

8 Das LMU-ifo Economics Business Data Center

HEIKE MITTELMEIER

8.1 Allgemeines zum EBDC

Das Economics & Business Data Center (EBDC) wurde im Jahr 2008 als Kooperation der LMU München und des ifo Instituts mit Unterstützung von LMUexcellent gegründet. Das Ziel war und ist es unter anderem, neue Felder für die wirtschaftswissenschaftliche Forschung zu erschließen. Für diesen Zweck stellt das EBDC innovative Datensätze deutscher Unternehmen für Forscher bereit, die nicht-kommerzielle, wissenschaftliche Themen bearbeiten. Neben den Mikrodaten der regelmäßigen ifo-Befragungen bietet es verknüpfte Datensätze, die sowohl Umfragedaten des ifo Instituts als auch externe Bilanzdaten der Firmendatenbanken Amadeus und Hoppenstedt enthalten. Damit werden den unternehmensspezifischen Erwartungen, Einschätzungen und Plänen aus den ifo-Daten realisierte Größen aus den Bilanz- und Strukturdaten der Unternehmensdatenbanken gegenübergestellt. Derzeit stehen das EBDC Business Expectations Panel (Fokus: konjunkturelle Faktoren/Bilanzdaten), das EBDC Business Investment Panel (Fokus: Investitionen/Bilanzdaten) und das EBDC Business Innovation Panel (Fokus: Innovationen/Bilanzdaten) zur Verfügung. Die Datensätze werden sowohl in zeitlicher Dimension als auch inhaltlich kontinuierlich erweitert.¹

Zu den Aufgaben des EBDC gehören auch die Beschaffung und Verwaltung wichtiger Datenquellen für Forschung und Lehre, die zentrale Bereitstellung, Aktualisierung und Dokumentation von Datenbanken externer Anbieter, der Erwerb von entsprechenden Support-Tools sowie die Archivierung der Forschungsdaten. Das EBDC hält in dieser Hinsicht eine geeignete Infrastruktur vor und leistet Unterstützung im Hinblick auf die softwaretechnische Wissensvermittlung. Da es im Rahmen des ifo-Datenpools auch Zugang zu den seit 1949 regelmäßig deutschlandweit erhobenen und zum Teil seit 1964 verfügbaren Umfragedaten des ifo Instituts bietet, ergeben sich für empirisch arbeitende Wissenschaftler hohe Synergieeffekte.

Die Daten werden außerdem anonymisiert und nur mit Zeitverzögerung (Schutzfrist je nach Panel unterschiedlich) zur Verfügung gestellt. Die nachfolgenden Ausführungen geben einen Überblick über Datenbasis, Umfang und Zugang zu den EBDC-Unternehmenspanels und bieten darüber hinaus Informationen zur Verknüpfungstechnik des Probabilistic Record Linkage.

¹ Frühere Ausführungen zum EBDC finden sich bei Abberger u. a. (2007a), Becker und Wohlrabe (2008) sowie Seiler (2012).

8.2 Nutzung der Mikrodaten zu Forschungszwecken

8.2.1 Datenschutz im EBDC

Das EBDC sieht sich als Dienstleister, welcher Professoren, Forscher und Doktoranden bei ihren Forschungsprojekten unterstützt. Diese müssen nicht-kommerzielle, wissenschaftliche Themen behandeln, die auf der Basis der EBDC-Daten empirisch bearbeitet werden. Aufgrund der hohen Vertraulichkeitsanforderungen bzw. der Verpflichtung zur Geheimhaltung einzelner Befragungsergebnisse sowie der Unternehmensidentität der Firmen des ifo-Umfragepanels können die Daten nur in den Räumlichkeiten des ifo Instituts/EBDC genutzt werden. Hier wird ein Rechner ohne Zugang zu Internet, Drucker oder anderen externen Datenträgern zur Verfügung gestellt. Die Analysen dürfen nur in Anwesenheit eines EBDC-Mitarbeiters durchgeführt werden. Dieser wird nach Beendigung des Aufenthalts auch sicherstellen, dass nur aggregierte Forschungsergebnisse zugesendet werden, die keine Rückschlüsse auf einzelne Unternehmen oder die Panelzusammensetzung insgesamt zulassen.

