ECONSTOR
*Make Your Publications Visible.*

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Möhrle, Sascha; Löffler, Caroline; Altmeyer, Michaela; Wittwer, Christina

**Article**

# Ein nachrichtenbasierter Sentiment-Indikator für die deutsche Konjunkturanalyse

ifo Schnelldienst

**Provided in Cooperation with:**
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Möhrle, Sascha; Löffler, Caroline; Altmeyer, Michaela; Wittwer, Christina (2022) : Ein nachrichtenbasierter Sentiment-Indikator für die deutsche Konjunkturanalyse, ifo Schnelldienst, ISSN 0018-974X, ifo Institut - Leibniz-Institut für Wirtschaftsforschung an der Universität München, München, Vol. 75, Iss. 11, pp. 43-46

This Version is available at:
https://hdl.handle.net/10419/272072

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

*Sascha Möhrle, Caroline Löffler, Michaela Altmeyer* und *Christina Wittwer**

# Ein nachrichtenbasierter Sentiment-Index für die Konjunkturanalyse

**IN KÜRZE**

**In einer aktuellen Forschungskooperation zwischen der KfW und dem ifo Institut wird das Potenzial von Unternehmensnachrichten aus Zeitungen für die Konjunkturanalyse untersucht. Mit Hilfe einer KI-basierten Anwendung zur Nachrichtenaufbereitung konstruieren wir einen neuen Stimmungsindikator (Sentiment-Index) für die deutsche Konjunktur. Dieser neue Indikator hat einen hohen Gleichlauf mit traditionellen Stimmungsindikatoren aus Unternehmensbefragungen und weist eine positive Korrelation mit volkswirtschaftlichen Zielgrößen wie dem Wachstum des Bruttoinlandsprodukts auf. Auf Basis der vorliegenden Resultate soll in Zukunft das Potenzial von Zeitungsartikeldaten für die Konjunkturprognose weiter erforscht werden.**

In den letzten Jahren haben neue, unkonventionelle Datenquellen die Verfügbarkeit von Konjunkturindikatoren in Deutschland erheblich verbessert. Dazu gehören beispielsweise Stromdaten (Lehmann und Möhrle 2022), Lkw-Mautdaten (Cox et al. 2020) und Internet-Suchergebnisse (Götz und Knetsch 2019). In Zeiten schneller und tiefgreifender wirtschaftlicher Veränderungen, wie etwa in der Coronakrise, liefern diese Indikatoren politischen Entscheidungsträgern und Unternehmen wichtige Informationen zum Zustand der Konjunktur, da amtliche Daten üblicherweise mit einiger Verzögerung veröffentlicht werden. In Bezug auf neue Datenquellen erfreuen sich insbesondere Textdaten in Form von Zeitungsartikeln einer stark steigenden Beliebtheit in der Konjunkturprognose. Zeitungsdaten haben mehrere Vorteile: i) Sie bilden breite Teile der Gesamtwirtschaft ab, ii) sind in hoher Frequenz (meist täglich) verfügbar, iii) haben im Vergleich zu amtlichen Daten keine (wesentlichen) Publikationsverzögerungen und iv) gehen üblicherweise durch einen Qualitätssicherungsprozess. Eine Reihe von wissenschaftlichen Analysen hat bereits die Prognosefähigkeit von Informationen aus Zeitungsartikeln für verschiedene Länder gezeigt, beispielweise für Norwegen (Thorsrud 2020), Italien (Aprigliano 2022) und die USA (Ellingsen et al. 2022). Im deutschen Kontext hingegen gibt es bisher nur wenig Evidenz zur Vorhersagekraft von Textdaten für die Konjunkturanalyse. In einer neuen Forschungskooperation zwischen der KfW und dem ifo Institut versuchen wir, diese Lücke zu schließen, indem wir einen neuen Stimmungsindikator (Sentiment-Index) für die deutsche Konjunktur konstruieren.

