. Mijnsen, Wien 06.04.2021 in NZZ vom 16.05.2021),

Schlechtere ökonomische Verhältnisse, verbunden mit ungünstigeren Wohnverhältnissen, können jedoch ausschlaggebend dafür sein, dass sich Abstandsregeln bei Migranten weniger gut einhalten lassen. Möglich ist auch, dass sich die Verhaltensweisen von Migranten aus verschiedenen Ländern deutlich von denen der einheimischen Bevölkerung unterscheiden. Intensiveren Kontakten mit anderen Menschen, vor allem mit Personen aus den gleichen Herkunftsländern, aus der gleichen Community, wird möglicherweise ein höherer Stellenwert beigemessen. Wenn der Migrantenanteil gleicher ursprünglicher Nationalität in einem Bundesland besonders hoch ist, dann ist die Neigung zur Integration, zur Aufgabe von Gewohnheiten aus dem Ursprungsland geringer als anderswo.

Wenn die genannten Argumente im Zusammenhang mit Migranten tragfähig sind und Migranten tendenziell geringere Schulbildung besitzen und ärmer sind, dann wäre zu erwarten, dass geringe Schulbildung und Armut eher mit einer höheren Gefahr der Ansteckung verbunden sind. Die Ergebnisse weisen jedoch nicht in diese Richtung, sondern liefern einen starken Hinweis für das Gegenteil. Eine Erklärung hierfür könnte sein, dass Personen mit besserer Schulbildung mehr als andere beruflich und privat mobil sind, mehr Kontakte als diejenigen ohne Schulabschluss aufweisen, also insofern auch einer höheren Infektionsgefahr ausgesetzt sind. Hinzukommt, dass ältere Menschen eher als jüngere keinen Schulabschluss besitzen und unabhängig davon ihre Mobilität einschränken.

In ähnliche Richtung lässt sich für ärmere Personen argumentieren. Sie können sich z. B. Auslandsreisen, bei denen sie auf Corona-Hotspots treffen, weniger leisten. Für Migranten trifft das weniger zu, auch wenn sie ökonomisch schlecht gestellt sind. Die meist verbleibende Bindung an ihre Heimatländer veranlasst sie trotz geringer Mittel, Reisen in ihre Heimat und zu den dort verbliebenen Verwandten und Bekannten zu unternehmen. Die Aussage von Jens Spahn (vgl. Tagesschau.de 25.05.2021), Verwandtschaftsbesuche in der Türkei und auf dem Balkan hätten 2020 phasenweise jede zweite Neuinfektion ausgelöst, hat scharfe Kritik von türkischer und albanischer Seite hervorgerufen.

Bundesländer mit einem relativ hohen Bevölkerungsanteil ohne Schulbildung, das trifft eher auf ländliche Regionen mit noch stärkerer Heimatverbundenheit zu als auf städtische Regionen, weisen eine vergleichsweise geringe Mobilität auf.

Von den erfassten Persönlichkeitsmerkmalen weist das RLASSO-Auswahlverfahren die Gewissenhaftigkeit (CONSC) als wichtiges Merkmal aus. Wenn bei einer Person das Merkmal Gewissenhaftigkeit schwach ausgeprägt ist, womit Unbekümmertheit und Nachlässigkeit zum Ausdruck kommen, sollte eine hohe Ansteckungsgefahr verknüpft sein, weil weniger auf die Einhaltung von Maßnahmen zur Vermeidung von Infektionen geachtet wird. Die Empirie spricht

für das Gegenteil. Dies Ergebnis lässt sich mit der häufig von Persönlichkeitsforschern vertretenen Meinung vereinbaren, dass Gewissenhaftigkeit gleichzeitig das Produkt einer motivierenden und einer disziplinierenden psychischen Kraft ist. Motivierend, weil solche Menschen sich ganz auf eine Arbeit konzentrieren können, und disziplinierend, weil sie lockende Ablenkungen asketisch abblocken können (Bonelli 2014, S. 62ff). Die Gefahren von Corona-Infektionen werden von solchen Menschen einfach ignoriert.

Bei den auf RLASSO basierenden Spezifikationen und den daraus folgenden Schätzungen sollte geprüft werden, inwieweit Verhaltenseigenschaften Einfluss auf die Nichtverhaltenseinflüsse nehmen. Vermutet wird z.B., dass der Einfluss von OPEN und GEDULD vernachlässigt werden kann, da diese Verhaltensvariablen im unstringierten Modell mit MIGRANTEN, SCHULE und ARMUT als weitere Regressoren insignifikant sind. Der statistische Einfluss insgesamt lässt sich schnell mit einem einfachen F-Test prüfen, d.h. die Residuenquadratsumme des restringierten Modells ohne Verhaltensvariablen wird mit der des erweiterten, relativ unrestringierten Modells verglichen, jeweils gewichtet mit den Freiheitsgraden. Es zeigt sich bei  $F=9.87 > F(3; 5)$  ein nicht zu vernachlässigender Einfluss. Unklar ist damit jedoch noch, inwieweit die einzelnen Regressoren des restringierten Modells betroffen sind. Eine Überprüfung hierzu ist mit Hilfe des Gelbach-Ansatzes möglich (Gelbach 2016). Hierfür ergibt sich Tabelle 10, oberer Teil.

Die Zerlegung der Veränderungen der Koeffizienten der Nichtverhaltensvariablen MIGRANTEN, SCHULE und ARMUT durch Berücksichtigung der Variablen OPEN und GEDULD macht deutlich, dass weder der Einfluss der Ausländervariablen noch der Schul- und Armutsvariablen auf  $\ln INFI\_BEV$  durch die beiden Verhaltensvariablen berührt wird, wie Tabelle 10 zu entnehmen ist. Anders sieht es aus, wenn statt OPEN und GEDULD die Bedeutung der Verhaltensvariablen CONSC und NEURO analysiert wird. Hier zeigt sich, dass der Einfluss der Migrantenvariablen auf  $\ln INFI\_BEV$  überschätzt wird, wenn CONSC und NEURO unberücksichtigt bleiben. Letztere üben einen gemeinsamen signifikanten Einfluss auf den Regressor MIGRANTEN aus. Der Koeffizient für MIGRANTEN sinkt beim Übergang vom restringierten zum unrestringierten Modell von 0.0594 auf 0.0494. Der absolute Veränderungseffekt beläuft sich also auf 0.01. In Bezug auf SCHULE und ARMUT lässt sich nichts dergleichen feststellen (vgl. Tabelle 10, unterer Teil).

Im nächsten Schritt wird geprüft, ob die regionalen Unterschiede in erster Linie auf beobachtete Determinanten oder vorrangig auf unbeobachtete Einflüsse zurückzuführen sind. Dies lässt sich mit Hilfe der aus der Diskriminierungstheorie bekannten Blinder-Zerlegung prüfen (Blinder 1973). Nach Tabelle 11 werden bei der gewählten Spezifikation nur 33,8% der Corona-Infektionsunterschiede zwischen nördlichen/östlichen Bundesländern auf der einen Seite ( $NO\_SW==1$ ) und südlichen/westlichen Bundesländern auf der anderen Seite ( $NO\_SW==0$ ) durch das Modell erklärt. Der große Rest (66,2%) ist auf nicht erklärte Variation zurückzuführen.