Zugang zu den EBDC Business Panel, kann über ein Formular im Internet beantragt werden,² in dem auch das Forschungsvorhaben kurz beschrieben werden soll, ebenso wie der Projektzeitraum und eventuelle Verknüpfungen. Auf Anfrage sendet das EBDC auch ein Testpaket per E-Mail, welches einen anonymisierten EBDC-Testdatensatz im Stata-Format sowie die Variablenliste zum jeweiligen Original-Panel enthält. Die Forschungsvorhaben werden vom EBDC ausdrücklich unterstützt und sind daher kostenfrei – der Zugang zu den EBDC-Daten ist allerdings von der Genehmigung des Vorhabens und der Verfügbarkeit eines freien Arbeitsplatzes abhängig.

8.2.2 Die EBDC Business Panels

In den EBDC Business Panels werden die Mikrodaten der ifo Konjunkturumfragen, der ifo Investitionsumfrage sowie der ifo Innovationsumfrage³ mit den Bilanzdaten der Unternehmensdatenbanken Amadeus und Hoppenstedt verknüpft. Damit liegen drei umfangreiche Datensätze deutscher Unternehmen vor, welche durch die Integration verschiedenartiger Datenbestände die gleichzeitige Untersuchung von erwarteten bzw. geplanten und realisierten Größen ermöglichen. Die regelmäßig aktualisierten Panels umfassen sowohl historische als auch aktuelle Daten, wobei das EBDC Business Expectations Panel im Jahr 1980, das EBDC Business Investment 1964 und das EBDC Innovation Panel 1982 beginnt. Im folgenden Kapitel soll auf die Datenbasis der EBDC Business Panels und die Unternehmensdatenbanken sowie deren Umfang und Verfügbarkeit eingegangen werden.⁴ Außerdem wird kurz die Verknüpfungstechnik des Probabilistic Record Linkage, auf dem die Erstellung der EBDC Business

² <http://www.cesifo-group.de/de/ifoHome/facts/EBDC/Access-Concept.html>

³ Vgl. Beschreibungen der ifo-Erhebungen in den Kapiteln 3.2, 3.3, 3.4, 3.5, 5.1, 5.3, 7.1, 7.2, 7.3. dieses Buches.

⁴ Die ifo-Erhebungen werden über den ifo-Datenpool zwar ebenso im EBDC der Forschung zur Verfügung gestellt; da die Erhebungen in den entsprechenden Kapiteln dieses Buches schon beschrieben werden, wird hier auf eine explizite Darstellung verzichtet.

Panel basiert, erläutert.

Firmendatenbank Amadeus

Die Amadeus-Firmendatenbank ist neben den ifo Umfragedaten und der Hoppenstedt-Datenbank eine weitere zentrale Quelle für die EBDC Business Panels. Sie ist ein Produkt der Bureau van Dijk Electronic Publishing GmbH (BvDEP), eines der führenden europäischen Anbieter globaler Unternehmensinformationen, und enthält Geschäfts- und Finanzinformationen zu mehr als 11 Millionen Unternehmen aus 41 Ländern Europas, wobei derzeit ca. 1,6 Mio. deutsche Unternehmen erfasst sind.

Für die Firmendatenbestände werden Informationen marktführender lokaler Institutionen und renommierter Unternehmen der jeweiligen Länder herangezogen. Die Abschlussdaten für deutsche Unternehmen stammen von Creditreform bzw. der Creditreform Rating AG, die zur Creditreform Gruppe gehört. Diese stellt seit mehr als 125 Jahren Bonitätsauskünfte zu Kunden-Lieferanten-Beziehungen bereit und ist heute europaweit Marktführer für Bonitätsinformationen. Zentrale Quellen der Amadeus-Datenbank sind die MARKUS-Datenbank, welche i.A. Gesellschaftsinformationen deutscher Handelsregisterunternehmen mit einem Bonitätsindex von maximal 499 enthält,⁵ und die DAFNE-Datenbank, die Jahresabschlüsse, Beteiligungsdaten etc. aller publizierenden deutschen Unternehmen beinhaltet (Creditreform Rating AG). Im Unterschied zur DAFNE-Datenbank (Rohdatenformat) liegen die Daten bei Amadeus jedoch in einem einheitlichen, standardisierten Bilanzformat vor, welches von nationalen bzw. internationalen Rechnungslegungsvorschriften abstrahiert. So besteht jeder Unternehmensbericht aus insgesamt 23 Bilanzpositionen, 25 Positionen der Gewinn- und Verlustrechnung, 20 Finanzkennzahlen und zahlreichen deskriptiven Informationen wie bspw. Branchencodes, Gesellschafterstrukturen, Aktien- oder Kursinformationen. Für die Unternehmen der EBDC-Unternehmenspanels wurden hieraus über 50 Positionen ausgewählt oder neu berechnet; Beteiligungs-, Aktien- und Kursinformationen wurden jedoch zunächst vernachlässigt.⁶ Die Abschlussinformationen für nicht börsennotierte Unternehmen sind spätestens nach 2-3 Jahren in den Amadeus-Datenbanken und damit auch in den EBDC Business Panels verfügbar.