## KARINA – EINE KI-BASIERTE ANWENDUNG ZUR NACHRICHTENAUFBEREITUNG

Die vorliegende Analyse stützt sich auf historische Datensätze, die durch die KI-Anwendung der KfW namens KARINA aufgebaut wurden. KARINA steht für »KI basierte Anwendung zur Risikofrüherkennung und intelligenten Nachrichtenaufbereitung«. Diese Plattform dient der KfW, Kreditrisiken unter anderem aus Zeitungsartikeln und Pressemitteilungen frühzeitig für Bestands- und Neukunden zu identifizieren. Die Artikel werden von Dow Jones Factiva bereitgestellt und stehen seit Dezember 2019 zur Verfügung. Die KfW-Mitarbeiter*innen haben die Möglichkeit, englisch- und deutschsprachige Nachrichten für 3 Mio. internationale Unternehmen und Finanzinstitute und für ihre konkreten Portfolien abzurufen. KARINA verarbeitet automatisiert bis zu 15 000 Zeitungsartikel pro Tag, die von 9 000 Nachrichtenquellen stammen. Dabei ordnet die KI die Artikel den im Text erwähnten Unternehmen und Finanzinstitutionen zu. Die hierfür notwendigen Nachrichten- und Unternehmensinformationen werden über Nachrichten- und Unternehmensdatenanbieter automatisiert bereitgestellt und in KARINA verarbeitet. Außerdem werden risikokritische Schlagwörter identifiziert und die Relevanz des Artikels für das jeweilige Unternehmen festgestellt. Für die Analyse besonders relevant ist jedoch die Bestimmung von eher positiver oder negativer Grundstimmung (Sentiment) in den Zeitungsartikeln.

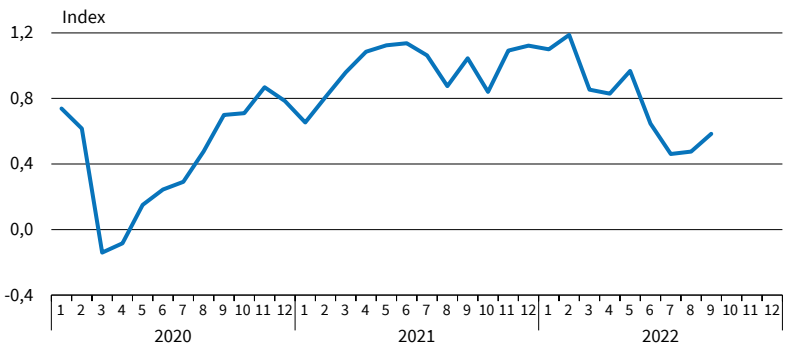
Zur Bestimmung des Sentiments hat die KfW ein komplexes Tool entwickelt. Hierbei kommen verschiedene öffentlich verfügbare KI-Modelle (spaCy und Google BERT) zum Einsatz, die über »supervised learning« weitertrainiert wurden.[1] In der laufenden

* Sascha Möhrle, ifo Institut; Caroline Löffler, Michaela Altmeyer und Christina Wittwer, Kreditanstalt für Wiederaufbau (KfW).
Wir danken Mario Ackermann und Alexander Wilms für ihre Unterstützung bei der Datenaufbereitung sowie Robert Lehmann für hilfreiche Kommentare.

[1] Die Sentiment-Modelle basieren auf Google-BERT-Modellen, die ein mächtiges Sprach- und Semantikverständnis besitzen. Diese wurden für den konkreten Anwendungsfall mit gelabelten Daten, die aus Sätzen aus Finanzartikeln bestehen, für die das Sentiment (positiv, neutral, negativ) vorgegeben wurde, nachtrainiert.

Abb. 1

**Verlauf des Sentiment-Index**




Quelle: KfW; Berechnungen der Autor*innen. © ifo Institut

Anwendung wird das Sentiment für jede Entität (Unternehmen oder Finanzinstitutionen) auf Satzebene bestimmt. Dabei werden nur Sätze in unmittelbarer Nähe der Entitätsnennung berücksichtigt Der Sentiment-Score berechnet den aggregierten Wert für den gesamten Nachrichtenartikel. Dieser ist auf einer Skala zwischen –10 (sehr negativ) und +10 (sehr positiv) definiert, wobei ein Wert von 0 eine neutrale Stimmung widerspiegelt. Da der Sentiment-Score für jede Kombination aus Artikel und darin identifiziertem Unternehmen vorliegt, können wir im Gegensatz zu bisherigen Analysen sehr genau Unternehmensneuigkeiten messen und diese flexibel auf Branchenebene oder gesamtwirtschaftliche Ebene aggregieren.

