

Eren Aydin, Wolfgang Maennig

# Das Hamburger Dieserverbot: ein Beispiel für die Schwierigkeit von Politikevaluierung

**Hamburg erließ Durchfahrverbote für bestimmte Dieselfahrzeuge in stark belasteten Straßenabschnitten mit dem Ziel, die Schadstoffemissionen zu reduzieren. Der Erfolgsnachweis ist schwierig. Relativ kleine Änderungen bezüglich der analysierten Zeiträume, Kontrollgruppen und Kontrollvariablen können zu signifikant unterschiedlichen Ergebnissen führen. Diese Untersuchung ist ein Beispiel für die Schwierigkeit von Ex-post-Evaluierungen politischer Maßnahmen.**

Ende der 1990er Jahre wurden in der EU Richtlinien für die Emission von Schadstoffen eingeführt. Die für den Automobilverkehr relevanteste Richtlinie 1999/30/EG des Rats der Europäischen Union legt für Stickstoffdioxid einen Kalenderjahresmittelwert von  $40 \mu\text{g}/\text{m}^3$  und ein Stundenlimit von  $200 \mu\text{g}/\text{m}^3$  als Obergrenze fest, die bis zu 18-mal in einem Kalenderjahr überschritten werden darf. Der Grenzwert im Jahresmittel für Feinstaub beträgt ebenfalls  $40 \mu\text{g}/\text{m}^3$ , wobei ein Limit von  $50 \mu\text{g}/\text{m}^3$  maximal 35-mal innerhalb eines Kalenderjahrs überschritten werden darf. Diese Grenzwerte wurden in mehreren deutschen Städten überschritten und führten Ende der 2000er Jahre zur Einführung von Umweltzonen und jüngst vereinzelt zu Durchfahrverboten, beispielsweise in Berlin und Hamburg. Hamburg reagierte als erste deutsche Stadt auf zu hohe Stickstoffdioxidwerte in Form von Durchfahrverboten, ohne vorher eine Umweltzone eingeführt zu haben.

In Hamburg werden die Schadstoffwerte zurzeit durch 15 Luftmessstationen gemessen. Darunter befinden sich vier sogenannte verkehrsnahen Messstationen in der Habichtstraße, der Kieler Straße, der Max-Brauer-Allee und der Stresemannstraße. Sämtliche verkehrsnahen Messstationen verzeichneten von 2003 bis 2018 Stickstoffdioxidmesswerte, die den Grenzwert für das Jahresmittel überschritten. Bundesweit ist ein Trend abnehmender Emissionen zu erkennen. Die vier Hamburger verkehrsnahen Messstationen weisen jedoch vergleichsweise hohe Messwerte auf.

Hamburg kündigte am 30.6.2017 Durchfahrverbote ab dem 31.5.2018 an, um die Stickstoffdioxidbelastung an den entsprechenden Stationen zu reduzieren (Behörde

für Umwelt und Energie, 2018). Die Durchfahrtsbeschränkung umfasst dabei Straßenbereiche an zwei der vier Messstationen: einen Abschnitt von ca. 580 m in der Max-Brauer-Allee und eine Strecke von ca. 1600 m in der Stresemannstraße. Die Straßenabschnitte waren von 2006 bis 2015 regelmäßig am stärksten belastet. Allerdings zeigte die Messstation in der Habichtstraße seit 2015 die höchste Belastung. In der Max-Brauer-Allee sind „Kraftwagen und sonstige mehrspurige Kraftfahrzeuge mit Dieselantrieb bis einschließlich Emissionsklasse Euro 5/V“ von dem Fahrverbot betroffen. Die Stresemannstraße ist für Fahrzeuge der gleichen Emissionsklassen eingeschränkt, jedoch ausschließlich ab einem zulässigen Gesamtgewicht von mehr als 3,5 t einschließlich ihrer Anhänger und Zugmaschinen. Ausgenommen vom Fahrverbot sind in beiden Straßenabschnitten Busse und Anliegerverkehr (vgl. Abbildung 1).

## Fahreinschränkungen und Luftqualität

Evaluierungen für unterschiedliche Fahreinschränkungen liegen im internationalen Bereich vor. Beispielsweise zeigen die in Asien und Lateinamerika verbreitete Form einer Fahreinschränkung durch eine sogenannte Nummernschildregelung (Davis, 2008; Xu et al., 2015), Maut-Zonen wie in London (Ellison et al., 2013) und Umweltzonen in Deutschland (Wolff, 2014; Gehrsitz, 2017) die Vielfalt im

**Eren Aydin, M. Sc., ist wissenschaftlicher Mitarbeiter an der Universität Hamburg.**

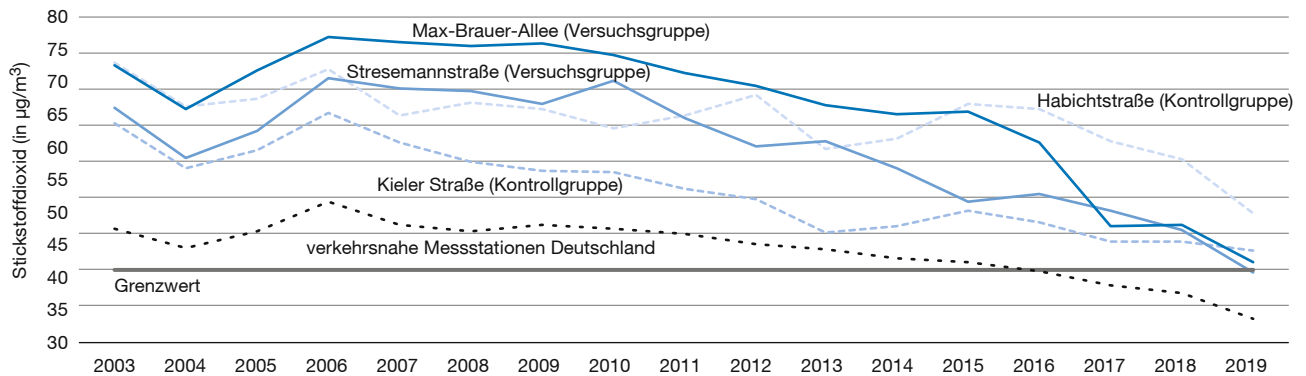
**Prof. Dr. Wolfgang Maennig ist Professor an der Universität Hamburg.**

© Der/die Autor(en) 2020. Open Access: Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<https://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht.

Open Access wird durch die ZBW – Leibniz-Informationszentrum Wirtschaft gefördert.