Bilanzdatenbank Hoppenstedt

Die Hoppenstedt-Bilanzdatenbank ist ein Produkt der Hoppenstedt Firmeninformationen GmbH, welche seit 2013 Teil der Unternehmensgruppe Bisnode ist. Wesentliche Geschäftsfelder sind neben der Bereitstellung von Firmeninformationen der Adressverkauf, Kredit- und Risikoanalysen sowie die Publikation von Fachzeitschriften. Je nach Kundenstatus und Nutzungsrechten kann daher aus einer Vielzahl an Datenbanken gewählt werden. Für die Firmendatenbestände werden Informationen aus externen Quellen, wie z.B. Bundesanzeiger,

⁵ Dieses wird von Creditreform bewertet.

⁶ Diese können jedoch über die historischen Datenbanken im EBDC exportiert werden.

Handelsregister, Wirtschaftspresse oder Geschäftsberichte, herangezogen und bei Bedarf auch im direkten Dialog ermittelt.

Die Hoppenstedt-Bilanzdatenbank,⁷ welche aktuell mehr als 3,5 Mio. Abschlüsse von über 1 Mio. deutschen Unternehmen aus den Bereichen Industrie, Handel, Dienstleistungen, Versicherungen und Banken beinhaltet, liefert die Finanzinformationen für die EBDC Business Panels. Erfasst sind hier die seit 2005 publizierten Abschlüsse, die Informationen über große Unternehmen gehen teilweise bis ins Jahr 1987 zurück. Die erhobenen Daten zu Bilanzen sowie Gewinn- und Verlustrechnungen einzelner Firmen sind jeweils in unterschiedlichen Detailtiefen abrufbar (Normbilanz: maximal verfügbare Positionen nach der jeweiligen Rechnungslegungsvorschrift; Verkürzte Bilanz: ca. 90; Kurzbilanz: ca. 30 Positionen), zudem ermöglichen separate, eng am jeweiligen Original orientierte Bilanzschemata die Berücksichtigung der einzelnen Abschlussarten nach HGB, IAS und US-GAAP.

Verknüpfungsmethode – Probabilistic Record Linkage

Zur Verknüpfung der ifo Umfragen auf der einen und der Bilanzdatenbanken Amadeus und Hoppenstedt auf der anderen Seite wird auf die Adressinformationen der Unternehmen in den einzelnen Datenbeständen zurückgegriffen. Auf diese Weise lassen sich jeweils zwei Zuordnungstabellen (ifo-Amadeus und ifo-Hoppenstedt) bilden, welche anschließend im entsprechenden EBDC Business Panel zusammengeführt werden können.

Im Folgenden soll beispielhaft das Record Linkage der Adressdaten der ifo Konjunkturumfragen mit denen aus der Firmendatenbank Amadeus beschrieben werden, wobei hier zusätzlich zum normalen Verknüpfungsprozess, der beim Record Linkage mit Hoppenstedt bzw. im Falle der ifo Investitionsumfrage analog verläuft, ein sogenannter „Goldstandard“ erzeugt wurde. Dieser Goldstandard wird benötigt, um die Match- bzw. Nonmatch-Gewichte für jede Adressvariable zu ermitteln. Die Verknüpfung der Adressdatensätze des ifo-Datenpools (Basis: gesamte Konjunkturumfragen inkl. Bau, Handel, Dienstleister, Industrie) mit den deutschen Firmen in der Amadeus-Datenbank erfolgte dabei mit Hilfe der Matching-Software MTB (Merge Toolbox)⁸, welche am „Center for quantitative Methods and Survey Research“ der Universität Konstanz entwickelt wurde.