Ein Spezifikationsfehler bei der Informationsverdichtung durch LARS oder LASSO kann dadurch zustande kommen, dass unbeobachtete, regionale Heterogenität unberücksichtigt bleibt. Ob daraus weniger zuverlässige Unterschiede resultieren, soll durch Paneldatenanalysen geprüft werden. Fixed Effects-Schätzungen sind hier jedoch nicht angebracht, da aufgrund der Datenlage zeitinvariante Regressoren speziell bei Persönlichkeitsmerkmalen zugrunde gelegt sind, die als Regressoren unberücksichtigt bleiben und sich nicht von den Individualeffekten trennen lassen. Random Effects-Schätzungen unterscheiden sich aufgrund der Messung der Regressoren nicht von den gepoolten Schätzungen.

Es können aber auch Regionaleffekte (Bundesländereffekte) auftreten, die sich durch regionale Dummy-Variablen erfassen lassen. Schätzungen hierzu finden sich in Tabelle 12, Spalte (1) und (2). Zwischen der der Schätzung von Spalte (2) in Tabelle 9 und Spalte (1) in Tabelle 12 zeigen sich keine gravierenden Unterschiede.

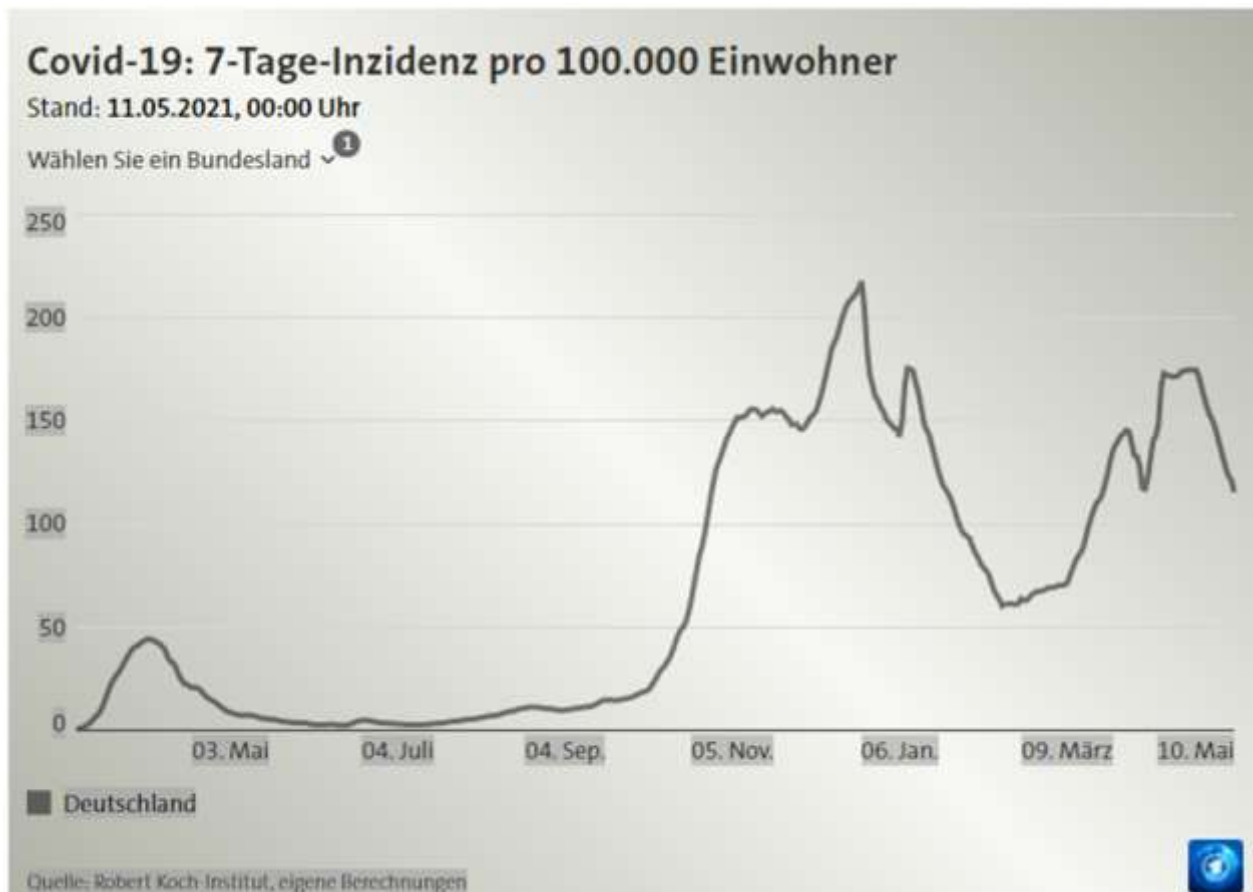
Eine Alternative bieten Between Effects-Modelle. Die Bedingungen hierfür sind nur partiell erfüllt. Ein Großteil der Regressoren bildet zwar regionale Durchschnittswerte ab. Für den Regressanden werden jedoch tagesspezifische Bundesländerwerte verwendet. Ergebnisse hierzu finden sich in Tabelle 12, Spalte (3) und (4). Jackknife Standardfehler liegen diesen Schätzungen zugrunde. Der zentrale Unterschied gegenüber den Random Effects-Schätzungen ist die weitgehend fehlende Signifikanz der Regressoren.

## **6 Veränderungen im Verlauf der Pandemie**

Im nächsten Schritt wird untersucht, ob die bisher berichteten Ergebnisse zeitinvariant sind oder ob Strukturbrüche vorliegen. Hierzu werden zunächst, monatsweise getrennt (März 2020 bis April 2021) für ein Fünftageintervall in der Mitte des Monats, die RLASSO-Spezifikationen ermittelt (Tabelle 13). Es zeigen sich auf dieser Ebene zwar Veränderungen, aber keine klaren Strukturbrüche. Erwähnt werden sollte, dass zu Beginn der Pandemie (März 2020) die geringere Betroffenheit der nördlichen und östlichen Bundesländer gegenüber dem Rest noch nicht durchschlägt. Ansonsten ist der Zusammenhang statistisch gesichert und weist im Zeitablauf eine zunehmende Bedeutung auf. Je höher der Migrantenanteil in einem Bundesland ist, umso höher fällt die Zahl der Infektionen aus. Die Bedeutung dieses Faktors nimmt jedoch im Verlauf der Pandemie ab und ab 2021 werden keine signifikanten Zusammenhänge mehr ausgewiesen. Ganz am Ende des Beobachtungszeitraums ist sogar ein Vorzeichenwechsel zu beobachten. Während bis Herbst 2020 armutsgefährdete Familien einer geringeren Infektionsgefahr ausgesetzt waren, wird für die spätere Zeit ein Vorzeichenwechsel offenbar. Daraus kann man aber keine statistisch gesicherte Zunahme der Infektionsgefährdung der ärmeren Bevölkerung ablesen. Der einzige stabile, keinen systematischen Veränderungen unterliegende Faktor ist das Persönlichkeitsmerkmal CONSC. Menschen, die zu Perfektionismus neigen, die sich nicht von

äußeren Gegebenheiten wie der Corona-Pandemie beeinflussen lassen, sind danach einem größeren Infektionsrisiko ausgesetzt.

Insbesondere vor dem Hintergrund des Verlaufs der Pandemie mit einer ersten, zweiten und dritten Welle (W1, W2 und W3), unterbrochen jeweils durch eine Phase verminderter Neuinfektionen (Zwischenphase 1 und 2 (ZP1, ZP2), ist eine Unterteilung in fünf Phasen (W1, ZP1, W2, ZP2, W3) angebracht, wie die nachstehende Graphik des RKI nahelegt.