Abbildung 1

## Jahres-Durchschnittswerte von Stickstoffdioxid-Messungen an Verkehrsstationen in Hamburg



Messstationen an Straßen mit Fahrverbot (ab 31.5.2018) sind mit durchgezogenen Linien abgebildet, Messstationen an Straßen ohne Fahrverbot sind mit gestrichelten Linien abgebildet. Die graue gestrichelte Linie zeigt den Durchschnitt aller bundesweit verteilten verkehrsnahen Messstationen in Städten.

Quelle: Umweltbundesamt.

Design solcher Maßnahmen im internationalen Vergleich. Veränderungen der Luftqualität können sich auch auf die menschliche Gesundheit (Health Organization and Office for Europe, 2013, Bauernschuster et al., 2017), den Immobilienmarkt (Chay und Greenstone, 2005; Xu et al., 2015) oder die Produktivität von Menschen (Zivin und Neidell, 2012; Lichter et al., 2017) auswirken.

In einigen deutschen Städten wurden ab Mitte der 2000er Jahre Umweltzonen eingeführt; Fahrzeugen wurde entsprechend ihrer Emissionsklasse das Fahren innerhalb solcher Zonen untersagt. Die Umweltzonen umfassen in der Regel einen großen Teil des Stadtzentrums. Wolff (2014) vergleicht die Feinstaubdaten von Städten mit und ohne Umweltzone vor und nach Einführung der jeweiligen Umweltzonen in der Zeit zwischen 2005 bis 2008 und schlussfolgert, dass sich die Feinstaubemissionen im bundesweiten Durchschnitt in Städten mit Umweltzone um 9 % stärker verringert haben.<sup>1</sup> Gehrsitz (2017) nutzt Daten von 2005 bis 2012 und kommt zu einer Reduktion von ca. 4 % in Feinstaubwerten und 3,4 % in Stickstoffdioxidwerten in deutschen Städten mit Umweltzone.

Die folgenden Schätzungen ergänzen diese Literatur, weil das Hamburger Durchfahrverbot ausschließlich für bestimmte, letztlich vergleichsweise kurze Straßenabschnitte gilt, statt das Stadtzentrum flächendeckend zu

erfassen. Die Literatur verwendet primär die gleichzeitig auftretende lokale und temporäre Variation in Bezug auf die Einführung von Umweltzonen, um kausale Schlussfolgerungen ziehen zu können. Da für die Hamburger Maßnahme keine simultane zeitliche Variation für die Einführung des Durchfahrverbots besteht, sondern ausschließlich eine lokale<sup>2</sup>, sind kausale Schlussfolgerungen ohne entsprechende Robustheitstests mit Bedacht zu treffen.

### Daten und empirische Strategie

Wir nutzen die Daten der Messnetze des Bundes und der Länder in Form von Schadstoff-Tagesmittelwerten für Stickstoffdioxid und Feinstaub, die vom Umweltbundesamt zur Verfügung gestellt wurden. Es standen die Daten sämtlicher Luftmessstationen Deutschlands zur Verfügung, die von 2003 bis einschließlich 2018 aktiv waren. Tabelle 1 stellt für diesen Zeitraum die durchschnittlichen Stickstoffdioxid- und Feinstaubwerte vor und nach Einführung des Dieselvebots am 31.5.2018 dar, an den beiden direkt betroffenen Messpunkten an der Max-Brauer-Allee und der Stresemannstraße (Versuchsgruppe) sowie an den beiden nicht vom Durchfahrverbot tangierten Hamburger Messstationen (Kontrollgruppe HH). Auch die durchschnittlichen Schadstoffwerte an sämtlichen anderen deutschen verkehrsnahen Luftmessstationen sind dargestellt (Kontrollgruppe DE). Ein einfacher Vergleich ergibt sich durch die Betrachtung der Differenz aus der absoluten Veränderung (vorher/nachher) zwischen Versuchs- und Kontrollgruppe. Ein Vergleich mit sämtlichen verkehrsnahen Messstationen in Deutschland als Kontrollgruppe impliziert einen um 10,67  $\mu\text{g}/\text{m}^3$  (17,3 %) stärkeren Rückgang der Stickstoffdi-

<sup>1</sup> In einem Robustheitstest wird ein potenzieller Effekt mit einer feineren Klassifizierung für Messstationen innerhalb und außerhalb von Umweltzonen statt Städten mit und ohne Umweltzone durchgeführt. In Städten mit Umweltzone findet Wolff keine statistisch signifikanten Effekte für Messstationen außerhalb von Umweltzonen. Dies bedeutet, dass die Verringerung von Emissionen innerhalb einer Umweltzone nicht durch mögliche Erhöhungen außerhalb der Umweltzone, beispielsweise durch ein Umfahren der Zone kompensiert werden.

<sup>2</sup> Die Maßnahme wurde an unterschiedlichen Straßenabschnitten (lokale Variation) zeitgleich (keine simultane temporäre Variation) eingeführt.

**Tabelle 1**  
**Tagesdurchschnittliche Stickstoffdioxid- und Feinstaubwerte vor und nach dem 31.5.2018**

	Kontrollgruppe HH		Kontrollgruppe DE		Versuchsgruppe	
	vorher	nachher	vorher	nachher	vorher	nachher
Stickstoffdioxid (Tagesmittel in µg/m³)	56,971	49,811	43,785	36,036	61,519	43,103
Veränderung in %	-12,57		-17,70		-29,94	
Beobachtungen	11.106	428	721.604	27.339	11.041	426
Feinstaub (Tagesmittel in µg/m³)	29,075	22,751	26,842	20,310	25,765	20,606
Veränderung in %	-21,75		-24,34		-20,02	
Beobachtungen	4.947	212	734.803	25.772	10.824	420

Anmerkungen: Die Kontrollgruppe HH erfasst die verkehrsnahen Messstationen in Hamburg, die nicht von der Fahrzeugschranke betroffen sind (Kieler Straße und Habichtstraße). Die Kontrollgruppe DE erfasst sämtliche verkehrsnahen Messstationen in Deutschland (ohne Hamburger Messstationen). Die Versuchsgruppe erfasst die verkehrsnahen Messstationen in Hamburg, die von der Fahrzeugschranke betroffen sind (Stresemannstraße und Max-Brauer-Allee). Zugrunde gelegt werden Tagesmittelwerte für Stickstoffdioxid und Feinstaub in µg/m³ vom 1.1.2003 bis zum 31.12.2018.

Quelle: eigene Berechnungen auf Basis von: Umweltbundesamt.

oxidwerte an den beiden Hamburger Messstationen in der Versuchsgruppe, nachdem das Durchfahrverbot eingeführt wurde.<sup>3</sup> Beim Feinstaub ergibt ein solcher Vergleich allerdings einen Anstieg von 1,37 µg/m³ (5,3 %). Da Feinstaub beispielsweise auch durch Reifenabrieb entsteht und nicht einzig als Produkt eines Motors auftritt, könnte man in diesem Kontext auf mögliche Effekte der Fahrzeugflotte spekulieren. Falls die Zahl an Fahrzeugen auf den Straßen mit einer erhöhten Feinstaubbelastung einhergehen würde, wären Maßnahmen, die primär auf der Motorisierung eines Fahrzeugs basieren, möglicherweise weniger effektiv.