MTB ermöglicht die Zuordnung unterschiedlicher Datensätze bei großem Datenumfang und unterschiedlichen Adressaufbereitungen und wird eingesetzt, wenn kein eindeutiger Schlüssel wie z.B. eine Firmenkennziffer vorliegt. Auf der Basis des Probabilistischen Record Linkage, das auf der Theorie von Newcombe u. a. (1959), aufbaut und von Fellegi und Sunter (1969) formalisiert wurde, werden gleiche Namens-/Adressdaten zusammengefügt. Der Grad der Übereinstimmung (die „Ähnlichkeit“) der Variablen wird hierbei mit Hilfe von Wahrscheinlichkeiten ermittelt.

⁷ Zum Zeitpunkt der Datenerhebung für das EBDC waren bei Hoppenstedt ca. 120.000 Unternehmen verfügbar.

⁸ Homepage mit Downloadmöglichkeit: <http://record-linkage.de/-Downloads--software.htm>

Diese „Ähnlichkeit“ errechnet sich aus dem Quotienten aus der Wahrscheinlichkeit, dass für gleiche Firmen die Variable x aus beiden Datenmengen als übereinstimmend erkannt wird (M-Wahrscheinlichkeit), und der Wahrscheinlichkeit, dass für ungleiche Firmen die Variable x aus beiden Datenmengen als übereinstimmend gewertet wird (u-Wahrscheinlichkeit). Im Idealfall lautet der Quotient $1/0$, wobei Abweichungen zwischen den verschiedenen Variablen unterschiedlich bewertet werden müssen.

Aus diesem Grund wird für jede Variable der Logarithmus der M-Wahrscheinlichkeit (= Match-Gewicht) bzw. u-Wahrscheinlichkeit (= Nonmatch-Gewicht) berechnet, um im Falle eines Matches bzw. Nonmatches die Qualitätsvariable („Quality“) der Verknüpfung zu erhöhen bzw. zu reduzieren. Die Quality-Variable bildet dabei den Grad der Ähnlichkeit einer Datensatzverknüpfung ab und ergibt sich aus der Summe der Gewichte der einzelnen Adressvariablen.

Da die Parameter (Gewichte) des Probabilistic Record Linkage empirisch gewonnen werden mussten, wurde aus der Menge der Datensätze zunächst eine gesichert zuzuordnende Teilmenge gebildet – ein sogenannter „Goldstandard“. Als Goldstandard wird generell eine Verknüpfung bezeichnet, die eine eindeutige Zuordnung der Datensätze zweier Datenbanken ermöglicht. Das erste Record Linkage erfolgte deshalb über Telefonnummer, Faxnummer und E-Mail-Adresse. Diese Variablen waren jedoch nicht in allen Datensätzen befüllt, und da auch erhebliche Unterschiede bei der Systematik der Telefon-/Faxnummerneingabe bestanden, war eine Verknüpfung der gesamten Datenmenge auf diese Weise nicht möglich. Es ließen sich so 40% der ifo-Einträge direkt einem Unternehmen aus Amadeus zuordnen. Dabei zeigte sich allerdings, dass auch Datensätze als Matches identifiziert wurden, die eigentlich nicht zusammengehören. Dies ist häufig dann der Fall, wenn innerhalb einer Firma mehrere Einheiten existieren – z.B. Verwaltungsgesellschaft, Holding und Geschäftsführung –, die sich alle am selben Standort befinden und so möglicherweise über dieselbe zentrale Telefon- oder Faxnummer oder E-Mail-Adresse verfügen. Einige der verknüpften Goldstandardpaare waren somit also keine „True Matches“ im eigentlichen Sinne, sie waren nur „True Matches“ im Sinne der Zugehörigkeit zu einem größeren Unternehmensverbund.⁹

Innerhalb dieser verknüpften Datenmenge wurden nun die Gewichte für die einzelnen Adressvariablen berechnet. Dafür wurde die Zahl der richtigen bzw. falschen Übereinstimmungen zur Grundgesamtheit ins Verhältnis gesetzt. Um das Ergebnis nicht zu verzerren, wurden die Gewichte zusätzlich für verschiedene Aufbereitungsvarianten der Variablen ermittelt. Damit verringerte sich das Risiko, das Ergebnis aufgrund häufig vorkommender Namenssequenzen zu positiv bzw. aufgrund unterschiedlicher Schreibweise des gleichen Namens zu negativ zu bewerten.