Eine eindeutige Abgrenzung der Phasen ist nicht möglich. Zum Teil zeigen sich schon Tendenzen einer neuen Welle, wenn sich dies in den Indikatoren noch nicht äußert. So wurde am 25.09.2020 im Bayrischen Rundfunk (BR24) die Frage diskutiert, ob wir uns wegen steigender Infektionszahlen schon am Beginn der zweiten Welle befinden oder ob es immer noch die erste Welle ist. Karl Lauterbach hat am 20.02.2021 in der Tagesschau geäußert, dass wir uns bereits am Beginn der dritten Welle befinden. Nach Augenschein der RKI-Graphik lässt sich eine vorläufige Einteilung vornehmen:

W1: bis Ende Mai 2020



ZP1: Juni 2020 bis August 2020

W2: September 2020 bis Ende Januar 2021

ZP2: Anfang Februar 2021 bis Mitte März 2021

W3: ab 18.03.2021

Alternativ werden die Zwischenphasen vernachlässigt, und es wird nur zwischen drei Wellen unterschieden (W1\*, W2\* und W3\*) mit der folgenden Einteilung:

W1\*: bis Ende Mai 2020

W2\*: ab Anfang Juni 2020 bis Anfang März 2021

W3\*: ab 08.03.2021

Wegen der Unschärfe der Abgrenzung zwischen den verschiedenen Phasen ist es angebracht, auf Strukturbrüche zu testen. Getestet wird nur für Deutschland insgesamt auf tagesweiser Basis der internationalen Daten der Johns-Hopkins-University, und zwar zwischen Welle 1\* und Welle 2\* für den Zeitraum 25.04.2020 bis 25.09.2020 und zwischen Welle 2\* und Welle 3\* für den Zeitraum zwischen 30.01.2021 und 17.03.2021.

Tabelle 14 ist zu entnehmen, dass zwischen W1\* und W2\* aufgrund eines Wald-Tests am 10. August 2020 ein Strukturbruch besteht, also nicht schon am Tiefpunkt der 7-Tage-Inzidenz, sondern erst später. Dann kann man den Anstieg nicht mehr als zufällig interpretieren. Zwischen der zweiten und der dritten Welle wird ein Bruch am 7. März 2021 angezeigt. Auch hier wieder das Muster, dass der Bruch nicht am Tiefpunkt der 7-Tage-Inzidenz, sondern mit einem gewissen Lag angezeigt wird.

Eine offene Frage ist, ob die Bestimmungsgründe für das regionale Infektionsfälle für den gesamten Zeitraum unverändert bleiben oder ob es wellenspezifische Strukturmuster gibt. Ergebnisse hierzu finden sich in Tabelle 15. Folgendes bleibt festzuhalten:

- (1) Die Unterschiede zwischen den nord-östlichen Bundesländern und den süd-westlichen hat von Welle zu Welle zugenommen.
- (2) Bundesländer mit hohem Migrantenanteil waren während der ersten und zweiten Welle eindeutig stärker von der Pandemie betroffen als andere.
- (3) In den Bundesländern mit einem vergleichsweise hohen Anteil an Einwohnern ohne Schulabschluss lag über alle drei Wellen hinweg die Infektionsbetroffenheit höher als in anderen Bundesländern. Die Bedeutung war jedoch rückläufig.

- (4) Für den Faktor Armutsgefährdung der Familien lassen sich ähnliche Aussagen treffen wie für den Faktor Schulbildung. In der dritten Welle ist der Zusammenhang mit den Corona-Ansteckungen statistisch nicht mehr gesichert.
- (5) In Bundesländern mit einem vergleichsweise hohen Anteil an emotional labilen, aber gewissenhaften und ungeduldigen Menschen werden mehr COVID-19-Fälle beobachtet als anderswo. Die Bedeutung der Faktoren Neurotizismus und Ungeduld hat von Welle zu Welle zugenommen. Zu Beginn der Pandemie hat sich noch kein statistisch eindeutiger Effekt gezeigt.

## **7 Fazit**

Der Beitrag hat Regionen als Betrachtungsebene gewählt. Das ist weder gleichbedeutend mit Analysen auf individueller oder betrieblicher Ebene noch auf gesamtwirtschaftlicher oder internationaler Ebene. Bei Bundesländern steht die Meso-Betrachtung im Vordergrund, in die trotz allem sowohl Elemente von der Mikro- als auch von der Makroebene eingehen.

Die Untersuchung hat herausgearbeitet, welche möglichen Einflüsse eher vernachlässigbar und welche nicht. Zu letzteren gehören vor allem die Agglomeration von Regionen, Migrationshintergrund, Schulbildung, Armut und aggregierte Persönlichkeitsmerkmale. Zwischen diesen bestehen Zusammenhänge. Eine saubere Abgrenzung ist nur schwer möglich. Entgegen verbreiteter Meinung sind Armutsgefährdung und Schulbildung ohne Abschluss nicht die entscheidenden Infektionstreiber. Je mehr Kooperationsbereitschaft und emotionale Labilität ausgeprägt sind, umso höher ist die Gefahr der Ansteckung. Zustandsabhängigkeit scheint für die Entwicklung ab der zweiten Welle wichtiger zu sein als unbeobachtete Heterogenität. Das heißt jedoch nicht, dass unbeobachtete Einflüsse bedeutungslos wären. Die Blinder-Zerlegung und Panelschätzungen liefern hierfür Anhaltspunkte. Ein signifikanter Zusammenhang zwischen Infektionen und Tests wird durch die Schätzungen abgebildet. Je mehr Tests in einem Bundesland durchgeführt werden, umso mehr Infektionen werden aufgedeckt. Eine weniger klare Verknüpfung zeigt sich zwischen Impfungen und der Zahl der Infektionen. Das ausgewiesene Vorzeichen entspricht zwar den Erwartungen – je mehr geimpft wird, umso weniger Neuinfektionen werden beobachtet. Die zugrunde gelegte Spezifikation ist aber ausschlaggebend dafür, ob sich ein signifikanter Zusammenhang zeigt.

Als wichtigstes Ergebnis der dynamischen Analyse bleibt festzuhalten, dass die Bedeutung des Migrantenanteils für die Ausbreitung der Neuinfektionen im Laufe der Pandemie abgenommen hat, sei es, weil sich die Migranten zunehmend besser informiert haben, ihr Verhalten angepasst haben oder weil sie im Vergleich zum Anfang besser aufgeklärt worden sind. Auch die Bedeutung des Faktors Armut ist Veränderungen unterworfen. Während armutsgefährdete

Familien in der ersten Phase der Pandemie einer vergleichsweise geringen Infektionsgefahr ausgesetzt waren, kehrt sich dieses Ergebnis später um. Die geringere Betroffenheit der nördlichen und östlichen Bundesländer gegenüber dem Rest ist in der zweiten und dritten Welle deutlicher hervorgetreten.

Insgesamt zeigen die empirischen Analysen eine deutliche Spezifikations- und Pandemiedauerabhängigkeit. Es bedarf einer sehr sorgfältigen Analyse, um zu belastbaren Ergebnissen zu kommen. Unbeobachtete, sich im Zeitablauf verändernde Einflüsse lassen den Eindruck eines diffusen Erscheinungsbildes aufkommen und erschweren das Aufdecken eindeutiger Zusammenhänge.