Solche einfachen Vorher-Nachher-Vergleiche kontrollieren jedoch nicht für gleichzeitig zeitlich und räumlich variierende Faktoren, die möglicherweise einen Einfluss auf die Schadstoffwerte ausüben, sodass verzerrte Ergebnisse der Wirksamkeit des Dieselvebotes drohen. Beispielsweise könnten meteorologische Einflüsse, wie Wind, Temperatur und Regen solche Verzerrungen hervorrufen. Multivariate Regressionsschätzungen können hingegen ein unverzerrtes Ergebnis liefern und nach Robustheitstests eine kausale Argumentation stützen.

Für die entsprechenden multivariaten Analysen orientieren wir uns methodisch zunächst an Wolff (2014) und Gehr-

sitz (2017). Wir verwenden vergleichbare Wetterdaten<sup>4</sup>, die vom Deutschen Wetterdienst (DWD) bereitgestellt werden und von eigenständigen Wettermessstationen gemessen werden. Wir verknüpfen die Luftmessstationen mit den nächstgelegenen Wettermessstationen, um meteorologische Begebenheiten in der Nähe der Luftmessstationen so gut wie möglich abzubilden. Um das Verkehrsaufkommen zu erfassen, werden die Daten der Bundesanstalt für Straßenwesen (BaSt) genutzt. Hierbei werden Stundendaten von automatischen Verkehrszählstationen an Autobahnen und Bundesstraßen auf Tagesebene aggregiert. Es wird der Durchschnitt der fünf nächstgelegenen Verkehrszählstationen jeder Luftmessstation zugeordnet. Potenziellen Verzerrungen des Schätzergebnisses, die auf unterschiedliche Entwicklungen des Verkehrsaufkommens zwischen den Vergleichsgruppen zurückzuführen wären, wird durch die Inklusion dieser Daten vorgebeugt. Der zugrunde liegende Datensatz nutzt die täglichen Informationen vom 1.1.2003 bis 31.12.2018, sodass für jede Luftmessstation bis zu 5.840 Datenpunkte vorhanden sein können, wenn die Luftmessstation jeden Tag aktiv war.

Es wird eine multivariate Differenz-in-Differenzen-Schätzung mit fixen Effekten für die Analyse zugrunde gelegt:

$$(1) \quad y_{it} = \beta_{it}(POST_t \times TREAT_i) + \gamma_{it} X_{it} + \tau_m + \sigma_i + \varepsilon_{it}$$

Dabei entspricht  $y_{it}$  dem gemessenen Tagesmittelwert für Stickstoffdioxid bzw. Feinstaub in µg/m³ an Station  $i$  an Tag  $t$ . Die Variable  $POST_t$  ist ein Indikator der gleich 1 ist für alle Messstationen für jeden Tag ab dem 31.5.2018, also nach Einführung des Durchfahrverbots, und somit das Zeitfenster nach der Maßnahme identifiziert. Die Variable  $TREAT_i$  ist ein Indikator der gleich 1 ist für die beiden vom Durchfahrverbot betroffenen Messstationen in Hamburg für jeden im Datensatz vorhandenen Tag und identifiziert somit die Versuchs- sowie Kontrollgruppe. Der Koeffizient  $\beta_{it}$  der Interaktion beider Indikatoren gibt den gewünschten Differenz-in-Differenzen-Schätzer, der die Abweichung in der Versuchsgruppe nach Einführung der Maßnahme relativ zur Kontrollgruppe zeigt. Ergänzt wird dieser Schätzer durch mehrere Kontrollvariablen, die in der Gleichung durch den Vektor  $X_{it}$  dargestellt sind. Außerdem werden zwei Gruppen von fixen Effekten verwendet. Fixe Effekte für die einzelnen Monate ( $\tau_m$ ) erfassen unbeobachtete temporär variierende Einflüsse, wie beispielsweise volkswirtschaftliche Trends oder politische Maßnahmen auf Bundesebene, wie z. B. die Abwrackprämie. Der fixe Effekt ( $\sigma_i$ ) erfasst unbeobachtete regionale/lokale Einflussgrößen

3 Veränderung Kontrollgruppe DE (43,785-36,036) – Veränderung Versuchsgruppe (61,519-43,103) = 10,667. Prozentuale Veränderung bemessen an durchschnittlichen Messwert vor Einführung der Maßnahme (61,52 µg/m³).

4 Tagesmittel der Temperatur, Tagesmaximum und Tagesminimum der Lufttemperatur in 2 m Höhe, Tagesmittel der relativen Feuchte, Niederschlagshöhe, Windgeschwindigkeit, Tagesmittel Dampfdruck und Luftdruck, Tagesmaximum Windspitze, Tagesmittel Bedeckungsgrad, Indikator Tag mit Niederschlag (ja/nein), tägliche Sonnenscheindauer.

**Tabelle 2**  
**Dieselvebot und relative Veränderungen der Stickstoffdioxid- und Feinstaubmesswerte**

Stickstoffdioxid (in µg/m³)				
	(a)	(b)	(c)	(d)
$POST_i \times TREAT_i$	-10,667*** (1,016)	-6,008*** (0,423)	-7,080*** (0,807)	-4,024*** (0,843)
$(TREND_i - \rho) \times POST_i \times TREAT_i$	-	-	-	-0,028*** (0,012)
$TREND_i$	-	-	0,020*** (0,002)	0,021*** (0,002)
$TREND_i \times TREAT_i$	-	-	0,0004 (0,0003)	-
Beobachtungen	760.410	703.193	703.193	703.193
Bestimmtheitsmaß	0,016	0,696	0,696	0,484
Feinstaub (in µg/m³)				
$POST_i \times TREAT_i$	1,372* (0,780)	3,160*** (0,385)	1,911*** (0,443)	4,705*** (0,592)
$(TREND_i - \rho) \times POST_i \times TREAT_i$	-	-	-	-0,026*** (0,008)
$TREND_i$	-	-	-0,004** (0,002)	-0,005*** (0,002)
$TREND_i \times TREAT_i$	-	-	0,0004* (0,0002)	-
Beobachtungen	771.819	713.390	713.390	713.390
Bestimmtheitsmaß	0,006	0,496	0,496	0,496
Kontrollvariablen	Nein	Ja	Ja	Ja
Fixe Effekte	Nein	Ja	Ja	Ja

