



Hamburgisches
WeltWirtschafts
Institut

Reihe Edition HWWI Band 6

Onlinedaten und Konsumententscheidungen
Voraussagen anhand von Daten aus Social Media und Suchmaschinen

Deniz Dilan Karaman Örsal

In:

Neuvermessung der Datenökonomie

herausgegeben von Thomas Straubhaar

Seite 157–172

Hamburg University Press
Verlag der Staats- und Universitätsbibliothek Hamburg
Carl von Ossietzky

Impressum

BIBLIOGRAFISCHE INFORMATION DER DEUTSCHEN NATIONALBIBLIOTHEK

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen National- bibliografie; detaillierte bibliografische Daten sind im Internet über <https://portal.dnb.de> abrufbar.

LIZENZ

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Das Werk steht unter der Creative-Commons-Lizenz Namensnennung 4.0 International (CC BY 4.0, <https://creativecommons.org/licenses/by/4.0/legalcode.de>). Ausgenommen von der oben genannten Lizenz sind Teile, Abbildungen und sonstiges Drittmaterial, wenn anders gekennzeichnet.



ISSN 1865-7974

ONLINE-AUSGABE

Die Online-Ausgabe dieses Werkes ist eine Open-Access-Publikation und ist auf den Verlagswebseiten frei verfügbar. Die Deutsche Nationalbibliothek hat die Online-Ausgabe archiviert. Diese ist dauerhaft auf dem Archivserver der Deutschen Nationalbibliothek (<https://portal.dnb.de>) verfügbar.

DOI <https://doi.org/10.15460/HUP.HWWI.6.212>

ISBN

Print: 978-3-943423-91-4

EPUB: 978-3-943423-94-5

SATZ Hamburg University Press

COVERGESTALTUNG Hamburg University Press unter Verwendung eines Fotos von Free-Photos auf Pixabay (<https://pixabay.com/images/id-768432>)

DRUCK UND BINDUNG Books on Demand (Norderstedt)

VERLAG

Hamburg University Press, Verlag der Staats- und Universitätsbibliothek Hamburg
Carl von Ossietzky, Hamburg (Deutschland), 2021
<https://hup.sub.uni-hamburg.de>

Inhalt

- 7 **Vorwort**
- 9 **Einleitung**
Datenwirtschaft: Was ist neu und anders?
Thomas Straubhaar

Teil 1: Das Produktivitätsparadox der Datenökonomie

- 29 **Die digitale Revolution: Der große Übergang in die Datenökonomie**
Henning Vöpel
- 41 **Der Rückgang des Produktivitätsfortschritts: Worum geht es?**
Thomas Straubhaar
- 61 **Das Produktivitätspuzzle – eine kritische Bewertung**
Felix Roth
- 83 **Zwei Rätsel der Produktivität – eine empirische Beobachtung**
Henrique Schneider

Teil 2: Neue empirische Verfahren für die Datenökonomie

- 101 **Nowcast als Forecast**
Neue Verfahren der BIP-Prognose in Echtzeit
Christina Heike Maaß
- 129 **Textdaten**
Anwendungen und Herausforderungen
Silke Sturm

- 157 **Onlinedaten und Konsumententscheidungen**
Voraussagen anhand von Daten aus Social Media und Suchmaschinen
Deniz Dilan Karaman Örsal
- 173 **Implizite Motive in der politischen Kommunikation**
Niklas Scheffer, Silke Sturm und Zahurul Islam
- 199 **Verfasserinnen und Verfasser**

Onlinedaten und Konsumententscheidungen

Voraussagen anhand von Daten aus Social Media und Suchmaschinen

Deniz Dilan Karaman Örsal

Einleitung

Dieser Beitrag beschäftigt sich mit dem Einsatz von Onlinedaten bei der Prognose grundlegender makroökonomischer Indikatoren wie privatem Konsum und Konsumentenvertrauensindikatoren. Dabei werden aktuelle Veröffentlichungen zusammengefasst und neue Anwendungsmöglichkeiten von Onlinedaten aufgezeigt.

Mit der breiten Nutzung des Internets und der rasanten Entwicklung der Digitalisierung stehen in den letzten Jahren neue und große Datenquellen zur Verfügung. Mit Hilfe von Kurznachrichten, überwiegend aus den sozialen Netzwerken (wie zum Beispiel Twitter, Facebook, Instagram, YouTube), oder aggregierten Indizes, die anhand von Suchanfragen konstruiert werden (wie unter anderem Google Trends, Google Insights), werden neue Vorhersagemodelle entwickelt, um wirtschaftliche Indikatoren frühzeitig beziehungsweise in Echtzeit zu prognostizieren.

In der Literatur wird auf einen Zusammenhang zwischen der Stimmung der Konsument:innen und dem privaten Konsum hingewiesen. Da der private Konsum 60 % bis 70 % des Bruttoinlandsprodukts ausmacht, können Veränderungen in der Verbraucherstimmung Veränderungen im privaten Konsum signalisieren.

Die Stimmung der Verbraucher:innen wird im Verlauf von Rezessionen akribisch beobachtet, da jede wesentliche Änderung oder jedes Fehlen einer Änderung als wichtiges Zeichen eines nahe gelegenen Wendepunkts oder einer Verlängerung des Tiefs angesehen wird.¹

¹ Vuchelen (2004).

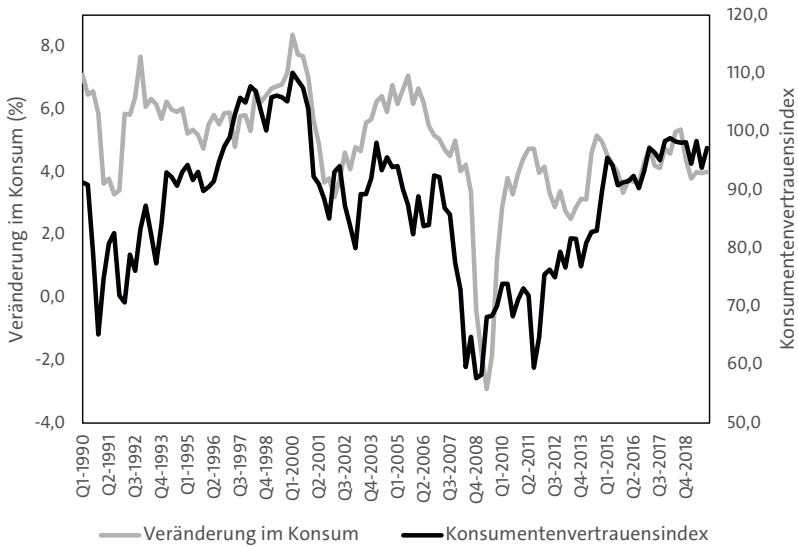


Abb. 1: Veränderungen im privaten Konsum und Konsumentenvertrauensindex in den USA
 Quelle: Eigene Darstellung, basierend auf Federal State Bank of St. Louis, <https://fred.stlouisfed.org/>; University of Michigan <http://www.sca.isr.umich.edu/tables.html>.