## DATENAUFBEREITUNG

Bevor die Sentiment-Informationen für eine Zeitreihenanalyse auf monatlicher Ebene genutzt werden können, müssen die Daten zunächst aufbereitet und angepasst werden. Für die Analyse werden nur Artikel über inländische Unternehmen in deutscher Sprache verwendet. Zudem werden nur Artikel einbezogen, in denen das genannte Unternehmen ein zentrales Thema ist, so dass beiläufige Erwähnungen vernachlässigt werden. Dies wird mit Hilfe des von der KfW erstellten Relevanz-Scores sichergestellt. Außerdem werden Artikel mit juristischem Bezug ausgeklammert, die mit hoher Wahrscheinlichkeit keine Konjunktursignale liefern.[2]

Zudem spielen zwei Beobachtungen eine wichtige Rolle: Zum einen gibt es recht große Unterschiede in der Berichterstattung unterschiedlicher Quellen. Beispielsweise berichten Verbandszeitschriften deutlich positiver als große Tageszeitungen. Darüber hinaus liegen auch systematische Unterschiede in der Berichterstattung zwischen verschiedenen Unternehmen und Branchen vor. Da nicht für alle Quellen und Unternehmen Zeitungsartikel für den gesamten Beobachtungszeitraum zur Verfügung stehen, könnte es dadurch zu systematischen Unterschieden des Sentiment-Index kommen (»composition effect«), die nicht auf Konjunktursignale zurückzuführen sind. Um diesen Faktoren Rechnung zu tragen, wird für die Schätzung des monatlichen Sentiments ähnlich wie in Shapiro et al. (2020) ein regressionsbasierter Ansatz verwendet, der für die jeweilige Nachrichtenquelle und das Unternehmen kontrolliert.[3] Die übrig gebliebene Variation über die Zeit wird als Schätzung für den monatlichen Sentiment-Index verwendet.

Nach den vorbereitenden Schritten bleiben rund 324 000 Nachrichtenartikel für ca. 10 000 Unternehmen aus 550 Quellen übrig, auf deren Basis das durchschnittliche Sentiment auf Monatsbasis berechnet wird. Die Zusammensetzung der Artikel auf Branchenebene entspricht in etwa deren Anteilen an der Bruttowertschöpfung, wodurch die Zeitungsartikel weitgehend repräsentativ für die deutsche Wirtschaft sind.

## DER SENTIMENT-INDIKATOR ÜBER DIE ZEIT

Abbildung 1 zeigt die zeitliche Entwicklung des aggregierten Sentiment-Index auf monatlicher Ebene zwischen Dezember 2019 und September 2022. Zunächst einmal fällt auf, dass sich der Index über den gesamten Zeitraum nahe dem neutralen Stimmungsniveau von 0 bewegt, obwohl jedem Zeitungsartikel ein Sentiment-Score zwischen –10 und +10 zugeordnet wird. Im Durchschnitt scheint also das Sentiment nicht besonders stark in jegliche Richtung zu variieren. Nichtsdestotrotz können die relativ gesehen kleinen Veränderungen über die Zeit informativ sein.

Der zeitliche Verlauf des Sentiment-Indikators bildet wichtige wirtschaftliche Veränderungen im Untersuchungszeitraum ab. Der Indikator bricht im März 2020 im Zuge der Coronakrise deutlich ein und signalisiert somit frühzeitig die Auswirkungen der Coronakrise. Danach steigt der Index kontinuierlich an, mit einer kurzen Ausnahme während der 2. Corona-Welle Ende 2020. Im Zuge des Ukraine-Krieges sinkt der Sentiment-Indikator dann ab März 2022 und leitet einen Abwärtstrend ein.

Die schnelle Reaktion des Sentiment-Index auf die Coronakrise und den Ukraine-Krieg deutet bereits an, dass der Indikator potenziell wertvolle Informationen zur konjunkturellen Lage liefert. Um diese Vermutung statistisch zu untermauern, wird im Folgenden der Gleichlauf mit wirtschaftlichen Kennzahlen und anderen zur Verfügung stehenden Konjunkturindikatoren betrachtet. Die Verfügbarkeit der Daten für knapp drei Jahre ist für eine Prognoseübung momentan jedoch noch zu kurz. Daher wird in der folgenden Analyse auf Korrelationsanalysen zurückgegriffen. Sobald eine

[2] Artikel mit juristischem Bezug werden über Schlagwörter (bspw. »Gerichtsverhandlung«) selektiert. Eine Einbeziehung solcher Artikel verschlechtert die Ergebnisse signifikant und bietet daher keinen Mehrwert für die Konjunkturanalyse.