Aus den daraus resultierenden Ähnlichkeitsgewichten, die vor dem eigentlichen MTB-Lauf den Variablen zugeordnet wurden, errechnete das Programm die Quality. Wurden für die Verknüpfung eines Unternehmens aus beiden Datenbeständen für alle Variablen positive

⁹ Es wurde eine zusätzliche Handkontrolle durchgeführt.

Matches ermittelt, so ergab sich die Quality als maximales Gesamtgewicht aus der Summe der einzelnen Übereinstimmungsgewichte: z.B. $10,6 + 8,6 + 4,9 + 12,9 = 37,0$. Wurde dagegen beispielsweise eine der Variablen von MTB als Nonmatch identifiziert - und damit das (negative) Nonmatch-Gewicht in die Gleichung eingesetzt - verringerte sich die Quality für die jeweilige Verknüpfung entsprechend.

Der beschriebene Abgleich der Variablen aus dem ifo- und dem Amadeus-Panel wurde mittels einer String-Ähnlichkeitsfunktion durchgeführt, welche N-Gramme der Länge 2 (= Bigramm mit Leerzeichen vor und hinter allen Strings) aus den jeweiligen Variablen miteinander vergleicht, indem ein Raster der Länge 2 über den String gelegt wird. Unterschiede in den Variablenausprägungen wurden linear nach Bigramm-Ähnlichkeit gewichtet, wobei MTB eine hohe Übereinstimmung bei einem kürzeren Namen niedriger bewertet als eine hohe Übereinstimmung bei einem längeren Namen. Die Festlegung eines für alle Variablen gültigen Jaro-Faktors (= Gewichtsanzpassung) auf den Wert „2“ führte weiterhin zu einer schnelleren Vergabe des vollen Übereinstimmungsgewichts bei hoher Übereinstimmung und damit zu einer besseren Differenzierung der Matches von den Nonmatches. Für die Auswertung des MTB-Laufs wurde schließlich ein Schwellenwert hinsichtlich der Quality-Variable definiert, wobei allerdings nicht auf die Bewertung des Programms zurückgegriffen wurde, welches eine Verknüpfung ab einer gewissen Quality als Match identifiziert. In unserem Fall war es zweckmäßiger, höhere Werte anzusetzen und für einen großen Quality-Bereich eine erneute Handkontrolle durchzuführen. Hiermit wurde wieder das Ziel verfolgt, den Fehler falsch positiver Treffer zu vermeiden – Pärchen können nicht als True Match qualifiziert werden, nur weil sie über einem Schwellenwert liegen (Ort im Namen kann die Ähnlichkeit erhöhen, ebenso die Rechtsform). Zum anderen können sich unterhalb des definierten Schwellenwerts in bestimmten Blöcken noch einige True Matches befinden, die wegen fehlender Informationen zu stark abgewertet wurden (gleicher Name, Straße leer).

Ergebnisse

Auf diese Weise wurden die Unternehmensadressen der ifo Konjunkturumfrage mit den Unternehmensadressen aus der Amadeus/Hoppenstedt-Firmendatenbank verglichen. Da die Adresszuordnung des BEP auf Fragebogenebene (questionnaire_id) stattfindet, können einem Unternehmen aus Amadeus durchaus auch zwei oder mehr ifo-Einträge zugeordnet sein, die jeweils für die Fertigung eines anderen Produktes an derselben Adresse stehen. Abbildung 8.1 verdeutlicht nochmals den Entstehungsprozess des EBDC Business Expectations Panel. Da in der ifo Investitionsumfrage nicht zwischen Produkten unterschieden wird, ist jeder ifo-Adresseintrag genau einem Eintrag bei Amadeus bzw. Hoppenstedt zugeordnet und wiederum über die BIP-Company-ID identifizierbar. Allgemein gilt: Es werden je Firma die Bilanzinformationen von dem Anbieter mit den meisten Bilanz-Jahrgängen bevorzugt.