Abbildung 1 gibt einen ersten Eindruck davon, wie sich der private Konsum und das Konsumentenvertrauen in den USA zwischen 1990 und 2019 verändert haben und inwieweit die Verläufe der beiden Variablen sich ähneln. Es ist deutlich erkennbar, dass der umfragebasierte Konsumentenvertrauensindex der University of Michigan den enormen Rückgang im privaten Konsum im Zuge der Großen Rezession früher erkennen konnte. Anders ausgedrückt lässt sich schlussfolgern, dass es möglich sein kann, mit der Analyse der Verbraucherstimmung die Veränderungen in den Konsumententscheidungen zu modellieren beziehungsweise vorauszusagen. Da allerdings die umfragebasierten Konsumentenvertrauensindizes verzögert veröffentlicht werden und ihre Ermittlung sehr zeit- und kostenintensiv ist, wurden in den letzten Jahren Daten sowohl aus den Sozialen Medien als auch Anfragen bei den Onlinesuchmaschinen verwendet, um die Stimmung der Konsumenten zeitgenauer zu messen.

Daher werden im nächsten Abschnitt die aktuellsten Studien zusammengefasst, die Daten aus den Sozialen Medien anwenden, um Konsumententscheidungen und Konsumentenvertrauen zu prognostizieren. Der nachfolgende

Abschnitt liefert eine Übersicht über die jüngsten Untersuchungen des Konsumverhaltens, die sich auf Onlinesuchanfragen stützen. Anschließend wird in dem letzten Abschnitt auf die Grenzen und Möglichkeiten der Analyse mit Onlinedaten aufmerksam gemacht und auf die möglichen Herausforderungen hingedeutet.

Soziale Medien und Konsumententscheidungen

Seit einigen Jahren werden Textdaten aus Sozialen Medien, Blogs, und Foren auch in der Erforschung der Konsumententscheidungen verwendet. Dabei werden überwiegend Kurznachrichten aus den Social-Media-Plattformen wie Twitter und Facebook benutzt. Seit 2004 stellt Facebook einen Online-Social-Networking-Dienst zur Verfügung. Mit seinen 2,7 Milliarden Nutzer:innen weltweit² ist Facebook die populärste Social-Networking-Plattform überhaupt. Da aber die Facebook-Posts nur durch von Nutzer:innen zugelassenen Personen gelesen oder kommentiert werden können und für Drittpersonen nicht zur Verfügung stehen, ist es nicht zielführend, ausgehend von den Posts eine allgemeine für die Konsumententscheidungen relevante Meinung ableiten zu wollen. Dies ist nur möglich wenn die Nutzer:innen in Facebook-Gruppen interagieren beziehungsweise zu Werbungen auf der Facebook-Plattform reagieren beziehungsweise Kommentare abgeben. Darüber hinaus ist es auch aus datenschutzrechtlichen Gründen nicht möglich, die Facebook-Nachrichten, die Drittpersonen nicht zur Verfügung stehen, zu analysieren. Daher konzentriert sich dieser Abschnitt auf die Kurznachrichten aus der Social-Media-Plattform Twitter.

Das Social-Networking-Unternehmen Twitter bietet seit 2006 einen Onlinedienst für Mikroblogging. Registrierte Nutzer:innen dürfen auf der Twitter-Plattform kurze Nachrichten (die sogenannten Tweets) senden und mit anderen Nutzer:innen interagieren, wobei die nicht registrierten Nutzer:innen diese Tweets nur lesen dürfen. Die Länge der Tweets ist seit November 2017 auf höchstens 280 Zeichen begrenzt, mit der Ausnahme der Tweets in chinesischer, japanischer und koreanischer Sprache, welche nur aus 140 Zeichen zusammengesetzt sein dürfen. Die 340 Millionen Twitter-Nutzer:innen weltweit senden

² Vgl. Statista, <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>, aufgerufen am 23.12.2020.

täglich durchschnittlich 500 Millionen Tweets.³ Damit ist Twitter weltweit eine der größten Social-Media-Plattformen.

Tag für Tag steigen die Zahl der Nutzer:innen und die Zahl der Tweets an. Social-Media-Plattformen beeinflussen den Alltag in vielfacher Weise und mit steigender Tendenz. Mittlerweile ist Twitter als wichtiges Medium für Politiker:innen, Journalist:innen und Unternehmen unerlässlich geworden. Durch Twitter haben Forscher:innen Zugang zu einer riesigen Datenmenge, mit der sie „wertvolle Einblicke in das Verhalten von Menschen, ihre Stimmungen, ihre Konsummuster, ihre Sprache und ihre Abstimmungen“ bekommen können. Daher wird es wichtiger, notwendiger und erkenntnisreicher denn je, Twitter-Daten zu analysieren.⁴

In letzter Zeit haben Tweets vermehrt Anwendung gefunden, um die führenden Aktienmarktindizes⁵ wie Dow Jones Industrial Average, NASDAQ Composite und S&P 500 oder Rendite, Volatilität und Handelsvolumen⁶ für S&P 500 vorherzusagen. Die Untersuchungsergebnisse haben gezeigt, dass sich die Vorhersagegenauigkeit durch die Einbeziehung der Stimmung in den sozialen Netzwerken verbessert.

Es gibt einige Studien, die sich mit der Prognose der Konsumentenvertrauensindikatoren anhand von Twitter-Kurznachrichten befassen⁷. Die Stimmung der Twitter-Nutzer:innen wird dabei mit den umfragebasierten Konsumentenvertrauensindizes verglichen. Um diese umfragebasierten Konsumentenvertrauensindizes zu konstruieren, wird monatlich eine zufällige Stichprobe von Verbraucher:innen befragt. Dabei beantworten die Interviewten Fragen zu

- ihrer finanziellen Situation in den letzten und in den nächsten zwölf Monaten,
- der allgemeinen wirtschaftlichen Lage in den letzten und in den nächsten zwölf Monaten,
- Preistrends in den letzten und in den nächsten zwölf Monaten,
- ihren geplanten großen Einkäufen derzeit und in den nächsten zwölf Monaten,
- Einsparungen derzeit und in den nächsten zwölf Monaten,
- Ihren Erwartungen der Arbeitslosigkeit in den nächsten 12 Monaten.