[3] Die Schätzgleichung lautet: $Sentiment_{a,t} = \beta_0 + \beta_1 Unternehmen_a + \beta_2 Quelle_a + \gamma_t + \varepsilon_a$, wobei $a$ den Artikel und $t$ den Zeitpunkt des Artikels repräsentiert. $\gamma_t$ repräsentiert den monatlichen »Time Fixed Effect«, der den monatlichen Sentiment-Index bildet.

längere Historie zur Verfügung steht, werden wir diese Ergebnisse auch in einer Prognoseübung (»out-of-sample forecast«) untersuchen.

## KORRELATION MIT IFO UNTERNEHMENSBEFRAGUNGEN

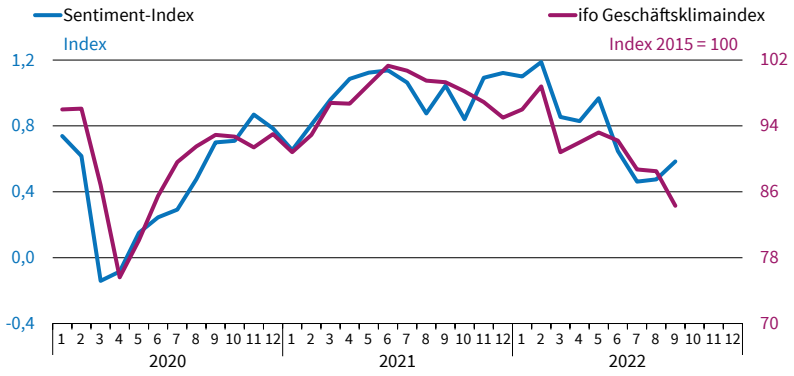
Im ersten Schritt der Analyse soll der auf Basis von Zeitungsartikeln geschaffene Sentiment-Indikator mit einem der wichtigsten Konjunkturindikatoren der deutschen Wirtschaft, dem ifo Geschäftsklimaindex, verglichen werden. Das ifo Institut befragt monatlich rund 8 000 Unternehmen nach ihrer aktuellen wirtschaftlichen Situation (»Geschäftslage«) und ihren Erwartungen für die nächsten sechs Monate (»Geschäftserwartungen«), die jeweils zur Hälfte in den Geschäftsklimaindex einfließen. Der ifo Geschäftsklimaindex spiegelt somit die Stimmung aus Sicht von Führungskräften in deutschen Unternehmen wider. Die Analyse von Lehmann (2022) zeigt, dass die ifo-Befragungsdaten sehr gute Prognoseeigenschaften für die deutsche Konjunktur besitzen. Eine starke Korrelation zwischen dem Sentiment-Indikator und den Befragungsergebnissen würde bestätigen, dass unsere Messung der Nachrichtenstimmung kein reines Rauschen ist und somit fundamentale Konjunktursignale widerspiegelt.

Abbildung 2 zeigt, dass der Sentiment-Indikator einen hohen Gleichlauf mit der Zeitreihe des ifo Geschäftsklimaindex aufweist, der jeweils zur Hälfte aus der Geschäftslage und den Geschäftserwartungen besteht. Tabelle 1 stellt die vorlaufenden Eigenschaften des Sentiment-Indikators gegenüber den ifo-Zeitreihen in Form von Kreuzkorrelationen dar. Die Tabelle bestätigt die visuelle Darstellung aus Abbildung 2: Im Zeitraum zwischen Dezember 2019 und September 2022 scheint insbesondere der Gleichlauf zwischen Sentiment-Index und ifo Geschäftsklima hoch zu sein. Der Korrelationskoeffizient beträgt 0,85.

Zudem zeigt Tabelle 1, dass der Sentiment-Index tendenziell vorlaufend gegenüber der ifo Geschäftslage ist. Hier wird die höchste Korrelation mit der Geschäftslage im nachfolgenden Monat erreicht (Koeffizient: 0,84). Der textbasierte Indikator scheint also vorlaufende Eigenschaften für die zukünftige Lage der Unternehmen im Aggregat aufzuweisen. Außerdem

Abb. 2

**Verlauf des Sentiment-Index und ifo Geschäftsklimaindex (Gesamtwirtschaft)**




Quelle: KfW; ifo Konjunkturumfragen; Berechnungen der Autor*innen. © ifo Institut

ist der Sentiment-Index tendenziell gleichlaufend mit den Geschäftserwartungen der Unternehmen (Korrelation: 0,70).