Bei weiteren Untersuchungen sollte geprüft werden, inwiefern die Ergebnisse sensitiv auf eine Veränderung der Messung der verwendeten Variablen reagieren. Aufzuklären ist, ob auf Landkreisebene, bei Verlängerung des Beobachtungszeitraums, bei alternative Datensätzen und Verwendung weiterer Variablen die gefundenen Zusammenhänge robust sind. Die Warnungen der Gesundheitsbehörde in den USA (30.03.2021 - FR), des Gesundheitspolitikers Lauterbach und des Virologen Streeck (02.06.2021 - ntv) vor einer vierten Welle im Herbst sollten ernst genommen werden. Auf regionaler Ebene ist zu untersuchen, ob und gegebenenfalls wo und ab wann deutliche Hinweise für eine neue Welle zu erkennen sind.

## Literatur

Akbarpour M, Cook C, Marzuoli A, Mongey, S, Nagaraj A, Saccarolak M, Tebaldi P, Vasserman S und Yang H (2020), Socioeconomic Network Heterogeneity and Pandemic Policy Response, Becker Friedman Institute for Economics, Working Paper Series, Vol. 11 No. 17

Belloni A, Chen D, Chernozhukov V, Hansen C (2012), Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain, *Econometrica* 80:6, 2369-2429

Benitez J, Courtemanche C, Yelowitz A (2020), Racial and Ethnic Disparities in COVID-19: Evidence from Six Large Cities, IZA Discussion Paper No. 13521

Bonelli R M (2014), Perfektionismus: Wenn das Soll zum Muss wird. Pattloch-Verlag, München.

Blinder A. (1973), Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources* 18, 436-455

Brown C S, Ravallion M (2020), Inequality and the Coronavirus: Socioeconomic Covariates of Behavioral Responses and Viral Outcomes across US Counties, NBER Working Paper 27549.

Burdett A, Davillas A and Etheridge B (2021), Weather, Psychological Wellbeing and Mobility during the first wave of the COVID-19 Pandemic, IZA DP No. 14119

Desmet K, Wacziarg, R (2020), Understanding Spatial Variation in COVID-19 across the United States, NBER Working Paper 27329

Galasso V, Pons V, Profeta P, Becher M, Brouard S, Foucault M (2020), Gender Differences In COVID-19 Related Attitudes And Behavior: Evidence From A Panel Survey In Eight OECD Countries, NBER Working Paper 27359

Efron B, Hastie T, Johnstone I, Tibshirani R (2004), Least Angle Regression, *Annals of Statistics* 32:2, 407-451

Gelbach J B (2016), When Do Covariates Matter? And Which Ones, and How Much? *Journal of Labor Economics* 34:2, 509-543

Goldstein J R, Lee R D (2020), Demographic Perspectives on Mortality of COVID-19 and Other Epidemics, NBER Working Paper 27043

Heckman J J (1978), Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence against the Hypothesis of Spurious State Dependence, *Annales de l'INSEE* 30, 227-269

Heckman J J (1981), Heterogeneity and State Dependence, in: Studies in Labor Markets, ed. by S. Rosen, University of Chicago Press, 91-139

Hübler O (2014), Estimation of standard errors and treatment effects in empirical economics—methods and applications, *J Labour Market Res.* 47:43–62

Kaiser H F, Dickman K W (1959), Analytical determination of common factors, *American Psychologist* 14, 425

Knittel C R, Ozaltun B (2020), What Does and Does Not Correlate With COVID-19 Death Rates, NBER Working Paper 27391

Krekel C, Swanke S, De Neve J-E, Fancourt D (2020), Are Happier People More Compliant? Global Evidence from Three Large-Scale Surveys during Covid-19 Lockdowns, IZA Discussion Paper 13690

McLaren J (2020), Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census Data, NBER Working Paper 27407

Moulton B R (1990), An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Rev. Econ. Stat.* 72, 334–338

Ost F (1984), Faktorenanalyse, in: Fahrmeir, L. und A. Hamerle, *Multivariate statistische Verfahren*, Berlin, de Gruyter

Papageorge N W, Zahn M V, Belot M, van den Broek-Altenburg E and Choi S (2020), Socio-Demographic Factors Associated with Self-Protecting Behavior during the COVID-19 Pandemic, IZA Discussion Paper 13333

Qui Y, Chen X and Shi W. (2020), Impacts of Social and Economic Factors on the Transmission of Coronavirus Disease 2019 (COVID-19) in China, IZA Discussion Paper 13165

Sa F (2020), Socioeconomic Determinants of COVID-19 Infections and Mortality: Evidence from England and Wales, IZA Policy Paper No.159

Toya H and Skidmore M (2021), A Cross-Country Analysis of the Determinants of Covid-19 Fatalities, CESifo Working Papers 9028

## Tabelle 1: Definition, Messung und Datenquelle regionaler Variablen

ABKÜRZUNG	ERKLÄRUNG	MESSUNG	QUELLE
ALQ	Arbeitslosenquote	in Prozent	Stat. Bundesamt 2019
BIP	Bruttoinlandsprodukt	in Mill. EUR pro Kopf	Stat. Bundesamt 2019
LEBERW	Lebenserwartung	in Jahren	Stat. Bundesamt 2019
FRAUEN	Frauenanteil	in Prozent	Stat. Bundesamt 2016
UNI	Hochschulreife	in Prozent	Stat. Bundesamt 2020
SCHULE	ohne Schulabschluss	in Prozent	Stat. Bundesamt 2020
LZUF	Lebenszufriedenheit	0 (unzufrieden) ...,10	Stat. Bundesamt 2019
STRAF	Straftaten	absolute Zahl	Stat. Bundesamt 2020
VSCHULD	Verschuldung	in EURO	Stat. Bundesamt 2020
BETREU	Betreuungsquote Kinder <3J.	in Prozent	Stat. Bundesamt 2020
ABSCHIEB	Abschiebungen	absolute Zahl	Stat. Bundesamt 2019
ARMUT	Armutgefährdungsquoten	in Prozent	Stat. Bundesamt 2019
J_ALQ	Jugendarbeitslosenquote	in Prozent	Stat. Bundesamt 2020
STERBE	Sterbefälle	absolute Zahl	Stat. Bundesamt 2019
ASYL	Asylanträge	absolute Zahl	Stat. Bundesamt 2020
AUFKLÄR	Aufklärungsquote Straftaten	in Prozent	Stat. Bundesamt 2020
GEFANG	Gefangenenzahl in JVA	absolute Zahl	Stat. Bundesamt 2020
TEMP	Jahrestemperatur	Grad Celsius	Stat. Bundesamt 2019
SONNE	Sonnenschein pro Jahr	in Stunden	Stat. Bundesamt 2019
REGEN	Niederschlagsmenge	ltr/qm	Stat. Bundesamt 2019
L_GRENZE	Auslandsgrenze	in km	Statista 2017
CDU	Wähleranteil letzte Landtagswahl	in Prozent	Statista bis 2020
SPD	Wähleranteil letzte Landtagswahl	in Prozent	Statista bis 2020
GRÜNE	Wähleranteil letzte Landtagswahl	in Prozent	Statista bis 2020
FDP	Wähleranteil letzte Landtagswahl	in Prozent	Statista bis 2020
LINKE	Wähleranteil letzte Landtagswahl	in Prozent	Statista bis 2020
AfD	Wähleranteil letzte Landtagswahl	in Prozent	Statista bis 2020
JUGEND	Jugendliche 15-25 Jahre	in Prozent	Zensus 2011