³ Statista, online unter <https://de.statista.com/themen/99/twitter/>, zuletzt aufgerufen am 7.7.2021.

⁴ Puschmann et al. (2014).

⁵ Bollen/Mao/Zeng (2011); Zhang (2017); Zhang/Fuehres/Gloor (2011); Oliveira/Cortez/Areal (2017).

⁶ Oliveira et al. (2017).

⁷ Daas/Puts (2014); Shayaa et al. (2018); Shayaa et al. (2017).

Mit Hilfe einer Sentimentanalyse⁸ wird die Stimmung der Konsumenten basierend auf ihren Kurznachrichten in den Sozialen Medien (Tweets, Facebook-Posts) gemessen.

Die Sentimentanalyse, auch Opinion Mining genannt, ist ein Forschungsgebiet, das sich mit der Analyse der Meinungen, Gefühle, Einschätzungen, Einstellungen und Emotionen der Menschen gegenüber Objekten und ihren Attributen, die sie in geschriebenem Text ausdrücken, befasst. Diese Objekte können Produkte, Dienstleistungen, Organisationen, Einzelpersonen, Veranstaltungen oder Themen sein.⁹

In Tabelle 1 werden die jüngsten Studien aus der Literatur zusammengefasst, die Textdaten überwiegend aus Sozialen Medien nutzen, um Konsumentenvertrauen beziehungsweise Konsumententscheidungen vorauszusagen.

Wie Tabelle 1 offenbart, benutzen verschiedene Studien unterschiedliche Such- und Schlagwörter, um die Kurznachrichten beziehungsweise Einträge in den Foren zu „scrapen“ (das heißt, zu schürfen). Wie oben erläutert wurde, wird dabei die Sentimentanalyse eingesetzt. Durch die Korrelationsanalyse wird die Stärke des Zusammenhangs zwischen dem umfragebasierten Konsumentenvertrauensindex und dem Sentiment Score – basierend auf den Sozialen Medien – betrachtet. Ein Vergleich der Ergebnisse der verschiedenen Studien verdeutlicht, dass die jeweils gewählten Schlagwörter bei der Berechnung des Sentiments eine sehr große Rolle spielen. Die Schlagwörter, die mit den Fragen aus den Konsumentenbefragungen hergeleitet werden können, liefern eine höhere Korrelation. Außerdem ist es bei der Berechnung des Sentiments wichtig, wie der Sentiment Score definiert wird. Überwiegend verwendeten die veröffentlichten Studien bisher für die Berechnung des Sentiment-Scores, die Prozentsätze der Wörter mit negativen und positiven Sentiments, die im Textkorpus für die Sentimentanalyse zur Verfügung stehen. Als Alternative ließen sich in Zukunft Sentiment-Wörterbücher einsetzen, solange die Lexika mit den kodierten Werten für die Wörter, wie zum Beispiel Adjektive und Adverbien, zur Verfügung stehen. Diese Adjektive und Adverbien und einige für das Sentiment relevante Nomen und Verben werden dann mit Zahlen bewertet, die im Intervall von $[-1, 1]$ liegen, um die Sentiment-Wörterbücher zu bilden.

⁸ Näheres zur Sentimentanalyse ist in dem Beitrag von Sturm (2021) zu finden.

⁹ Liu (2015).

Tab. 1: Studien mit Textdaten aus Sozialen Medien, Foren und Blogs

AUTOREN	ZIEL	METHODE	SCHLAGWÖRTER/ LAND	ERGEBNISSE
Salampasis/ Paltoglou / Giahanou (2011)	Ansichten der Konsument:innen zu Produkten und Marken analysieren	Sentimentanalyse (lexikonbasiert) – täglicher Prozentsatz der negativen, positiven oder neutralen Stimmung bei den Twitter-Daten	35 verschiedene Lebensmittelprodukte und -marken, wie „McDonalds“, „milk“, „seafood“	Die Stimmung der Konsument:innen gegenüber einem Produkt ändert sich kurzfristig nicht, sofern es keinen triftigen Grund gibt.
Daas / Puts (2014)	Zusammenhang zwischen dem umfrage- und Social-Media-basierten Konsumentenvertrauen in den Niederlanden	Sentimentanalyse, Korrelation, Granger-Kausalitätsanalyse, Kointegrationanalyse	Tweets, Facebook-Posts, Blog-einträge aus den Niederlanden. Schlagwörter: „economy“, „job“, „jobs“, „consumer“, „confidence“, „finance“, „spending“, Personalpronomen auf Niederländisch und Englisch	Textbasiertes Stimmungsmaß ist hoch korreliert und kointegriert mit dem offiziellen Konsumentenvertrauensindex.
O'Connor et al. (2010)	Vergleich von Umfragewerten (für Konsumentenvertrauen und US-Wahlen) mit der Stimmung in Twitter-Daten.	Sentimentanalyse (lexikonbasiert) – positive und negative Werte der Wörter.	Konsumentenvertrauen: „economy“, „job“, „jobs“ Wahlen: „obama, mc-cain“ zwischen 2008 und 2009	Ab Mitte 2009 kann Stimmung aus Textnachrichten das Konsumentenvertrauen besser voraussagen als vorher.
Pekar / Binner (2017)	Modellierung von Konsumausgaben anhand von Kaufabsichten in Social Media	saisonale ARIMA, AdaBoost, Gradientenverstärkung, Word2vec	Tweets mit den Kombinationen von („I“, „we“), („will“, „I“, „be going to“, „be looking to“, „want to“, „wanna“, „gonna“), („buy“, „shop for“, „get oneself“) aus den USA	Verbesserung der Prognosegenauigkeit um 11 % bis 18 %.
Shayaa et al. (2017)	Analyse des Kaufverhaltens der Konsument:innen in Malaysia via Social Media	Sentimentanalyse, überwachtes maschinelles Lernen, Korrelationsanalyse	Kurznachrichten aus Twitter, Foren, Mainstream-Medien, Blogs, Facebook und Onlinekommentare beim Kauf von Computern und Telefonen 2015 bis 2016.	schwache Korrelation zwischen der textbasierten Stimmung und dem umfragebasierten Konsumentenvertrauensindex
Shayaa et al. (2018)	Zusammenhang zwischen dem Konsumentenvertrauensindex in Malaysia und der Stimmung aus Social Media	Sentimentanalyse, Korrelationsanalyse	Tweets von 2015 bis 2019, die die Wörter „car“ oder „holiday“ auf Englisch und Malaysisch beinhalten.	Schwache Korrelation zwischen der Textbasierten und der umfragebasierten Stimmung.
Conrad et al. (2019)	Zusammenhang zwischen dem University of Michigan Index zum Konsumentenvertrauen (ICS) und der Stimmung aus Twitter-Daten	Korrelationsanalyse, Kointegration	Tweets vom 1. Januar 2007 bis zum 27. Juni 2014 aus den USA, die das Wort „jobs“ beinhalten	Die Ergebnisse in (O'Connor u. a. [2010]) können ab 2011 nicht repliziert werden.