Schätzergebnisse linearer Regressionen für Unterschiede der Veränderungen von Stickstoffdioxid- und Feinstaubmesswerten zwischen den vom Dieselvebot betroffenen Hamburger Messstationen und der Kontrollgruppe aller deutscher Messstationen Differenz-in-Differenzen-Schätzer; verschiedene Spezifikationen für einen Datenzeitraum vom 1.1.2003 bis 31.12.2018. Spezifikation (a) verwendet ausschließlich den Differenz-in-Differenzen-Schätzer ohne Kontrollvariablen und fixe Effekte in der Schätzgleichung. Spezifikation (b) fügt Kontrollvariablen (Wetterdaten, Verkehrsaufkommen, Schulferien/Feiertage) sowie fixe Effekte hinzu. Spezifikation (c) kontrolliert für potenziell unterschiedliche Trends zwischen Versuchs- und Kontrollgruppe vor Einführung des Verbots. Spezifikation (d) erfasst zusätzlich Trendveränderungen nach Einführung des Verbots. Robuste Standardfehler sind unterhalb der Schätzergebnisse in Klammern aufgeführt. Signifikanz: \* p<10%, \*\* p<5%, \*\*\* p<1%.

Quelle: eigene Berechnungen, auf Basis von: Umweltbundesamt, Deutscher Wetterdienst, Bundesanstalt für Straßenwesen.

auf Messstation-Ebene  $i$ , wie beispielsweise den Standort einer Messstation, und ist über die gesamte Periode konstant. Mögliche Messunterschiede etwa zwischen hafennahen Messstationen in Hamburg oder Messstationen, die in Tälern liegen, werden hierdurch erfasst. Der Fehlerterm ist in der Gleichung mit  $\varepsilon_{it}$  abgebildet.

**Ergebnisse**

Tabelle 2 zeigt in Spalte (a) zunächst das Ergebnis der Schätzgleichung (1) ohne Kontrollvariablen und ohne fixe

Effekte. Dieser Ansatz entspricht dem zuvor in Tabelle 1 gezeigtem einfachen Vorher-Nachher-Vergleich, weshalb das Schätzergebnis ebenfalls mit den Zahlen aus Tabelle 1 reproduziert werden kann. Die geringen Bestimmtheitsmaße verdeutlichen, dass bei dieser Vorgehensweise relevante Einflussfaktoren nicht einbezogen wurden.

In Spalte (b) werden die Kontrollvariablen und fixen Effekte, wie in Schätzgleichung (1) aufgeführt, hinzugefügt. Der geschätzte Effekt des Hamburger Dieselvebts in Bezug auf Stickstoffdioxid beträgt -6 µg/m³, ist statistisch hochsignifikant und entspricht einem Rückgang von ca. 10% relativ zur Kontrollgruppe, nachdem das Durchfahrverbot eingeführt wurde. Das Schätzergebnis für Feinstaub impliziert hingegen einen Anstieg (3,16 µg/m³ Anstieg in der Versuchsgruppe nach Einführung der Maßnahme, relativ zur Kontrollgruppe) und ist statistisch ebenfalls hochsignifikant. Die in dieser Methodik grundlegende Annahme impliziert eine identische Entwicklung der Schadstoffwerte in der Versuchsgruppe im Vergleich zur Kontrollgruppe, wenn die Maßnahme nicht eingeführt worden wäre. Da das entsprechende Szenario für die Versuchsgruppe per definitionem unbeobachtet bleibt, kann diese Annahme nicht direkt getestet werden. Wir prüfen alternativ, ob die Schadstoffwerte für die Versuchs- sowie Kontrollgruppe vor der Maßnahme unterschiedliche Trends aufweisen, um die Sinnhaftigkeit dieser Annahme zu stützen. Hierzu wird ein linearer Trend in der Schätzgleichung ergänzt, der explizit zwischen Kontroll- und Versuchsgruppe variieren kann:

$$(2) \quad y_{it} = \beta_{1it}(POST_i \times TREAT_i) + \theta_{it}(TREND_i \times TREAT_i) + \delta_i TREND_i + \gamma_{it} X_{it} + \tau_m + \sigma_i + \varepsilon_{it}$$

Dabei stellt der neue Term  $(TREND_i \times TREAT_i)$  die Interaktion aus dem zuvor bekannten Indikator für die Versuchsgruppe  $(TREAT_i)$  und dem linearem Trend  $TREND_i$  dar, der beginnend am 1.1.2003 einem Wert gleich 1 entspricht und für jeden folgenden Tag um 1 erhöht wird. Der Koeffizient  $\theta_{it}$  spiegelt hier die Abweichung vom täglichen linearen Trend in der Versuchsgruppe relativ zur Kontrollgruppe wider. Der Koeffizient  $\delta_i$  gibt den linearen Trend über alle Messstationen in der Kontrollgruppe wieder. In Spalte (c) sind Ergebnisse der Regression (2) abgetragen. Die grundlegende Interpretation des Effekts des Dieselvebts bleibt unverändert. Bei der Interpretation der in (c) dargestellten Trendwerte (signifikant positiv für Stickstoffdioxid und signifikant negativ für Feinstaub) ist zu beachten, dass die Monats-Fixen-Effekte einen erheblichen Teil der in der Realität zu beobachtenden Verringerung der Emissionen (vgl. Abbildung 1) auffangen.<sup>5</sup> Die Trends von Versuchs- und Kontrollgruppe sind nicht signifikant

5 Ohne die fixen Effekte ist der Trend für beide Schadstoffe negativ.

unterschiedlich, sodass die Annahme eines allgemeinen Trends nicht wiederlegt werden kann.<sup>6</sup>

Sowohl Schätzgleichung (1) als auch Schätzgleichung (2) lassen zu, dass es mit der Einführung des Dieselvebots zu einem Level-Shift kommt, d. h. einer einmaligen Veränderung des Niveaus des Trends. Möglich ist aber auch, dass es sowohl zu einem Level- als auch zu einer Trendverschiebung kommt, also zu einer gleichzeitigen Veränderung des Trends. Eine simultane Inklusion einer Trend- und Levelverschiebung spiegelt sich in der Schätzgleichung (3) wider:

$$(3) \quad y_{it} = \beta_{1it}(POST_t \times TREAT_i) + \varphi_{it}(TREND_t - \rho) \times POST_t \times TREAT_i + \delta_t TREND_t + \gamma_{it} X_{it} + \tau_m + \sigma_i + \varepsilon_{it}$$

Dabei stellt  $\rho$  den Zeitpunkt der Einführung des Durchfahrverbots dar. Somit ist der Term  $(TREND_t - \rho) \times POST_t \times TREAT_i$  im Zeitraum vor der Maßnahme gleich null (da  $POST_t = 0$ ) und am Tag der Maßnahme gleich null (da  $(TREND_t - \rho) = 0$ ). Ab dem ersten Tag nach Einführung der Maßnahme steigt er mit jedem weiteren Tag. Das Schätzergebnis (d) impliziert kleinere, aber dennoch signifikante einmalige Niveauperänderungen des Trends bei Stickstoffdioxid, die durch eine signifikante Trendverschiebung um  $-0,028 \mu\text{g}/\text{m}^3$  pro Tag begleitet werden. Bei den Feinstaubwerten kommt es zu einem nunmehr größeren Anstieg des Trendlevels, der jedoch durch eine negative Trendverschiebung begleitet wird.