ALTE	ältere Personen >= 65 Jahre	in Prozent	Zensus 2011
HH-GRÖSSE	Haushaltsgröße	Personen pro Haushalt	Mikrozensus 2018
AUSLAND	Ausländeranteil	in Prozent	GENESIS 2016
B-INDEXX	regionaler Beliebtheitsindex	Stimmenzahl 2015/16	YouGovInstitut
GESUND	Gesundheitszustand gut	1=ja; 0= sonst	SOEP 2018
KRANK	Gesundheitszustand schlecht	1=ja; 0=sonst	SOEP 2018
EXTRA	gesellig, aktiv, gesprächig	1-trifft nicht zu; 7-trifft voll zu	SOEP 2013
OPEN	aufgeschlossen, neugierig	1-trifft nicht zu; 7-trifft voll zu	SOEP 2013
AGRE	verträglich, kooperativ	1-trifft nicht zu; 7-trifft voll zu	SOEP 2013
NEURO	launenhaft, instabil	1-trifft nicht zu; 7-trifft voll zu	SOEP 2013
CONSC	gewissenhaft, effektiv	1-trifft nicht zu; 7-trifft voll zu	SOEP 2013
SELBEW	selbstbewusst	1=große Sorgen; 3=keine Sorgen	SOEP 2018
OPTIM	optimistisch	1=ganz und gar;...;4=nicht	SOEP 2018
GEDULD	geduldig	0=ungeduldig; 10=sehr geduldig	SOEP 2018
RISIKO	risikobereit	0=risikoscheu;10= risikobereit	SOEP 2018
IMPULS	impulsiv	0=gar nicht;10=stark	SOEP 2018

---

## Tabelle 2: Beschreibende Statistik für regionale Variablen

Variable	N	r	$\bar{x}$	s	min.	max.
BUNDESLAND	5,760		8.5	4.610172	1	16
PERIODE	4,576		143.4998	82.56954	1	286
CORONA-INFEKTIONEN	5,744		41187.14	79026.68	0	526695
TOTE	4,608	0.5851*	1156.46	1857.962	19	12424
R-Wert	4,544	-0.1492*	1.007	.250047	.54	2.88
VERDOPPLUNGSZEIT	4,571	-0.2881*	275.43	471.1449	4.4	9061.5
TESTS	4,528	0.4911*	951984.6	381928.4	326645	1578209
EINWOHNERZAHL	5,760	0.1123*	5.1875	4.78409	.7	17.9
BEV.DICHTE	5,760	0.1012*	6 89.43	1077.10	69	4090
AUSLANDSGRENZEN	5,760	0.0030	3.658	3.0602	0	8.8917
NORD/OST-SÜD/WEST	5,856	-0.0877*	.5000	.5000	0	1
WEST-OST	5,760	0.1594*	6875	.46355	0	1
FLÄCHE-STADTSTAAT	5,760	-0.0877*	.8125	.39034	0	1
HAUSHALTSGRÖSSE	5,760	0.0349*	1.9518	.08324	1.79	2.09
FRAUENANTEIL in %	5,760	0.0101	50.7062	.18191	50.4	51
ANTEIL JUGENDLICHE %	5,760	0.1450*	10.6875	1.0043	9	12
ANTEIL ÄLTERE in %	5,760	-0.1505*	21.1687	1.6911	19	24.8
MIGRANTEN	5,760	0.1848*	10.925	4.9954	4.5	18.5
INLANDSPRODUKT	5,760	0.1721*	39193.69	9842.6	27905	65603
ARBEITSLOSENQUOTE	5,760	-0.0518*	5.943	1.7472	3.2	10.2
ARBEITSLOSE JUGENDL.	5,760	-0.1154*	5.806	2.0297	2.5	9.3
VERSCHULDUNG	5,760	0.0755*	37650	39598.	1388	177670
ARMUT	5,760	-0.1104*	16.555	2.7762	11.7	22.7
TEMPERATUR	5,760	-0.0252	10.525	.51905	9.5	11.7
SONNE	5,760	0.0159	1807.188	105.45	1645	1970
REGEN	5,760	0.1304*	707.5	143.48	475	980
CDU	5,760	0.0418*	27.25	7.9549	11.2	40.7
SPD	5,760	-0.0175	23.287	10.389	7.7	39.2



GRÜNE	5,760	0.1334*	12.268	7.5931	4	30.3
FDP	5,760	0.0411*	6.3187	2.6089	3	12.6
LINKE	5,760	-0.0666*	9.9437	7.0331	2.8	31
AfD	5,760	-0.1111*	13.856	7.4932	5.3	27.5
STRAFTATEN	5,760	0.1125*	339900.1	287121	74719	1227929
AUFKLÄRUNGSQUOTE	5,760	-0.0494*	57.175	6.5883	44.7	67
GEFANGENENZAHL	5,760	0.1231*	3723.563	3736.5	697	14490
ASYLANTRÄGE	5,760	0.1023*	2955.625	2781.0	449	11578
ABSCHIEBUNGEN	5,760	0.1105*	1347.438	1578.3	93	6359
BETREUUNGSQUOTE	5,760	-0.1464*	40.225	11.733	28.2	58
ohne SCHULABSCHLUSS	5,760	-0.0272	7.4625	4.2890	4.7	18.5
UNIVERSITÄT	5,760	0.0456*	40.8375	5.847395	32.1	54.8
GESUNDHEIT gut	5,760	0.0373	45.48563	5.115132	37	59.41
GESUNDHEIT schlecht	5,529	0.0684*	20.01016	2.541344	14	24.27
LEBENSERWARTUNG	5,760	0.1469*	78.075	.875075	76.3	79.7
STERBEFÄLLE	5,760	0.0976*	58720	51809.92	7704	206479
LEBENSZUFRIEDEN	5,760	0.3221*	311.38	1361.181	6.76	10457
GESELLIG	5,760	0.4056*	25.20292	35.43701	14.709	194.3
AUFGESCHLOSSEN	5,760	0.2473*	13.7589	.8282921	12.43	15.56
VERTRÄGLICH	5,760	-0.0649*	14.34377	.6381826	12.43	15.65
VERLETZLICH	5,760	0.3559*	12.69513	.6991707	11.78	15.65
GEWISSENHAFT	5,760	-0.2612*	14.06465	.6201183	11.78	15.23
GEDULDIG	5,760	0.4231*	6.76289	2.293641	5.86	15.23
OPTIMISTISCH	5,760	-0.0742*	2.276363	.0938255	2.12	2.43
RISIKOFREUDIG	5,760	0.0005	4.229663	.377912	3.97	4.46
IMPULSIV	5,760	0.1413*	4.834106	.1585487	4.55	5.11
SELBSTBEWUSST	5,760	0.1186*	2.327325	.0645497	2.21	2.41
B-Index	5,760	0.1387*	21.66875	18.93715	-13	59.1