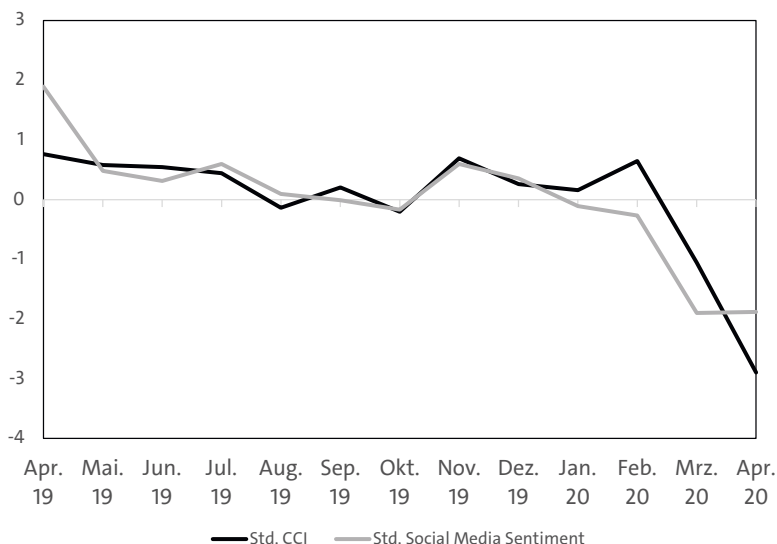


Abb. 2: Standardisierter Konsumentenvertrauensindex (CCI) in Deutschland
 Quelle: <https://stats.oecd.org> und Standardisiertes Sentiment aus Social Media in Deutschland, April 2019 bis April 2020. Quelle: Eigene Berechnungen.

Abbildung 2 zeigt, wie gut die Bestimmung des Konsumentenvertrauens anhand von Kurznachrichten aus Social Media sein kann, wenn das Sentiment mit Hilfe eines Sentiment-Wörterbuchs berechnet wird. Ziele der Analyse, die Abbildung 2 zugrunde liegt, sind die Berechnungen des Social-Media-Sentiments der Konsument:innen und dessen Vergleich mit dem umfragebasierten Konsumentenvertrauensindex in Deutschland. Dabei wurde das Social-Media-Sentiment mit der lexikonbasierten Sentimentanalyse ermittelt. Für die Sentimentanalyse wurden die wirtschaftlich relevanten Kurznachrichten aus Twitter in der Zeit vom April 2019 bis zum April 2020 mittels dem Sentiment-Wörterbuchs SentiWS¹⁰ für die deutsche Sprache analysiert. Insgesamt standen für diesen Zeitraum rund 502.743 Tweets zur Verfügung. Die ersten Ergebnisse haben einen Korrelationskoeffizienten von 0,83 zwischen dem Sentiment aus Social Media und dem umfragebasierten Konsumentenvertrauensindex ergeben. Dieses vielversprechende Ergebnis kann natürlich bekräftigt werden, wenn historische

¹⁰ Remus/Quasthoff/Heyer (2010).

Tweets für Forschungszwecke zur Verfügung stehen. Standardmäßig steht leider momentan nur 1 % der gesamten Tweets kostenlos zur Verfügung, und es ist nicht möglich Tweets die älter als 9 Tage sind. Allerdings gewährt Twitter seit einigen Monaten auf Antrag für Forschungszwecke kostenlos einen vollen Zugang zum Twitter-Archiv. Mit diesem akademischen Zugang ist es möglich, pro Antrag bis zu 10 Millionen Tweets monatlich zu scrapen.

Google Trends und Konsumentenverhalten

Google Trends¹¹ liefert seit 2004 „einen Zeitreihenindex vom Volumen der Suchanfragen, die die Nutzer:innen in einem bestimmten geografischen Gebiet bei der Onlinesuchmaschine Google eingeben“¹². Die Berechnung dieser Zeitreihenindizes erfolgt basierend auf den Suchanfragen je nach Bedarf jährlich, monatlich, wöchentlich und/oder täglich. Der Zeitpunkt mit den meisten Suchanfragen wird dem Wert 100 zugeordnet. Dies bedeutet wiederum, dass sich die historischen Indexwerte mit der Zeit ändern können.

Bei Google Trends werden für 605 Hauptkategorien¹³ und insgesamt für 1426 Unterkategorien aggregierte Indizes, die auf den Suchanfragen basieren, veröffentlicht. Allerdings kann es vorkommen, dass nicht für jede Kategorie ein Index in einem geografischen Gebiet kalkuliert wird, da in einigen Kategorien Suchanfragen fehlen können. Darüber hinaus ist zu berücksichtigen, dass die aggregierten Indizes ohne Bereinigung der saisonalen Schwankungen berechnet werden.

Da in den letzten Jahren die Konsument:innen zunehmend ihre Einkäufe und Preisvergleiche zunehmend via Internet durchführen, können Suchanfragen für Waren und Dienste bei den Suchmaschinen eine alternative Datenquelle werden, um private Konsumausgaben zu prognostizieren.

¹¹ Vgl. <https://trends.google.com/trends/>.

¹² Choi/Varian (2012).

¹³ Hauptkategorien sind Unter anderem Autos und Fahrzeuge, Beruf und Ausbildung, Bücher und Literatur, Computer und Elektronik, Essen und Trinken, Finanzen, Gesetz und Regierung, Gesundheit, Haus und Garten, Haustiere und wild lebende Tiere, Hobby und Freizeitbeschäftigungen, Immobilien, Internet und Kommunikation, Kunst und Unterhaltung, Mensch und Gesellschaft, Nachrichten, Naturwissenschaften, Online-Communitys, Referenz, Reisen, Schönheit und Fitness, Shopping, Spiele, Sport, Unternehmen und Industrie.

Neben der Erstellung von Prognosen können Daten aus Google Trends auch helfen, die Gegenwart besser zu verstehen.¹⁴ Zum Beispiel ist es mit Hilfe der Daten aus Suchmaschinen möglich, unter anderem Influenzawellen¹⁵, Inflationserwartungen¹⁶, Arbeitslosenquote¹⁷ beziehungsweise Anträge auf Arbeitslosigkeit¹⁸, Autoverkäufe¹⁹, Hausverkäufe²⁰, Zahl der Museumsbesucher²¹, Hauspreise²², Einnahmen von Kinofilmen²³ und Bruttoinlandsprodukt²⁴ zu prognostizieren.