### Robustheitstests

Die Qualität der Schätzungen aus Tabelle 2 könnte darunter leiden, dass die verglichenen Messstationen nicht hinreichend homogen sind. In den Kontrollgruppen der Spezifikationen (b) bis (d) aus Tabelle 1 sind beispielsweise auch Messstationen in Städten enthalten, die in Städten mit einer Umweltzone stehen. Wenn die Umweltzonen die Schadstoffbelastungen an den jeweiligen Messstationen verringern, verzerren die Schätzer den ökologischen Beitrag des Dieselvebts nach unten. Da in Hamburg keine Umweltzone existiert, könnte es sich anbieten, ausschließlich Stationen als Kontrollgruppe heranzuziehen, die ebenfalls von keiner anderen fahreinschränkenden Maßnahme betroffen sind. Spalte (e) in Tabelle 3 zeigt, dass bei Beschränkung auf eine Kontrollgruppe ohne Umweltzonen (bei ansonsten gleicher Spezifikation wie in Spalte (d) in Tabelle 2) der dem Hamburger Dieselvebot zuzuordnende Effekt stärker wird und ein Rückgang um

6 Der Koeffizient ist für die entsprechende Schätzung in Tabelle 2 Spalte (c) für den Schadstoff Feinstaub auf dem 10 %-Niveau statistisch signifikant. In homogeneren Vergleichsgruppen, die im nächsten Abschnitt diskutiert werden, wird der Schätzer für Feinstaub insignifikant.

Tabelle 3  
Robustheitstests I: Unterschiedliche Gruppen von Kontrollstationen und geschätzte Effekte

Stickstoffdioxid (in $\mu\text{g}/\text{m}^3$ )				
	(e)	(f)	(g)	(h)
$POST_t \times TREAT_i$	-4,398*** (1,066)	-5,614*** (1,279)	-5,509*** (1,326)	-5,645*** (0,971)
$(TREND_t - \rho) \times POST_t \times TREAT_i$	-0,035*** (0,013)	-0,031*** (0,014)	-0,031* (0,016)	0,010 (0,018)
$TREND_t$	0,016*** (0,002)	0,013*** (0,003)	0,016*** (0,004)	0,025** (0,007)
Beobachtungen	385.696	153.896	76.039	22.997
Bestimmtheitsmaß	0,485	0,5	0,493	0,503
Feinstaub (in $\mu\text{g}/\text{m}^3$ )				
$POST_t \times TREAT_i$	5,683*** (0,793)	4,585*** (1,307)	4,378** (1,742)	6,066*** (0,639)
$(TREND_t - \rho) \times POST_t \times TREAT_i$	-0,034*** (0,008)	-0,031*** (0,010)	-0,034** (0,012)	-0,038** (0,009)
$TREND_t$	0,002 (0,002)	-0,004 (0,003)	-0,013*** (0,003)	-0,018* (0,005)
Beobachtungen	367.233	139.765	65.725	16.401
Bestimmtheitsmaß	0,452	0,47	0,48	0,429
Gebiete ohne Umweltzone	Ja	Ja	Ja	Ja
Durchgängig aktive Stationen	Nein	Ja	Ja	Ja
Gebiete mit über 100.000 Einwohner*innen	Nein	Nein	Ja	Ja
Nur Stationen in Hamburg	Nein	Nein	Nein	Ja

Anmerkungen: Schätzergebnisse linearer Regressionen mit Kontrollvariablen und fixen Effekten für Unterschiede der Veränderungen von Stickstoffdioxid- und Feinstaubmesswerten zwischen den vom Dieselvebot betroffenen Hamburger Messstationen und verschiedenen Kontrollgruppen für einen Datenzeitraum vom 1.1.2003 bis 31.12.2018.

Quelle: eigene Berechnungen auf Basis von: Umweltbundesamt, Deutscher Wetterdienst, Bundesanstalt für Straßenwesen.

$-4,4 \mu\text{g}/\text{m}^3$  ( $-7,2\%$ ) für Stickstoffdioxid, für Feinstaub hingegen ein Zuwachs von  $5,7 \mu\text{g}/\text{m}^3$  ( $22,1\%$ ) geschätzt wird. Die negativen Trendveränderungen verschärfen sich auf  $-0,035$  (Stickstoffdioxid) bzw.  $-0,034$  (Feinstaub).

In einem weiteren Robustheitstest werden ausschließlich Luftmessstationen außerhalb von Umweltzonen herangezogen, die von 2003 bis 2018 jährlich an mindestens 250 Tagen aktiv waren (Spalte f). Der geschätzte Effekt des Hamburger Dieselvebts erhöht sich auf  $-5,6 \mu\text{g}/\text{m}^3$  ( $-9,2\%$ ) für Stickstoffdioxid; für Feinstaub ergibt sich ein Anstieg von  $4,5 \mu\text{g}/\text{m}^3$  ( $17,3\%$ ). Die Spezifikation (g) begrenzt die Auswahl an Messstationen weiter und verwendet ausschließlich Daten von Messstationen, die außerdem in Städten mit 100.000 oder mehr Einwohner\*innen stehen. Die Schätzergebnisse ändern sich hinsichtlich der Niveauperänderungen nicht wesentlich gegenüber

Schätzung (f), wohingegen die Trendverschiebung in Spezifikation (g) ungenauer geschätzt wird.

In der Robustheits-Spezifikation (h) werden die Werte der Messstationen an der Max-Brauer-Allee und an der Stresemannstraße ausschließlich mit den beiden anderen verkehrsnahen Messstationen in Hamburg verglichen. Auch in dieser letzten Spezifikation bleibt der Effekt für Stickstoffdioxid bei ca.  $-5,6 \mu\text{g}/\text{m}^3$  (-9%). Das Schätzergebnis für Feinstaub zeigt weiterhin einen positiven Effekt. Da die Zahl der Beobachtungen deutlich geringer ist (Feinstaub wird nur an einer der beiden verkehrsnahen Kontrollmessstationen in Hamburg gemessen), leidet die Schätzung unter geringeren Bestimmtheitsmaßen. Deshalb sollten diese Ergebnisse aus Spalte (h) vorsichtiger interpretiert werden. Insgesamt weisen die Schätzergebnisse (b) bis (h), welche die Zusammensetzung der Kontrollgruppe variieren, eine relativ große Robustheit hinsichtlich der Effektgröße des Dieserverbotes aus. Insgesamt erscheint die Schätzung (g), welche die Hamburger Entwicklung mit derjenigen in anderen Großstädten vergleicht, besonders relevant.