-----  
Quelle: Statistisches Bundesamt und Sozio-ökonomisches Panel

**Tabelle 3: OLS-Regressionen, getrennt für einzelne inhaltliche Blöcke**

VAR. t_rob	Koef.	t	t_rob	VAR.	Koef.	t
BEVD	.2024	8.76	3.28	AUSLAND	.0986	9.78 2.16
EINWZ	.0152	3.02	1.33	JUGEND	-.1676	-3.71 -1.11
WEST-OST	.3211	4.74	1.38	ALTE	.0273	1.27 0.36
NORD/OST-SÜD/WEST	-.4848	-6.95	-2.53	FRAUEN	.1372	1.16 0.38
NORD-SÜD	.0555	0.60	0.24	HH-GRÖSSE	.0558	5.31 1.44
L-GRENZE	-.0044	-0.72	-0.22			
cons	7.907	100.43	33.71		2.733	-0.44 -0.14
N	5,775				5,775	
R <sup>2</sup>	0.06				0.05	
TEMP	.2823	6.71	1.27	BIP	.0217	9.45 1.94
SONNE	.0011	6.50	1.45	ALQ	.1032	3.54 0.84
REGEN	.0022	14.50	2.92	ALQ_J	-.0422	-1.70 -0.37
				ARMUT	-.0783	-4.47 -1.22
cons	1.743	2.65	0.51		8.495	40.09 12.39
N	5,040				5,040	
R <sup>2</sup>	0.04				0.06	
ASYL	-.0723	-1.13	-0.24	LEBERW	.2450	10.16 3.07
ABSCHIEB	-.4016	-5.20	-1.53	GESUND	.0216	6.13 1.29
STRAF	-.6180	-1.25	-0.30	KRANK	.0548	7.50 1.48
AUFKLÄR	-.0310	-6.41	-1.12	STERBE	.3700	3.50 1.34
GEFANG	.3341	8.53	2.52			
VSCHULD	-.6230	-0.46	-0.13			
cons	9.9087	35.41	6.54		-12.8956	-6.84 -2.03

---

N	5,040				5,018		
R <sup>2</sup>	0.05				0.05		

---

UNI	.0239	7.65	2.07	CDU	.0143	2.13	0.63
SCHULE	-.0151	-3.59	-0.77	SPD	-.0069	-1.18	-0.41
BETREU	-.0223	-14.84	-3.54	GRÜNE	.0302	4.30	1.35
FDP	-.0180	-2.10	-0.41	LINKE	.0115	1.82	0.67
				AfD	-.0247	-3.39	-1.12
cons	8.4512	65.47	15.53		8.1562	14.36	4.89

---

N	5,040			5,040
R <sup>2</sup>	0.05			0.05

---

LZUF	-.2479	-2.96	-7.28
EXTRA	-.7377	-3.47	-1.28
OPEN	.0522	1.04	0.33
AGRE	.0249	0.40	0.14
NEURO	.1429	1.24	0.49
CONSC	.2642	2.93	1.09
OPTIM	1.8900	3.51	0.95
GEDULD	.8442	3.50	1.57
RISIKO	-.7934	-2.17	-0.92
IMPULS	1.5225	7.40	2.29
B-INDEX	.0155	6.85	1.45
cons	.8890	0.26	0.10

---

N	5,040
R <sup>2</sup>	0.08

---

**Tabelle 3a: Schätzung mit jeweils einer Variablen aus allen 9 Erklärungsblöcken**

```

-----
Linear regression                Number of obs   =    6,730
                                F(9, 15)         =    145.73
                                Prob > F           =    0.0000
                                R-squared          =    0.0416
                                Root MSE       =    1.7571
-----

```

lnINFI_BEV	Coef.	Robust		
		Std. Err.	t	P> t
NORD/OST-SÜD/WEST	-.1333292	.1073942	-1.24	0.233
MIGRANTEN	.0294469	.0098419	2.99	0.009
UNI	.0206309	.0091839	2.25	0.040
NIEDERSCHLAG	-.0020269	.0003734	-5.43	0.000
BIP .	.0000182	3.17e-06	5.75	0.000
CDU	.0122721	.0084891	1.45	0.169
GEFANG	.0000514	.0000132	3.88	0.001
LEBERW	.1024999	.0423489	2.42	0.029
GEDULDIG	1.808913	.3166332	5.71	0.000
cons	-11.56592	3.938666	-2.94	0.010

**Tabelle 4: Todesfälle, bezogen auf die Bevölkerung der Bundesländer, als Anfangsbeobachtung in der zweiten und dritten Welle (State Dependence) und regionale durchschnittliche Lebenszufriedenheit vor Beginn der Pandemie als Heterogenitätsindikator**

	(1)	(2)	(3)	(4)
TOTE/EINWZ	0.035** (0.01)		0.029* (0.01)	0.001 (0.01)
LZUF	-0.021 (0.02)	-0.014 (0.02)		
NORD/OST-SÜD/WEST				-0.306** (0.12)
MIGRANTEN				0.041*** (0.01)
SCHULE				-0.024** (0.01)
ARMUT				-0.001 (0.02)
CONSC				0.277* (0.13)
Constant	9.179*** (0.11)	9.431*** (0.22)	9.309*** (0.24)	5.228** (2.11)
N	4240	4240	4240	4240
R <sup>2</sup>	0.011	0.007	0.014	0.070

\* p<.10, \*\* p<0.05, \*\*\* p<.01

**Tabelle 5: Einfluss von Tests und Impfungen als Maßnahmen gegen Corona auf COVID-19-Infektionen**

	(1)	(2)	(3)	(4)
TESTS	0.625*** (0.04)		0.625*** (0.04)	0.625*** (0.04)
IMPFUNGEN_r		0.012 (0.02)	0.012 (0.02)	0.020 (0.02)
TOTE/EINWZ				0.029 (0.02)
LZUF_r				-0.023 (0.03)
cons	8.472*** (0.10)	9.133*** (0.18)	8.380*** (0.20)	8.433*** (0.23)
N	4128	4128	4128	4128
R <sup>2</sup>	0.010	0.003	0.013	0.032

\* p<.10, \*\* p<0.05, \*\*\* p<.01

**Tabelle 6: Hauptkomponentenanalyse-Scoring-Koeffizienten bei orthogonaler Varimax-Rotation**

Variable	Comp1	Comp2	Comp3	Unexplained
BEVD	0.1053	0.2747	0.0106	.1991
EINWZ	0.0121	-0.0591	0.3122	.03789
WEST-OST	0.2733	0.0119	0.0202	.09084
NORD/OST-SÜD/WEST	-0.0073	0.2154	-0.0714	.451
NORD-SÜD	-0.1008	0.1626	-0.1032	.38
MIGRANTEN	0.2328	0.1189	0.0354	.2044
JUGEND	0.2549	-0.0487	0.0364	.1275
ALTE	-0.2275	-0.0557	-0.0576	.2656
FRAUEN	0.0669	0.1655	0.0010	.7031
HH-GRÖSSE	0.0351	0.0223	-0.0487	.9699
UNI	0.0848	0.2415	0.0201	.3993
S CHULE	-0.0201	0.1872	0.0656	.7057
BETREU	-0.2216	0.0956	-0.0512	.2024
TEMP	-0.0660	0.3058	0.0518	.1612
SONNE	-0.1729	-0.0312	0.0608	.6894
REGEN	0.2262	-0.1889	-0.0473	.1638
BIP	0.2597	0.0729	-0.0570	.2476
ALQ	-0.0366	0.2449	-0.0149	.4047
J_ALQ	-0.1523	0.2250	-0.0165	.2077
ARMUT	-0.1076	0.1900	0.0124	.5438
CDU	0.0282	-0.2118	0.0428	.5066
SPD	0.1421	0.1133	-0.0292	.666
GRÜNE	0.1921	0.0238	-0.0384	.6104
FDP	0.0636	0.0283	0.2242	.4505
LINKE	-0.1480	0.0543	-0.0911	.5176
AfD	-0.2698	-0.0199	0.0281	.2056
ASYL	-0.0035	-0.0088	0.3423	.02199