Im vorliegenden Beitrag liegt das Hauptaugenmerk auf dem Einsatz von Google-Trends-Daten um Konsumentenvertrauen- und Entscheidungen sowie privaten Konsum vorauszusagen. Um den privaten Konsum in den USA zu prognostizieren,²⁵ wurden 56 wirtschaftlich relevante Google-Trends-Kategorien herauskristallisiert. Die Vorhersagemodelle basieren dabei auf der Erkenntnis, dass der Konsumentenvertrauensindex den privaten Konsum in den USA erklären kann.²⁶ Anhand einer Faktorenanalyse wird die Zahl der Kategorien reduziert und die geschätzten Faktoren fließen anschließend in die Vorhersagemodelle ein. Die Ergebnisse zeigen, dass die Modelle, in denen die Faktoren aus den Google-Trends-Daten als erklärende Variable zusammen mit den für die Konjunktur relevanten makroökonomischen Variablen (das heißt reales persönliches Einkommen, S&P-500-Aktienmarktindex, Dreimonatszinssatz) auftauchen, bessere Prognosen liefern als die Baseline-Modelle. Hier muss auch darauf hingewiesen werden, dass die Studie den Zeitraum zwischen 2005 und 2009 betrachtet hat.

¹⁴ Choi/Varian (2009b); Choi/Varian (2009a), (2009b), (2012).

¹⁵ Ginsberg et al. (2009).

¹⁶ Guzmán (2011).

¹⁷ D'Amuri/Marcucci (2009); Askitas/Zimmermann (2009).

¹⁸ Choi/Varian (2009a).

¹⁹ Choi/Varian (2009b); Choi/Varian (2012); Carrière-Swallow/Labbé (2013).

²⁰ Choi/Varian (2009b); Choi/Varian (2012).

²¹ Volchek et al. (2018).

²² Wu/Brynjolfsson (2015).

²³ Asur/Huberman (2010).

²⁴ Liu/Xu/Fan (2018); Qian (2018).

²⁵ Vosen/Schmidt (2011).

²⁶ Wilcox (2007), Ludvigson (2004); Carroll/Fuhrer/Wilcox (1994).

In einer anderen Studie zur Prognose des privaten Konsums in den USA wurden neben den finanziellen Variablen wie S&P-500-Aktienmarktindex oder kurzfristige (Dreimonats) und langfristige (Zehnjahres)Zinssätze lediglich 25 Google-Trends-Kategorien benutzt.²⁷ Obwohl diese Studie nur den Zeitraum zwischen 2007 und 2009 analysiert, haben die Ergebnisse mit den Faktoren aus den Google-Trends-Kategorien wieder eine bessere Voraussage geliefert.

Ähnliche Ergebnisse mit Google-Trends-Daten wurden in einer anderen Studie auch für den Konsum in Deutschland erzielt.²⁸ Dabei wurden lediglich 41 Google-Trends-Kategorien benutzt und die Daten aus der Zeit von 2004 bis 2010 analysiert. Basierend auf diesen Erkenntnissen wird beim RWI – Leibniz-Institut für Wirtschaftsforschung der RWI-Konsumindikator²⁹ berechnet. Dem Institut zufolge weist der „RWI-Konsumindikator gegenüber dem privaten Verbrauch einen Vorlauf von einem Quartal“³⁰ auf.

In einer aktuelleren Studie³¹ fungierten 66 konsumrelevante Google-Trends-Kategorien zusammen mit den nachrichtenrelevanten Google-Trends-Daten³² als erklärende Variablen, um die Veränderung des privaten Konsums in den USA zu prognostizieren beziehungsweise in Echtzeit vorauszusagen (Nowcast³³). Dabei wurden „recession“ und „layoff“ als Suchwörter für die nachrichtenrelevanten Google-Trends-Daten benutzt. Die Analyse, basierend auf den monatlichen Daten für den Zeitraum zwischen 2009 und 2014, hat gezeigt, dass die nachrichtenrelevanten Google-Trends-Daten für die Veränderung des Konsums von Gebrauchsgütern nützlich sein können und die Modelle mit den konsumrelevanten Google-Trends-Kategorien Konsumabsichten erklären können. Darüber hinaus können konsum- und nachrichtenrelevante³⁴ Google-Trends-Daten zusammen die Prognosegüte verbessern und dabei helfen, den Konsum einen Monat im Voraus zu prognostizieren.

In einer Studie mit zwei Google-Trends-Kategorien (das heißt „motorcycle“, „birth control“) ließ sich die Voraussage des Konsumentenvertrauensindex

²⁷ Kholodilin/Podstawski/Siliverstovs (2010).

²⁸ Vosen/Schmidt (2012).

²⁹ RWI (2010), S.103. Online abrufbar unter <https://www.rwi-essen.de/forschung-und-beratung/wachstum-konjunktur-oeffentliche-finanzen/projekte/221/>.

³⁰ RWI-Konsumindikator, <https://www.rwi-essen.de/konsumindikator>

³¹ Woo/Owen (2019).

³² Google Trends, https://trends.google.de/trends/explore?cat=16&date=all_2008&gprop=news.

³³ Näheres zu „Nowcasts“ ist in dem Beitrag von Maaß (2021) zu finden.

³⁴ Google Nwes, <https://news.google.com>.

in Sri Lanka im Vergleich mit einem autoregressiven Prozess erster Ordnung um 5,2 % verbessern.³⁵ Diese zwei Google-Trends-Kategorien wurden mit Hilfe der Spike-und-Slab-Regression unter Berücksichtigung der insgesamt 1104 Google-Trends-Kategorien für die Zeit vom September 2012 bis zum April 2017 bestimmt.

Google-Trends-Daten werden auch verwendet, um Konsumentenentscheidungen und -verhalten für Mode vorauszusagen. In einer Studie wird zum Beispiel Google-Trends-Serie für die Suchanfrage „Burberry“ mit verschiedenen Modellen geschätzt.³⁶ Die Prognosegüte des bereinigten („enoised“) neuronalen Netzwerk-Modells (DNNAR) wird dabei mit den Prognosen, basierend auf den konventionellen Zeitreihenmodellen wie ARIMA, exponentielle Glättung oder ARMA mit saisonalen Trends, verglichen. Der Vergleich ergibt, dass das DNNAR-Modell bessere Ergebnisse liefert als die anderen Modelle. Allerdings weisen die Autoren darauf hin, dass die Google-Trends-Daten für die moderelevanten Suchanfragen hohe saisonale Schwankungen zeigen, wodurch die Prognosen für die moderelevanten Suchanfragen mit wenigen saisonalen Schwankungen schwer vorherzusagen wären.