### Variation der Beobachtungszeiträume

Wir hatten bislang die Daten von 2003 bis 2018 verwendet. Dies war der maximale Zeitraum, für den die Daten auch für die Kontrollvariablen vollständig zur Verfügung standen. Die dadurch zugrunde gelegten Zeitfenster vor Dieserverbot (185 Monate) und nach Dieserverbot (sieben Monate) sind unausgewogen. Im Folgenden testen wir mit alternativen Zeitfenstern. Eine Schätzung unter Verwendung der Daten ausschließlich aus 2018, dem Jahr der Einführung des Durchfahrverbots, ergibt eine signifikant stärkere Niveauperiode des Stickstoffdioxids aufgrund des Dieserverbotes, die sich auf rund  $-11 \mu\text{g}/\text{m}^3$  (ca. -18%) erhöht; hinzukommt eine signifikante Trendverschiebung. Bei Ausweitung des Beobachtungszeitraums auf 2017 und 2018 wird der Effekt jedoch statistisch insignifikant. Die Effekte des Dieserverbotes bleiben insignifikant, wenn schrittweise zusätzlich die Daten aus 2016, 2015, 2014 und 2013 herangezogen werden. Erst wenn auch Daten vor 2006 herangezogen werden, wird der Effekt statistisch signifikant. Je länger der Beobachtungszeitraum, desto größer fällt der geschätzte Level-Shift aus.<sup>7</sup> Abbildung 1 verdeutlicht den Grund: In den Jahren bis 2011 waren die Messwerte an den Straßen des Hamburger Dieserverbotes nicht nur besonders hoch, sondern die Rückgänge fielen auch klein aus. Ab 2013 (Stresemannstraße) bzw. 2015 (Max-Brauer-Allee) gingen die Emissionswerte stärker zurück (und zwar ohne Kenntnis

7 Der Schätzer für die Trendveränderungen in der Versuchsgruppe nach der Maßnahme relativ zur Kontrollgruppe zeigt sich hingegen auch hinsichtlich Veränderungen der Panellänge robust signifikant negativ.

Tabelle 4  
Robustheitstest II: Variation des Beobachtungszeitraums und Ergänzung mittels Daten aus 2019

Stickstoffdioxid (in $\mu\text{g}/\text{m}^3$ )				
	(i)	(j)	(k)	(l)
$POST_i \times TREAT_i$	-6,898*** (2,677)	-4,231 (3,447)	-11,668*** (3,796)	-7,336 (4,252)
$(Trend_i - \rho) \times POST_i \times TREAT_i$	-0,006 (0,006)	-0,011 (0,007)	0,004 (0,006)	0,026* (0,010)
$Trend_i$	0,015 (0,014)	0,039** (0,015)	0,037 (0,021)	0,063 (0,032)
Beobachtungen	7.957	12.186	2.916	4.373
Bestimmtheitsmaß	0,541	0,546	0,604	0,602
Feinstaub (in $\mu\text{g}/\text{m}^3$ )				
$POST_i \times TREAT_i$	0,694 (2,058)	-0,035 (1,287)	-0,020 (1,961)	0,110 (1,023)
$(Trend_i - \rho) \times POST_i \times TREAT_i$	0,007 (0,011)	-0,001 (0,007)	-0,004 (0,010)	-0,004 (0,004)
$Trend_i$	-0,082*** (0,019)	-0,038*** (0,001)	-0,016 (0,009)	-0,008 (0,003)
Beobachtungen	6.827	10.444	2.138	3.199
Bestimmtheitsmaß	0,419	0,441	0,453	0,435
Panelanfang	31.5.2017	31.5.2016	31.5.2017	31.5.2016
Panelende	30.5.2019	30.5.2019	30.5.2019	30.5.2019
Verkehrsaufkommen als Kontrollvariable	Nein	Nein	Nein	Nein
Nur Hamburger Messstationen	Nein	Nein	Ja	Ja

Anmerkungen: Schätzergebnisse linearer Regressionen für Unterschiede der Veränderungen von Stickstoffdioxid- und Feinstaubmesswerten zwischen den vom Dieserverbot betroffenen Hamburger Messstationen und Messstationen in Städten mit mehr als 100.000 Einwohner\*innen, verschiedene Beobachtungszeiträume. Robuste Standardfehler sind in Klammern angegeben.

Quelle: eigene Berechnungen auf Basis von: Umweltbundesamt, Deutscher Wetterdienst, Bundesanstalt für Straßenwesen.

eines kommenden Dieserverbotes). Die Einschließung länger zurückliegender Perioden lassen die aktuell niedrigeren Messwerte vergleichsweise „besser“ aussehen.

Die Wahl des „richtigen“ Vorher-Zeitraums ist schwierig. Unter rein statistischen Kriterien wäre ein Panel mit vergleichbar langen Vorher-/Nachher-Zeiträumen zu bevorzugen. Hierbei bietet es sich an, den Zeitpunkt der Intervention für die Bestimmung der Zeitfenster vor und nach der Maßnahme als Ankerpunkt zu verwenden, statt den Beginn/das Ende eines Kalenderjahres.

Entsprechend wird Gleichung (3) unter Hinzuziehung von aktuelleren Datenpunkten aus 2019 geschätzt. Zu bemerken ist, dass mangels Verfügbarkeit die Daten der automatischen Verkehrszählstationen nicht verwendet werden können und dass die Wetterdaten für 2019 noch kei-

nen vollständigen Qualitätsprozess durchlaufen haben.<sup>8</sup> Die Schätzungen sind somit weniger belastbar.