ABSCHIEB		-0.0083	-0.0209	0.3309		.07455
STRAF		-0.0029	0.0458	0.3545		.02528
AUFKLÄR		-0.0477	-0.2644	0.0130		.3253
GEFANG		0.0074	-0.0255	0.3219		.07509
VSCHULD		0.0122	0.1207	0.3143		.2009
LEBERW		0.1501	-0.1098	0.0787		.4342
GESUND		0.1224	0.0261	-0.0921		.827
KRANK		0.0501	0.0875	-0.0246		.897
STERBE		-0.0102	-0.0567	0.3241		.03403
LZUF		0.2620	-0.1345	-0.1071		.2115
EXTRA		0.1309	0.2049	-0.0428		.416
OPEN		0.0949	0.2015	-0.0744		.469
AGRE		0.0774	0.1638	-0.0810		.6205
NEURO		0.0239	0.1107	-0.1470		.6511
CONSC		0.0709	0.1746	0.0885		.6305
OPTIM		-0.2322	0.0014	-0.0369		.2978
GEDULDIG		0.0713	-0.2139	-0.2294		.4415
RISKY		0.0246	-0.1373	-0.0368		.8385
IMPULSIV		0.2070	-0.0326	0.0958		.2453
BINDEX		0.2708	0.0785	0.0085		.09059

-----



**Tabelle 7: Clusterrobuste Regression mit drei Hauptkomponenten  
als Regressoren aus der PCA mit VARIMAX-Rotation**

Source	SS	df	MS	Number of obs	=	6,618
				F(3, 6614)	=	32.83
Model	322.418669	3	107.4728	Prob > F	=	0.0000
Residual	21652.251	6,614	3.2736	R-squared	=	0.0147
				Adj R-squared	=	0.0142
Total	21974.6697	6,617	3.3209	Root MSE	=	1.8093

lnINFI_BEV	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
PC1	.0569547	.0071503	7.97	0.000	.0429378	.0709715
PC2	.0073909	.007675	0.96	0.336	-.0076547	.0224364
PC3	.0193926	.008784	2.21	0.027	.0021731	.036612
_cons	8.326181	.0222411	374.36	0.000	8.282581	8.36978

**Tabelle 8: Clusterrobuste Regression mit den Regressoren, die nach LARS ausgewählt werden**

	(1)	(2)	(3)	(4)
BEVD	0.001** (0.00)		0.001** (0.00)	0.001* (0.00)
NORD/OST-SÜD/WEST	-0.569 (0.43)	-0.334 (0.48)	-0.569 (0.43)	-0.378** (0.15)
MIGRANTEN	-0.092 (0.15)	0.091 (0.07)	-0.092 (0.15)	-0.241 (0.18)
UNI	-0.024 (0.06)	-0.029 (0.06)	-0.024 (0.06)	-0.002 (0.02)
BETREU	0.013 (0.06)	0.031 (0.05)	0.013 (0.06)	-0.037 (0.08)
REGEN	0.002 (0.00)	0.000 (0.00)	0.002 (0.00)	0.001 (0.00)
BIP	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	0.000** (0.00)
GEFANG	-0.000 (0.00)	-0.000 (0.00)	-0.000 (0.00)	
LEBERW	0.389 (0.65)	0.155 (0.63)	0.389 (0.65)	0.290 (0.23)
STERBE	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	
LZUF	-0.206*** (0.03)	-0.168*** (0.04)	-0.206*** (0.03)	-0.164** (0.06)
EXTRA	0.365 (2.57)	1.901 (1.87)	0.365 (2.57)	-1.145* (0.58)
AGRE	0.297 (0.23)	0.084 (0.21)	0.297 (0.23)	0.237 (0.67)
OPTIM	-2.135	-2.261	-2.135	2.077*

	(9.82)	(9.87)	(9.82)	(1.16)
GEDULDIG	-1.121	1.070	-1.121	-1.223
	(1.31)	(1.36)	(1.31)	(1.40)
IMPULSIV	0.643	-0.229	0.643	0.418
	(1.67)	(1.50)	(1.67)	(3.95)
B-INDEX	-0.014	-0.011	-0.014	-0.001
	(0.04)	(0.04)	(0.04)	(0.01)
SELFCON	-5.640	-6.490	-5.640	
	(6.90)	(5.71)	(6.90)	
KRANK				-0.003
				(0.17)
cons	-9.816	-19.542	-9.816	2.456
	(31.23)	(28.71)	(31.23)	(10.38)

---

N	5824	5824	5824	5534
R <sup>2</sup>	0.079	0.075	0.079	0.079

---

\* p<.10, \*\* p<0.05, \*\*\* p<.01

**Tabelle 9: Regressionen mit Spezifikation nach RLASSO**

	(1)	(2)	(3)
NORD/OST_SÜD/WEST	-0.349** (0.12)	-0.259** (0.09)	-0.397*** (0.09)
MIGRANTEN	0.045*** (0.01)	0.050*** (0.01)	0.039*** (0.01)
HH-GRÖSSE	0.006*** (0.00)		
SCHULE	-0.034*** (0.01)	-0.034*** (0.01)	
ARMUT	-0.020 (0.02)	-0.016 (0.01)	
CONSC	0.384*** (0.13)	0.316*** (0.10)	0.265** (0.10)
NEURO			0.225* (0.12)
cons	3.119 (2.04)	3.974** (1.52)	1.557 (2.38)
N	6922	6922	6922
R <sup>2</sup>	0.066	0.038	0.034

\* p<.10, \*\* p<0.05, \*\*\* p<.01

## Fortsetzung Tabelle 9: Regressionen mit Spezifikation nach RLASSO

	(4)	(5)	(6)
NORD/OST-_SÜD/WEST	-0.318*** (0.10)	-0.329* (0.16)	-0.339*** (0.07)
MIGRANTEN	0.052*** (0.01)		0.025*** (0.01)
HH-GRÖSSE		0.005*** (0.00)	
BETREU	-0.012* (0.01)		
LEBERW			0.127** (0.06)
AGRE			0.327*** (0.06)
cons	7.961*** (0.13)	8.973*** (0.23)	-6.336 (5.02)
N	6922.000	6922.000	6922.000
R <sup>2</sup>	0.030	0.043	0.036

\* p<.10, \*\* p<0.05, \*\*\* p<.01

## Tabelle 10: Gelbach-Zerlegung der RLASSO-Spezifikation

Verhaltensvariablen: OPEN GEDULD

```

-----
lnINFI_BEV |      Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
MIGRANTEN  |   .0003936   .0022095    0.18   0.859   - .0039369   .0047241
-----+-----
SCHULE     |   -.00222    .0019255   -1.15   0.249   - .005994    .0015539
-----+-----
ARMUT      |  -.0023454   .0039205   -0.60   0.550   - .0100294   .0053386
-----+-----
_cons      |   2.191523   1.141315    1.92   0.055   - .0454121   4.428459
-----

```

Verhaltensvariablen: CONSC2013 NEURO2013

```

-----
lnINFI_BEV |      Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
AUSLÄNDER  |   .010021    .0021797    4.60   0.000    .0057488    .0142932
-----+-----
oSCHULE    |   .0036114   .0027146    1.33   0.183   - .0017091    .0089318
-----+-----
ARMUT      |   .0045524   .0032073    1.42   0.156   - .0017338    .0108385
-----+-----
_cons      |   5.404831   1.174332    4.60   0.000    3.103183    7.706479
-----