Schlussfolgerungen

Textdaten aus Sozialen Medien bergen große Chancen und Herausforderungen für die Praxis. Die ersten vielversprechenden Ergebnisse zeigen, dass die Sentiment Scores, die auf der Basis von Kurznachrichten aus Sozialen Medien berechnet worden sind, in der Zukunft die umfragebasierten Konsumentenvertrauensindizes ersetzen können. Da die Durchführung und Analyse von Umfragen viel Zeit und Geld kostet, können durch die Nutzung von Textdaten aus Sozialen Medien viel Zeit und Ressourcen gespart werden. Eine der größten Herausforderungen dabei ist die Auswahl der richtigen Such- und Schlagwörter für die Praxis. Eine mögliche Lösung dafür könnte der Einsatz von Algorithmen beziehungsweise künstlicher Intelligenz sein. Dabei müssen die Algorithmen mit Daten zunächst trainiert werden. Mit der Zeit können die Algorithmen durch den Lerneffekt verbesserte Prognosen liefern.

Die Nützlichkeit von Google-Trends-Daten unter anderem zu Voraussage von Konsumententscheidungen hat sich schon längst erwiesen. Dennoch könnten

³⁵ Zhang (2017).

³⁶ Silva et al. (2019).

die Prognosen verbessert werden, indem die Daten von Onlinesuchanfragen und Textdaten aus Sozialen Medien zusammen in die Analyse einfließen. Es gibt schon einige erste Versuche in dieser Richtung.³⁷ Um beispielsweise die Autoverkäufe in Deutschland vorauszusagen, wird das Suchverhalten der Verbraucher:innen mit der Stimmung der Verbraucher:innen in Bezug auf eine Marke oder ein Produkt kombiniert.³⁸ Die Ergebnisse der Analyse haben gezeigt, dass die Modelle mit Google-Trends- und Twitter-Daten bessere Prognosen liefern als die Modelle, die nur Suchanfragedaten oder Textdaten aus Sozialen Medien verwenden.

Es muss hier auch noch einmal betont werden, dass Google-Trends-Daten sich mit der Zeit ändern können – mit entsprechenden Rückwirkungen auf die Berechnung von Indexwerten für einen bestimmten Zeitpunkt. Dies wiederum macht die Replizierbarkeit der Studien schwierig. Darüber hinaus wurden in den Studien wegen der hohen Saisonalität, die Google-Trends-Daten in Wachstumsraten umgewandelt, um saisonale Schwankungen zu eliminieren. Da mittlerweile längere Zeitreihen zur Verfügung stehen, wird es bald möglich sein, Google-Trends-Daten mit den konventionellen Methoden wie X12-ARIMA oder X-13-ARIMA-SEATS saisonal zu bereinigen. Dies wird eine einfachere Interpretation der Ergebnisse ermöglichen.

Die Wahl der richtigen Schlagwörter und Kategorien ist für die Analyse mit Google-Trends-Daten sehr wichtig. In einigen Studien wurde vorgeschlagen, Google Correlate zu nutzen, um die richtigen Schlagwörter zu finden.³⁹ Bei Google Correlate ist es möglich, anhand einer Zeitreihe einige relevante Google-Trends-Kategorien oder Unterkategorien zu bekommen. Diese vorgeschlagenen Google-Trends-Kategorien haben schlussendlich einen ähnlichen Verlauf wie die Zeitreihe, die analysiert werden soll. Allerdings können sich dabei Suchwörter ergeben, zu denen lediglich ein begrenzter ein Zusammenhang hergestellt werden kann. Daher ist es wichtig, die Ergebnisse herauszufiltern und mit Vorsicht zu interpretieren.

Eine weitere Möglichkeit wäre die Nutzung von Google-Analytics-Daten, die keine Indexdaten sind. Mit deren Hilfe wäre es möglich, replizierbare Analysen durchzuführen.

³⁷ Skodda/Benthaus (2015).

³⁸ Ebenda.

³⁹ Sawaengsuksant (2018).

Literatur

- Askitas, Nikolaos und Zimmermann, Klaus F. (2009): Google Econometrics and Unemployment Forecasting. In: *Applied Economics Quarterly*, Vol. 55 (2), S. 107–120, DOI <https://doi.org/10.3790/aeq.55.2.107>.
- Asur, Sitaram und Huberman, Bernardo A. (2010): Predicting the Future with Social Media. In: Hoeber, Orland (Hrsg.): *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, S. 492–499, DOI <https://doi.org/10.1109/WI-IAT.2010.63>.
- Bollen, Johan; Mao, Huina; Zeng, Xiaojun (2011): Twitter Mood Predicts the Stock Market. In: *Journal of Computational Science*, Vol. 2 (1), S. 1–8, DOI <https://doi.org/10.1016/j.jocs.2010.12.007>.
- Carrière-Swallow, Yan und Labbé, Felipe (2013): Nowcasting with Google Trends in an Emerging Market. In: *Journal of Forecasting*, Vol. 32 (4), S. 289–298, DOI <https://doi.org/10.1002/for.1252>.
- Carroll, Christopher D.; Fuhrer, Jeffrey C.; Wilcox, David (1994): Does Consumer Sentiment Forecast Household Spending? If so, why? In: *The American Economic Review*, Vol. 84 (5), S. 1397–1408. Online abrufbar unter <https://www.jstor.org/stable/2117779>.
- Choi, Hyunyoung und Varian, Hal (2009a): Predicting Initial Claims for Unemployment Insurance Using Google Trends. Technical report, Google, Inc. Online abrufbar unter <http://research.google.com/archive/papers/initialclaimsUS.pdf>.
- Choi, Hyunyoung und Varian, Hal (2009b): Predicting the Present with Google Trends. Technical Report, Google, Inc., DOI <https://doi.org/10.2139/ssrn.1659302>.
- Choi, Hyunyoung und Varian, Hal (2012): Predicting the Present with Google Trends. In: *Economic Record*, Vol. 88 (1), S. 2–9, DOI <https://doi.org/10.1111/j.1475-4932.2012.00809.x>.
- Conrad, Frederick G.; Gagnon-Bartsch, Johann A.; Ferg, Robyn A.; Schober, Michael F.; Pasek, Josh; Hou, Elizabeth (2019): Social Media as an Alternative to Surveys of Opinions About the Economy. In: *Social Science Computer Review*, DOI <https://doi.org/10.1177/0894439319875692>.
- D’Amuri, Francesco und Marcucci, Juri (2009): „Google it!“ Forecasting the US unemployment rate with a Google job search index. MPRA Paper No.18248, DOI <http://dx.doi.org/10.2139/ssrn.1594132>.
- Daas, Piet J. H. und Puts, Marco (2014): Social Media Sentiment and Consumer Confidence. *Statistics Papers Series*, No. 5, Europäische Zentralbank, Frankfurt am Main, DOI <https://doi.org/10.2866/11606>.