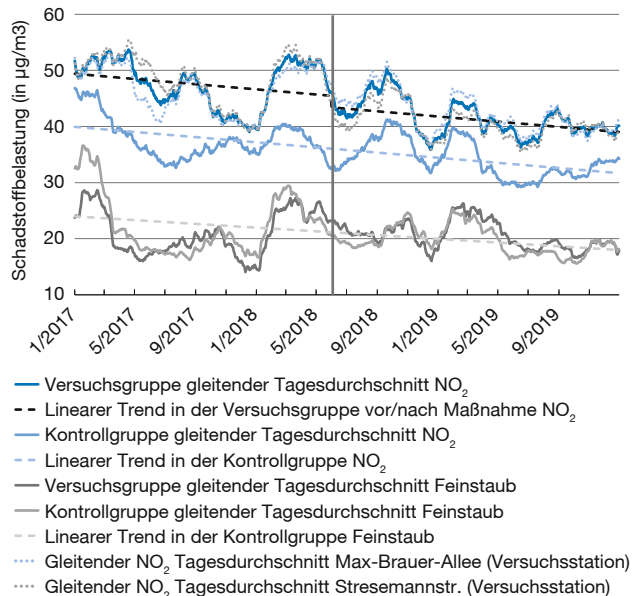
Tabelle 4 zeigt die Ergebnisse auf Basis der präferierten Spezifikationen (g) und (h), jedoch mit den aktuelleren und kürzeren Datenzeiträumen. Das Schätzergebnis aus Spalte (i) zeigt für den Beobachtungszeitraum von 365 Tagen vor und nach der Maßnahme einen signifikanten Level-Shift für Stickstoffdioxid; der Trend-Shift nach der Maßnahme ist nicht signifikant. Für Feinstaub sind die Ergebnisse insgesamt nicht signifikant. Wenn für das identische Beobachtungszeitfenster nur Hamburger Messstationen herangezogen werden (Spalte k), werden die Level-Effekte mit  $-11,7 \mu\text{g}/\text{m}^3$  (-19%) besonders groß. Die Erweiterung des Panels um weitere 365 Tage vor der Maßnahme führt sowohl im Vergleich zur Kontrollgruppe der bundesweit aufgestellten Messstationen (Spalte j) als auch bei der Kontrollgruppe ausschließlich der Hamburger Messstationen (Spalte l) zu keinerlei (politisch erwünschten) negativen Level-Effekten. Der Trendshift-Effekt für Stickstoffdioxid wird gar (politisch unerwünscht) positiv. Auch hier bestätigt sich ein entscheidender Einfluss der Auswahl des Beobachtungszeitraums.

Abbildung 2 illustriert, wie sich Tagesmittelwerte für Stickstoffdioxid von 2017 bis 2019 in der Hamburger Versuchsgruppe und in der Kontrollgruppe entwickelten. Der Übersichtlichkeit wegen ist der Verlauf durch gleitende Durchschnitte für Tagesmittelwerte für beide Gruppen abgebildet. Zur Berechnung des gleitenden Durchschnitts für einen Messwert an Tag  $t$  werden Tagesmittelwerte von jeweils 30 Tagen vor und nach Tag  $t$  herangezogen. Die vertikale graue Linie markiert den Zeitpunkt der Einführung der Maßnahme Ende Mai 2018. Der Level-Unterschied für en zwischen den Hamburger Versuchsmessstationen und den Kontroll-Messstationen in Großstädten über 100.000 Einwohner\*innen ist vor und nach der Maßnahme deutlich erkennbar. Für Feinstaub fanden wir hingegen weder signifikante Level- noch Trendverschiebungen.

Insgesamt halten wir für die von uns präferierte Schätzung mit dem Untersuchungszeitraum vom 31.5.2017 bis zum 30.5.2019 und der bundesweiten Vergleichsgruppe fest, dass die Stickstoffdioxidwerte in den Straßen des Hamburger Dieserverbots im Untersuchungszeitraum um einmalig  $6,89 \mu\text{g}/\text{m}^3$  stärker zurückgegangen sind als in der Vergleichsgruppe. Wir haben bei der Ableitung dieser Ergebnisse eine Reihe von möglichen anderen Einflussfaktoren außer dem Dieserverbot berechnet; dennoch ist es möglich, dass unsere Schätzung aufgrund der Nicht-

<sup>8</sup> Um die Kontrollgruppe weitestgehend unberührt zu lassen, verwenden wir, auch für Datenpunkte in 2019, weiterhin ausschließlich Messstationen, die bis 2018 jährlich an mindestens 250 Tagen aktiv waren

Abbildung 2  
Tagesdurchschnittliche NO<sub>2</sub>- und Feinstaubmesswerte



Anmerkungen: Gleitende Tagesdurchschnitte für Stickstoffdioxid- und Feinstaubmesswerte in Versuchs- und Kontrollgruppe. Zur Kontrollgruppe zählen verkehrsnaher Messstationen in Städten mit mindestens 100.000 Einwohner\*innen ohne Umweltzone, die an mindestens 250 Tagen im Jahr aktiv waren. Die Versuchsgruppe setzt sich aus den zwei von der Maßnahme betroffenen Messstationen in der Stresemannstraße und Max-Brauer-Allee zusammen. Die gepunkteten Linien stellen entsprechende Entwicklung für jeweils nur eine der beiden Versuchsmessstationen dar. Die vertikale graue Linie kennzeichnet die Einführung des Durchfahrverbots.

Quelle: eigene Berechnungen, Umweltbundesamt.

Berücksichtigung von weiteren exogenen Faktoren den Effekt des Dieserverbots überschätzt: So liegen uns beispielsweise keine Daten zu etwaigen unterschiedlichen Flottenmodernisierungen des ÖPNV in Hamburg und in anderen deutschen Großstädten oder beispielsweise unterschiedliche Entwicklungen von alternativen Fortbewegungsmethoden wie Bike-Sharing und E-Bikes vor. Die 2. Fortschreibung des Luftreinhalteplans der Stadt Hamburg umfasst zehn Maßnahmenpakete (Behörde für Umwelt und Energie, 2017), welche die Luftqualität verbessern sollen. Sollten diese Maßnahmen ab Mitte 2018 Wirkungen gezeigt haben, wären unsere Ergebnisse in Richtung einer Überzeichnung der Emissionsverringerung verzerrt.

### Wirtschaftspolitische Bewertung der Ergebnisse

Bereits vor der Einführung des Hamburger Dieserverbots ging die Schadstoffbelastung sowohl an den Hamburger Messstationen als auch bundesweit an verkehrsnahen Messstationen im Trend zurück. Hauptziel dieser Untersuchung war, die Wirkungen des Dieserverbots auf die Schadstoffemissionen anhand der Entwicklung geeigneter anderer Messstationen zu relativieren, um seine Wirkung

von anderen Einflussfaktoren zu isolieren. Die geschätzten Wirkungen nach Einführung des Hamburger Dieselvebots sind von der gewählten Spezifikation abhängig. Dabei ist die Wahl der Vergleichsgruppe weniger einflussreich; entscheidendere Veränderungen ergeben sich bei der Wahl des Untersuchungszeitraums. Schätzungen, die den Untersuchungszeitraum stark ausdehnen, zeigten einen signifikanten Rückgang der Stickstoffdioxid- und einen Anstieg der Feinstaubwerte. Für kürzere Untersuchungszeiträume, welche die Zeitfenster vor und nach der Maßnahme ausgewogener gestalten, ergibt sich ein signifikanter Effekt des Dieselvebots auf die Stickstoffdioxidemissionen erst, wenn Daten aus 2019 verwendet werden. Für Feinstaub ergeben sich in diesem Fall keine signifikanten Effekte.