```

## **Tabelle 11: Blinder-Zerlegung der RLASSO-Spezifikation in**

### **beobachtete Komponenten (endowment) und unbeobachtete Einflüsse**

Decomposition results for variables (as %s)

Variable	Attrib	Endow	Coef
MIGRANTEN	36.2	-4.2	40.4
SCHULE	-14.0	-4.9	-9.0
ARMUT	19.3	-5.8	25.1
CONSC	-110.4	2.9	-113.3
Subtotal	-68.8	-12.0	-56.8

Summary of decomposition results (as %)

Amount attributable:		-68.8
- due to endowments (E):		-12.0
- due to coefficients (C):		-56.8
Shift coefficient (U):		33.3
Raw differential (R) {E+C+U}:		-35.5
Adjusted differential (D) {C+U}:		-23.5
Endowments as % total (E/R):		33.8
Discrimination as % total (D/R):		66.2

U = unexplained portion of differential

(difference between model constants)

D = portion due to discrimination (C+U)

positive number indicates advantage to high group

negative number indicates advantage to low group

**Tabelle 12: RE- und BE-Panelschätzungen mit Bundesländer-Dummies für RLASSO-Spezifikationen**

	(1)	(2)	(3)	(4)
MIGRANTEN	0.071***	0.079***	0.070	0.078
(0.02)	(0.02)	(0.07)	(0.09)	
SCHULE	-0.033***	-0.026***	-0.032	-0.026
	(0.01)	(0.01)	(0.07)	(0.08)
ARMUT	0.006	0.014	0.006	0.015
	(0.02)	(0.02)	(0.06)	(0.02)
CONSC	0.464***	0.598***	0.463*	0.594***
	(0.12)	(0.14)	(0.23)	(0.13)
NEURO		0.176		0.172
		(0.11)		(0.23)
cons	0.898	-3.618	0.955	-3.468
	(1.98)	(3.52)	(3.35)	(6.47)
N	6922	6922	6922	6922

\* p<.10, \*\* p<0.05, \*\*\* p<.01



**Tabelle 13: Regressionen, basierend auf RLASSO-Spezifikationen und aufgliedert nach Monaten**

	März 2020	April 2020	Mai 2020	Juni 2020	Juli 2020
NORD/OST-SÜD/EST	-0.088 (0.10)	-0.240*** (0.03)	-0.204*** (0.04)	-0.192*** (0.04)	-0.205*** (0.04)
MIGRANTEN	0.083*** (0.01)	0.056*** (0.00)	0.067*** (0.00)	0.077*** (0.00)	0.079*** (0.00)
SCHULE	-0.039*** (0.01)	-0.047*** (0.00)	-0.052*** (0.00)	-0.054*** (0.00)	-0.052*** (0.00)
ARMUT	-0.013 (0.02)	-0.079*** (0.01)	-0.065*** (0.01)	-0.047*** (0.01)	-0.041*** (0.01)
CONSC	0.371*** (0.13)	0.484*** (0.04)	0.308*** (0.05)	0.216*** (0.05)	0.250*** (0.05)
cons	-1.240 (1.91)	1.440** (0.55)	3.782*** (0.67)	4.791*** (0.76)	4.250*** (0.77)
N	80	80	80	80	80
R <sup>2</sup>	0.615	0.958	0.934	0.913	0.912

\* p<.10, \*\* p<0.05, \*\*\* p<.01

## Fortsetzung 1 Tabelle 13: Regressionen, basierend auf RLASSO-Spezifikationen und aufgliedert nach Monaten

	August2020	September 2020	Oktober 2020	November 2020	Dezember 2020
NORD/OST-SÜD/WEST	-0.196*** (0.04)	-0.201*** (0.03)	-0.210*** (0.04)	-0.315*** (0.04)	-0.444*** (0.06)
MIGRANTEN	0.079*** (0.00)	0.085*** (0.00)	0.088*** (0.00)	0.077*** (0.00)	0.042*** (0.01)
SCHULE	-0.046*** (0.00)	-0.040*** (0.00)	-0.032*** (0.00)	-0.028*** (0.00)	-0.023*** (0.01)
ARMUT	-0.039*** (0.01)	-0.042*** (0.01)	-0.020*** (0.01)	0.005 (0.01)	0.015 (0.01)
CONSC	0.243*** (0.05)	0.228*** (0.04)	0.261*** (0.05)	0.298*** (0.06)	0.346*** (0.08)
cons	4.346*** (0.71)	4.669*** (0.65)	4.162*** (0.70)	4.029*** (0.79)	4.326*** (1.14)
N	80	80	80	80	80
R <sup>2</sup>	0.923	0.941	0.929	0.895	0.700

\* p<.10, \*\* p<0.05, \*\*\* p<.01

## Fortsetzung 2 Tabelle 13: Regressionen, basierend auf RLASSO-Spezifikationen und aufgliedert nach Monaten

	Januar 2120	Februar 2120	März 2120	April 2120
NORD/OST-SÜD/WEST	-0.396*** (0.07)	-0.358*** (0.07)	-0.329*** (0.06)	-0.317*** (0.06)
MIGRANTEN	0.009 (0.01)	0.003 (0.01)	-0.001 (0.01)	-0.003 (0.01)
SCHULE	-0.017** (0.01)	-0.015* (0.01)	-0.014* (0.01)	-0.012* (0.01)
ARMUT	0.013 (0.01)	0.014 (0.01)	0.016 (0.01)	0.018 (0.01)
CONSC	0.324*** (0.09)	0.302*** (0.09)	0.278*** (0.08)	0.253*** (0.08)
cons	5.420*** (1.29)	5.895*** (1.22)	6.352*** (1.16)	6.837*** (1.15)
N	80	80	80	80
R <sup>2</sup>	0.438	0.390	0.358	0.325

\* p<.10, \*\* p<0.05, \*\*\* p<.01

## Tabelle 14: Wald-Test auf Strukturbruch bei unbekanntem Bruchzeitpunkt

zwischen erster und zweiter Welle (W1\* und W2\*)

Geschätzter Strukturbruchzeitpunkt: 48

Ho: kein Strukturbruch

Test	Statistik	p-value
swald	822.8963	0.0000

zwischen zweiter und dritter Welle (W2\* und W3\*)

Geschätzter Strukturbruchzeitpunkt: 108

Ho: kein Strukturbruch

Test	Statistik	p-value
swald	52.1065	0.0000

**Tabelle 15: Wellenweise Regressionsschätzungen für COVID-19-Infektionen, basierend auf gemeinsamer RLASSO-Spezifikation**

	W1*	W2*	W3*
NORD/OST-SÜD/WEST	-0.175** (0.08)	-0.396*** (0.05)	-0.458*** (0.02)
MIGRANTEN	0.068*** (0.01)	0.041*** (0.00)	-0.016*** (0.00)
SCHULE	-0.049*** (0.01)	-0.023*** (0.01)	-0.010*** (0.00)
ARMUT	-0.045*** (0.01)	-0.022*** (0.01)	-0.002 (0.00)
NEURO	0.063 (0.08)	0.324*** (0.05)	0.398*** (0.02)
CONSC	0.375*** (0.09)	0.333*** (0.05)	0.338*** (0.02)
GEDULDIG	0.101 (0.25)	-0.390*** (0.15)	-0.657*** (0.06)
cons	0.468 (1.96)	2.803** (1.14)	5.075*** (0.45)
N	2538	3344	720
R <sup>2</sup>	0.087	0.119	0.596

\* p<.10, \*\* p<0.05, \*\*\* p<.01