Sowohl für den ifo Geschäftsklimaindex als auch für den Sentiment-Indikator lassen sich neben gesamtwirtschaftlichen auch branchenbezogene Analysen durchführen, was ein wesentlicher Vorteil gegenüber anderen Konjunkturindikatoren ist. Daher betrachten wir in Tabelle 2 auch die Kreuzkorrelationen zwischen den Indikatoren für das Verarbeitende Gewerbe, den Dienstleistungssektor und den Handel.[4] Die Analyse zeigt, dass der Sentiment-Indikator insbesondere die Stimmung im Verarbeitenden Gewerbe sehr gut wiedergibt. Die gleichlaufende Korrelation beträgt 0,84. Aber auch im Dienstleistungssektor und im Handel sind die Ergebnisse recht vielversprechend. Insgesamt zeigt der Vergleich mit den ifo Stimmungsindikatoren, dass der Sentiment-Indikator die Stimmung der deutschen Wirtschaft gut widerspiegelt und eine hohe Informationsgüte (»signal-to-noise ratio«) aufzuweisen scheint.

## KORRELATION MIT VOLKSWIRTSCHAFTLICHEN ZIELGRÖSSEN

In einem zweiten Schritt sollen nun die Korrelationen mit volkswirtschaftlichen Zielgrößen untersucht werden, die für die Konjunkturanalyse zentral sind. Das

[4] Für den Bausektor stehen zu wenige Nachrichtenartikel zur Verfügung, so dass auf eine Analyse hier verzichtet wird.

Tab. 1

**Korrelationskoeffizienten für den Zusammenhang zwischen Sentiment-Index und ifo Konjunkturindikatoren (Gesamtwirtschaft)**

| | Vorlauf (*t*) in Monaten | | |
|---|---|---|---|
| | 0 | 1 | 2 |
| ifo Geschäftsklima | 0,85 | 0,83 | 0,61 |
| ifo Geschäftslage | 0,68 | 0,84 | 0,80 |
| ifo Geschäftserwartungen | 0,70 | 0,55 | 0,25 |

Anmerkung: Die Tabelle zeigt die Korrelationskoeffizienten zwischen dem Sentiment-Index und verschiedenen ifo Konjunkturindikatoren. Ein Vorlauf von »0« bezieht sich auf die Korrelation mit den ifo Indikatoren im gleichen Monat (Gleichlauf des Sentiment-Index), »1« auf die Korrelation mit den ifo Indikatoren im nachfolgenden Monat (Vorlauf des Sentiment-Index von einem Monat) und »2« auf die Korrelation mit den ifo Indikatoren im übernächsten Monat (Vorlauf des Sentiment-Index von zwei Monaten).

Quelle: KfW; ifo Institut.

Tab. 2

**Korrelationskoeffizienten für den Zusammenhang zwischen Sentiment-Index und ifo Geschäftsklima auf Branchenebene**

| | Vorlauf (*t*) in Monaten | | |
|---|---|---|---|
| | 0 | 1 | 2 |
| Verarbeitendes Gewerbe | 0,84 | 0,81 | 0,66 |
| Dienstleistungssektor | 0,64 | 0,68 | 0,46 |
| Handel | 0,58 | 0,47 | 0,24 |

Anmerkung: Die Tabelle zeigt die Korrelationskoeffizienten zwischen dem Sentiment-Index und dem ifo Geschäftsklima auf Branchenebene. Ein Vorlauf von »0« bezieht sich auf die Korrelation mit dem ifo Geschäftsklima im gleichen Monat (Gleichlauf des Sentiment-Index), »1« auf die Korrelation mit dem ifo Geschäftsklima im nachfolgenden Monat (Vorlauf des Sentiment-Index von einem Monat) und »2« auf die Korrelation mit dem ifo Geschäftsklima im übernächsten Monat (Vorlauf des Sentiment-Index von zwei Monaten).

Quelle: KfW; ifo Institut.

sind in erster Linie das Bruttoinlandsprodukt (BIP) und die Industrieproduktion.