- Ginsberg, Jeremy; Mohebbi, Matthew H.; Patel, Rajan S.; Brammer, Lynnette; Smolinski, Mark S.; Brilliant, Larry (2009): Influenza Epidemics Using Search Engine Query Data. In: *Nature*, Vol. 457 (7232), S. 1012–1014, DOI <https://doi.org/10.1038/nature07634>.
- Guzmán, Giselle (2011): Internet Search Behaviour as an Economic Forecasting Tool: The Case of Inflation Expectations. In: *Journal of Economic and Social Measurement*, Vol. 36 (3), S. 119–167, DOI <https://doi.org/10.3233/JEM-2011-0342>.
- Kholodilin, Konstantin A.; Podstawski, Maximilian; Siliverstovs, Boriss (2010): Do Google Searches Help in Nowcasting Private Consumption? A Real-Time Evidence for the US. In: KOF Swiss Economic Institute Working Paper No. 256, ETH Zürich, DOI <http://dx.doi.org/10.2139/ssrn.1615453>.
- Liu, Bing (2016): *Sentiment Analysis. Mining, Opinions, Sentiments, and Emotions*. Cambridge University Press, Cambridge, UK, DOI <https://doi.org/10.1017/CBO9781139084789>.
- Liu, Taoxiong; Xu, Xiaofei; Fan, Fangda (2018): Forecasting Chinese GDP Using Online Data. In: *Emerging Markets Finance and Trade*, Vol. 54 (4), S. 733–746, DOI <https://doi.org/10.1080/1540496X.2016.1216841>.
- Ludvigson, Sydney C. (2004): Consumer Confidence and Consumer Spending. In: *Journal of Economic Perspectives*, Vol. 18 (2), S. 29–50, DOI <https://doi.org/10.1257/0895330041371222>.
- Maaß, Christina (2021): Nowcast als Forecast. Neue Verfahren der BIP Prognose in Echtzeit. In: Straubhaar, Thomas (Hrsg.): *Neuvermessung der Datenökonomie*, Hamburg University Press, Hamburg, S. 96.
- O'Connor, Brendan T.; Balasubramanyan, Ramnath; Routledge, Bryan R.; Smith, Noah A. (2010): From Tweets to Pools: Linking Text Sentiment to Public Opinion Time Series. In: AAI Publications, Fourth International Conference on Weblogs and Social Media, Palo Alto.
- Oliveira, Nuno; Cortez, Paulo; Areal, Nelson (2017): The Impact of Microblogging Data for Stock Market Prediction: Using Twitter to Predict Returns, Volatility, Trading Volume and Survey Sentiment Indices. In: *Expert Systems with Applications*, Vol. 73, S. 125–144, DOI <https://doi.org/10.1016/j.eswa.2016.12.036>.
- Pekar, Viktor und Binner, Jane (2017): Forecasting Consumer Spending from Purchase Intentions Expressed on Social Media. In: *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Association for Computational Linguistics, Stroudsburg, PA, DOI <https://doi.org/10.18653/v1/W17-5212>.

- Puschmann, Cornelius; Bruns, Axel; Mahrt, Merja; Weller, Katrin; Burgess, Jean (2014): Epilogue Why Study Twitter? In: Weller, Katrin; Bruns, Axel; Burgess, Jean; Mahrt, Merja; Puschmann, Cornelius (Hrsg.): *Twitter and Society*. Peter Lang, New York, S. 425–432, DOI <https://doi.org/10.3726/978-1-4539-1170-9>.
- Qian, Eric (2018): Nowcasting Indian GDP with Google Search Data. Online abrufbar unter <https://core.ac.uk/reader/210594347>.
- Remus, Robert; Quasthoff, Uwe; Heyer, Gerhard (2010): SentiWS – A Publicly Available German-language Resource for Sentiment Analysis. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta.
- RWI (2010): Analyse und Prognose des Spar- und Konsumverhaltens privater Haushalte. RWI Projektberichte, URN http://nbn-resolving.de/urn:nbn:de:hbz:061:3-15339_<zuletzt_aufgerufen_am_3.4.2021>.
- Salampasis, Michail; Paltoglou, Giorgos; Giahanou, Anastasia (2011): Using Social Media for Continuous Monitoring and Mining of Consumer Behaviour. In: Salampasis, Michail; Matopoulos, Aristides (Hrsg.): *Proceedings of the International Conference on Information and Communication Technologies for Sustainable Agri-production and Environment (HAICTA 2011)*, S. 192–202, DOI <https://doi.org/10.1504/IJEB.2014.057905>.
- Sawaengsuksant, Paphatsorn (2018): Standardized Approach in Developing Economic Indicators using Internet Searching Applications. IFC – Bank Indonesia International Workshop and Seminar on „Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data“.
- Shayaa, Shahid; Ainin, Sulaiman; Jaafar, Noor Ismawati; Zakaria, Shamsul Bahri; Phong, Seuk Wai; Yeong, Wai Chung; Al-Garadi, Mohammed Ali; Muhammad, Ashraf; Zahid Piprani, Arsalan (2018): Linking Consumer Confidence Index and Social Media Sentiment Analysis. In: *Cogent Business & Management*, Vol. 5 (1), S. 1–12, DOI <https://doi.org/10.1080/23311975.2018.1509424>.
- Shayaa, Shahid; Al-Garadi, Mohammed Ali; Piprani, Arsalan Zahid; Ashraf, Muhammad; Sulaiman, Ainin (2017): Social Media Sentiment Analysis of Consumer Purchasing Behavior vs Consumer Confidence Index. In: *Proceedings of the International Conference on Big Data and Internet of Thing – BDIOT2017*. New York, 2017, S. 32–35, DOI <https://doi.org/10.1145/3175684.3175712>.
- Silva, Emmanuel; Hassani, Hossein; Madsen, Dag; Gee, Liz (2019): Googling Fashion: Forecasting Fashion Consumer Behaviour Using Google Trends. In: *Social Sciences*, Vol. 8 (4), S. 111, DOI <https://doi.org/10.3390/socsci8040111>.