Diese Untersuchung ist ein Beispiel für die Schwierigkeit von Ex-post-Politikevaluierungen, selbst solchen, die ausschließlich technische Daten betrachten. Schon relativ geringe Änderungen bezüglich der Annahmen der „richtigen“ Vergleichszeiträume und (wenngleich hier weniger ausgeprägt) der Vergleichsgruppen und Kontrollvariablen können zu fundamental unterschiedlichen Ergebnissen führen. Hinzukommt, dass möglicherweise den Emissionsreduktionen in den untersuchten Messstationen in der Max-Brauer-Allee und der Stresemannstraße verschlechterte Emissionswerte in Nachbargebieten gegenüberstehen, etwa weil Dieselfahrer\*innen die gesperrten Zonen umfahren.

Den Veränderungen der Schadstoffemissionen in Hamburg sind aus ökonomischer Sicht die Kosten gegenüberzustellen, etwa in Form der Kosten Hamburger Verkehrsteilnehmender bei der Anschaffung emissionsärmerer Fahrzeuge. Und es wäre zu untersuchen, ob andere Maßnahmen zur Reduktion der Schadstoffe zur Verfügung stehen, die ein günstigeres Kosten-Nutzen-Verhältnis aufweisen. Unter politischen Gesichtspunkten wäre schließlich noch die Verteilung der Kosten und Nutzen auf die verschiedenen Hamburger Bevölkerungsgruppen zu bewerten, auch bei den alternativen Maßnahmen. Eine solche vollständige wirtschaftspolitische Ex-post-Analyse für das Hamburger Dieselvebot liegt bisher nicht vor. Allerdings sei darauf hingewiesen, dass auch 2019 immer noch an drei von vier Hamburger verkehrsnahen Messstationen Stickstoffdioxid-Jahresmittelwerte vorliegen,<sup>9</sup>

9 Die Stickstoffdioxidwerte an der Stresemannstraße weisen ein Jahresmittel von 39,62 µg/m<sup>3</sup> aus.

die über dem Grenzwert von 40 µg/m<sup>3</sup> liegen. Diese kontinuierlich hohe Schadstoffbelastung gilt als gesundheitsschädlich, was mit entsprechend hohen volkswirtschaftlichen Kosten verbunden ist. Die geschätzten Emissionsreduktionen an den Hamburger Versuchsmessstationen relativ zu den Kontrollmessstationen in anderen Großstädten dürften auch unter gesundheitsökonomischen Gesichtspunkten ein Schritt in die richtige Richtung sein, der allerdings nicht genügt und durch weitergehende und/oder andersartige Maßnahmen zu ergänzen wäre, die unter Umständen eine bessere wirtschaftspolitische Kosten/Nutzen-Relation für die Region aufweisen.

## Literatur

- Bauernschuster, S., T. Hener und H. Rainer (2017), When Labor Disputes Bring Cities to a Standstill: The Impact of Public Transit Strikes on Traffic, Accidents, Air Pollution, and Health, *American Economic Journal: Economic Policy*, 9(1), 1-37.
- Economic Policy, 9(1), 1-37. Behörde für Umwelt und Energie (2017), Luftreinhalteplan für Hamburg (2. Fortschreibung).
- Behörde für Umwelt und Energie (2018), Dieseldurchfahrtsbeschränkungen, [https://www.hamburg.de/durchfahrtsbeschaerungen/11067546/dieseldurchfahrtsbeschaerungen-faq/#anker\\_1](https://www.hamburg.de/durchfahrtsbeschaerungen/11067546/dieseldurchfahrtsbeschaerungen-faq/#anker_1) (21. Januar 2020).
- BaSt (Bundesanstalt für Straßenwesen) (2017) [https://www.bast.de/BAST\\_2017/DE/Verkehrstechnik/Fachthemen/v2-verkehrszaehlung/Stundenwerte.html](https://www.bast.de/BAST_2017/DE/Verkehrstechnik/Fachthemen/v2-verkehrszaehlung/Stundenwerte.html).
- DWD (Deutscher Wetterdienst) Climate Data Center (2018), Historische tägliche Stationsbeobachtungen (Temperatur, Druck, Niederschlag, Sonnenscheindauer, etc.) für Deutschland, Version v006
- Chay, K. Y. und M. Greenstone (2005), Does Air Quality Matter? Evidence from the Housing Market, *Journal of Political Economy*, 113(2), 376-424.
- Davis, L. W. (2008), The Effect of Driving Restrictions on Air Quality in Mexico City, *Journal of Political Economy*, 116(1), 38-81.
- Ellison, R. B., S. P. Greaves und D. A. Hensher (2013), Five years of London's low emission zone: Effects on vehicle fleet composition and air quality, *Transportation Research Part D: Transport and Environment*, 23, 25-33.
- Gehrsitz, M. (2017), The effect of low emission zones on air pollution and infant health, *Journal of Environmental Economics and Management*, 83, 121-144.
- Health Organization and Office for Europe (2013), Review of evidence on health aspects of air pollution-REVIHAAP Project Technical Report, <http://www.euro.who.int/pubrequest> (20. Januar 2020).
- Lichter, A., N. Pestel und E. Sommer (2017), Productivity effects of air pollution: Evidence from professional soccer, *Labour Economics*, 48, 54-66.
- Wolff, H. (2014), Keep Your Clunker in the Suburb: Low-emission Zones and Adoption of Green Vehicles, *Economic Journal*, 124(578), 1-16.
- Xu, Y., Q. Zhang und S. Zheng (2015), The rising demand for subway after private driving restriction: Evidence from Beijing's housing market, *Regional Science and Urban Economics*.
- Zivin, J. G. und M. Neidell (2012), The Impact of Pollution on Worker Productivity, *American Economic Review*, 102(7), 3652-3673.

**Title:** *The Hamburg Diesel Ban – An Example of the Difficulties of Policy Evaluation*

**Abstract:** *Hamburg has issued drive-through bans for certain diesel vehicles on sections of heavily trafficked roads with the aim of reducing pollutant emissions. It is difficult to prove the success of this initiative. Relatively small changes in the observed time periods, control groups and control variables can lead to significantly different results. This study is an example of the difficulty of ex post evaluations of policy measures.*

**JEL Classification:** H43, Q53, R48