Zunächst wird der Gleichlauf zwischen dem Sentiment-Index und dem BIP betrachtet. Da das BIP in quartalsweisen Wachstumsraten vorliegt, aggregieren wir die Textinformationen auf Quartalsebene und bilden die ersten Differenzen des Sentiment-Index gegenüber dem Vorquartal. Zwar ist die Analyse mit einer Zeitspanne zwischen 2. Quartal 2020 und 2. Quartal 2022 recht kurz[5], doch scheint der Sentiment-Index einen recht hohen Gleichlauf mit dem BIP-Wachstum aufzuweisen; der Korrelationskoeffizient beträgt 0,73. Im Vergleich dazu beträgt der Korrelationskoeffizient des Gleichlaufs zwischen ifo Geschäftsklimaindex und BIP-Wachstum 0,85.

Der Sentiment-Index für das Verarbeitende Gewerbe weist auch eine positive Korrelation mit der Industrieproduktion aus. Der Korrelationskoeffizient für den Gleichlauf beträgt 0,61 in Levels (0,68 für den nachfolgenden Monat) und 0,24 in ersten Differenzen (0,42 für den nachfolgenden Monat). Die schwächeren Korrelationen in ersten Differenzen könnten darauf zurückzuführen sein, dass die Vormonatsveränderungen des Sentiment-Index recht volatil sind. Durch eine noch bessere Aufbereitung und Filterung der Daten könnten sich die Resultate noch verbessern. Insgesamt scheint der Sentiment-Index ein vielversprechender Konjunkturindikator zu sein, der auch Analysen auf Branchenebene zulässt.

[5] Das 4. Quartal 2019 wird bei der Analyse ignoriert, da nur die Daten für einen Monat (Dezember 2019) zur Verfügung stehen. Aufgrund der Bildung von Differenzen beginnt die Analyse im 2. Quartal 2020. Für das BIP-Wachstum stehen zum Zeitpunkt der Erstellung dieser Analyse die Daten aus dem 3. Quartal 2022 noch nicht zur Verfügung.

## AUSBLICK

Auf Basis der bisherigen Ergebnisse scheint das nachrichtenbasierte Sentiment eine vielversprechende Ergänzung zu umfragebasierten Methoden und ein potenziell wertvoller Indikator für die Konjunkturanalyse zu sein. Da die Daten bisher nur für einen recht kurzen Zeitraum zur Verfügung stehen, sind weitere Analysen zur Bestätigung der bisherigen Resultate notwendig. Insbesondere sollten zu einem späteren Zeitpunkt die Ergebnisse durch ein Prognoseexperiment bestätigt werden. In der weiteren Analyse planen wir auch, die Texte mit Hilfe von »Topic-Modelling-Ansätzen« zu analysieren und regionale und branchenbezogene Aspekte stärker zu betrachten.

## LITERATUR


Aprigliano, V., S. Emiliozzi, G. Guaitoli, A. Luciani, J. Marcucci und L. Monteforte (2022), »The Power of Text-based Indicators in Forecasting Italian Economic Activity«, *International Journal of Forecasting*.

Cox, M., J. Triebel, S. Linz, C. Fries, L. F. Flores, A. Lorenz, D. Ollech, A. Dietrich, J. LeCrone und K. Webel (2020), »Täglicher Lkw-Maut-Fahrleistungsindex aus digitalen Prozessdaten der Lkw-Mauterhebung«, *WISTA – Wirtschaft und Statistik* (4), 63–76.

Götz, T. B. und T. A. Knetsch (2019), »Google Data in Bridge Equation Models for German GDP«, *International Journal of Forecasting* 35(1), 45–66.

Ellingsen, J., V. H. Larsen und L. A. Thorsrud (2022), »News Media versus FRED-MD for Macroeconomic Forecasting«, *Journal of Applied Econometrics* 37(1), 63–81.

Lehmann, R. (2020), »The Forecasting Power of the ifo Business Survey«, CESifo Working Paper No. 8291.

Lehmann, R. und S. Möhrle (2022), »Forecasting Regional Industrial Production with High-Frequency Electricity Consumption Data«, *CESifo Working Paper* No. 9917.

Shapiro, A. H., M. Sudhof und D. J. Wilson (2020), »Measuring News Sentiment«, *Journal of Econometrics* 228(2), 221–243.

Thorsrud, L. A. (2020), »Words Are the New Numbers: A Newsy Coincident Index of the Business Cycle«, *Journal of Business & Economic Statistics* 38(2), 393–409.