- Skodda, Christoph und Benthaus, Janek (2015): Investigating Consumer Information Search Behavior and Consumer Emotions to Improve Sales Forecasting. In: Proceedings of the 21st American Conference on Information Systems (AMCIS).
- Sturm, Silke (2021): Textdaten. Anwendungen und Herausforderungen. In: Straubhaar, Thomas (Hrsg.): Neuvermessung der Datenökonomie, Hamburg University Press, Hamburg, S. 126.
- Volchek, E.; Song, H.; Law, R.; Buhalis, D. (2018): Forecasting London Museum Visitors Using Google Trends Data. In: e-Review of Tourism Research. Online abrufbar unter <https://journals.tdl.org/ertr/index.php/ertr/article/view/121>.
- Vosen, Simeon und Schmidt, Torsten (2012): A Monthly Consumption Indicator for Germany based on Internet Search Query Data. In: Applied Economics Letters, Vol. 19 (7), S. 683–687, DOI <https://doi.org/10.1080/13504851.2011.595673>.
- Vosen, Simeon und Schmidt, Torsten (2011): Forecasting Private Consumption: Survey-based Indicators vs. Google Trends. In: Journal of Forecasting, Vol. 30 (6), S. 565–578, DOI <https://doi.org/10.1002/for.1213>.
- Vuchelen, Jef (2004): Consumer Sentiment and Macroeconomic Forecasts. In: Journal of Economic Psychology, Vol. 25 (4), S. 493–506, DOI [https://doi.org/10.1016/S0167-4870\(03\)00031-X](https://doi.org/10.1016/S0167-4870(03)00031-X).
- Wilcox, James A. (2007): Forecasting Components of Consumption with Components of Consumer Sentiment. In: Business Economics, Vol. 42 (4), S. 22–32, DOI <https://doi.org/10.2145/20070403>.
- Woo, Jaemin und Owen, Ann L. (2019): Forecasting Private Consumption with Google Trends Data. In: Journal of Forecasting, Vol. 38 (2), S. 81–91, DOI <https://doi.org/10.1002/for.2559>.
- Wu, Lynn und Brynjolfsson, Erik (2015): The Future of Prediction. In: Goldfarb, Avi; Greenstein, Shane M.; Tucker, Catherine E. (Hrsg.): Economic Analysis of the Digital Economy. National Bureau of Economic Research, Cambridge, MA, S. 89–118, DOI <https://doi.org/10.7208/9780226206981>.
- Zhang, Haiyang (2017): Incorporating Google Trends Data in Predicting Consumer Confidence in Sri Lanka. Online abrufbar unter <https://ssrn.com/abstract=3085010>.
- Zhang, Xue; Fuehres, Hauke; Gloor, Peter A. (2011): Predicting Stock Market Indicators Through Twitter „I hope it is not as bad as I fear“. In: Procedia – Social and Behavioral Sciences, Vol. 26, S. 55–62, DOI <https://doi.org/10.1016/j.sbspro.2011.10.562>.

Verfasserinnen und Verfasser

ISLAM, ZAHURUL, Professor an der NORDAKADEMIE Hochschule der Wirtschaft

KARAMAN ÖRSAL, DENIZ DILAN, Dr. rer. pol. (Humboldt-Universität zu Berlin), Universität Hamburg und außerplanmäßige Professorin an der Leuphana Universität Lüneburg.

MAASS, CHRISTINA HEIKE, M. Sc. in Economics, Universität Hamburg

ROTH, FELIX, Privatdozent für Volkswirtschaftslehre an der Universität Hamburg und Leiter des Projekts GLOBALINTO im Rahmen der Horizon-2020-Forschungsförderung der Europäischen Kommission

SCHEFFER, NIKLAS, cand. rer. pol. (Universität Potsdam), Universität Hamburg, Institut für Computer Aided Psychometric Text Analysis (CAPTA)

SCHNEIDER, HENRIQUE, Professor für Volkswirtschaftslehre an der Nordakademie, Hochschule der Wirtschaft, in Elmshorn und stellvertretender Direktor des Schweizerischen Gewerbeverbands sgv in Bern, Schweiz

STRAUBHAAR, THOMAS, Professor für Volkswirtschaftslehre, insbesondere Internationale Wirtschaftsbeziehungen der Universität Hamburg

STURM, SILKE, M. Sc (Universität Bayreuth), Universität Hamburg

VÖPEL, HENNING, Hamburgisches WeltWirtschaftsinstitut (HWWI) und Professor der Hamburg School of Business Administration (HSBA)



Hamburgisches
WeltWirtschafts
Institut

Reihe Edition HWWI

herausgegeben von Thomas Straubhaar

In der Edition HWWI (ISSN 1865-7974) erscheinen abgeschlossene, umfangreiche Projektergebnisse sowie Dissertationen zu Forschungsthemen, die vom HWWI bearbeitet werden. Folgende Titel sind bisher erschienen:

- Band 1: Thomas Straubhaar (Hg.): Bedingungsloses Grundeinkommen und Solidarisches Bürgergeld – mehr als sozialutopische Konzepte, 2008.
ISBN 978-3-937816-47-0, DOI <https://doi.org/10.15460/HUP.HWWI.1.69>.
- Band 2: Martin-Peter Büch et al. (Hg.): Sportfinanzierung – Spannungen zwischen Markt und Staat, 2009.
ISBN 978-3-937816-53-1, DOI <https://doi.org/10.15460/HUP.HWWI.2.70>.
- Band 3: Martin-Peter Büch et al. (Hg.): Zur Ökonomik von Spitzenleistungen im internationalen Sport, 2012.
ISBN 978-3-937816-87-6, DOI <https://doi.org/10.15460/HUP.HWWI.3.122>.
- Band 4: Martin-Peter Büch et al. (Hg.): Sport und Sportgroßveranstaltungen in Europa – zwischen Zentralstaat und Regionen, 2012.
ISBN 978-3-937816-88-3, DOI <https://doi.org/10.15460/HUP.HWWI.4.123>.
- Band 5: Seçil Paçacı Elitok, Thomas Straubhaar (eds.): Turkey, Migration and the EU: Potentials, Challenges and Opportunities, 2012.
ISBN 978-3-937816-94-4, DOI <https://doi.org/10.15460/HUP.HWWI.5.118>.
- Band 6: Thomas Straubhaar (Hg.): Neuvermessung der Datenökonomie, 2021.
ISBN (Print) 978-3-943423-91-4, (Epub) 978-3-943423-94-5,
DOI <https://doi.org/10.15460/HUP.HWWI.6.212>.

Die Online-Ausgaben der Reihe sind frei zugänglich als Open-Access-Publikation erschienen. Die Printversion kann über den Buchhandel oder direkt beim Verlag (<https://hup.sub.uni-hamburg.de>) bezogen werden.