Tabelle 8.1: EBDC Statistik (Stand 2016)

Umfrage	Zeitraum	Beobachtungen	Unternehmen	davon im EBDC Panel
<i>Konjunkturumfragen:</i>				
Industrie	ab 1980	1.602.450	14.131	6.014
Bau	ab 1991	941.870	4.273	2.072
Handel	ab 1990	797.491	6.971	2.930
Dienstleister	ab 2004	308.694	6.866	5.040
Versicherungen	ab 1999	46.213	156	-
<i>Investitionsumfrage:</i>				
Industrie	ab 1964	365.260	23.491	6.034
Handel	ab 2000	15.696	4.546	-
<i>Innovationsumfrage:</i>				
Industrie	ab 1982	41.996	7.862	3.060

Stand 2016, detaillierte Analysen nach Sektoren können auf Wunsch zugeschickt werden.

8.2.3 Aufbau der EBDC-Unternehmenspanels

Generell sind alle EBDC Business Panels ähnlich aufgebaut. Die Identifikation erfolgt über eine unternehmenseigene EBDC-ID, das Jahr der Beobachtung sowie weitere Datensatzspezifische Zeitvariablen. Den Abschlussinformationen in den EBDC-Unternehmenspanels liegen, wenn verfügbar, Einzel- an Stelle von Konzernabschlüssen zugrunde,¹⁰ wobei die Bilanzdaten der beiden Unternehmensdatenbanken nicht einfach übernommen wurden. Es wurde vielmehr ein neues EBDC-Bilanzschema entwickelt, welches sowohl Amadeus- als auch Hoppenstedt-Variablen integriert und von den bestehenden Unterschieden der ursprünglichen Datenbestände abstrahiert.¹¹ Das EBDC-Bilanzschema orientiert sich dabei an Bilanz- und GuV-Struktur des Handelsgesetzbuchs (HGB) und weist teilweise auch Variablen nach Gesamt- oder Umsatzkostenverfahren aus.¹² Für eine detaillierte und entsprechend gegliederte Übersicht zu den verfügbaren Bilanz- und GuV-Variablen sei auf die jeweilige Variablenliste des entsprechenden EBDC Business Panels verwiesen.

EBDC Business Expectations Panel

Für das EBDC Business Expectations Panel (BEP) ergibt sich durch die Verknüpfung eine Kombination aus Monats- (ifo Konjunkturumfrage) und Jahresdaten (Bilanz). Die BEP-ID setzt sich

¹⁰ Die jeweilige Abschlussart wird durch die Variable „reporting_basis“ angezeigt. Limited financial data bedeutet hierbei, dass die Bilanzinformationen nicht veröffentlicht, sondern meist individuell erfragt wurden.

¹¹ Das Umrechnungsschema, die zur Überführung der ursprünglichen Variablen in die neu generierten EBDC-Bilanzvariablen verwendet wurde, kann am EBDC eingesehen werden. Auch ist eine detaillierte Standard-Bilanz verfügbar, welche zu Orientierungszwecken herangezogen werden kann.

¹² Auf Wunsch und in Ausnahmefällen können die EBDC-Panels auch mit den ursprünglichen Bilanzvariablen aus Amadeus und Hoppenstedt zur Verfügung gestellt werden.

aus drei Bestandteilen zusammen: aus der Variablen „bep_company_id“, einer durchlaufenden Unternehmensnummer, der Variablen „questionnaire_id“, welche die Fragebögen pro Unternehmen durchnummeriert, und der Variablen „sector_id“, welche die Sektoren enthält, die abgefragt werden. In der Konjunkturumfrage Verarbeitendes Gewerbe (KTVG) wird dabei nach Produkten gefragt, in der Konjunkturumfrage Dienstleister (KTDL) nach Dienstleistungssparten, in der Konjunkturumfrage Handel (KTHAN) nach Produktgruppen und in der Konjunkturumfrage Bau (KTBAU) nach Bausparten. Ein Fragebogen kann Fragen für mehrere Sparten enthalten.

Der Datensatz ist nach BEP-ID, Jahr (year), Monat (month) sortiert. Jede Meldung ist über diese drei Variablen identifizierbar. In den einzelnen Monaten sind jeweils die Umfrageergebnisse aus der ifo Konjunkturumfrage enthalten, so dass für jedes Jahr bis zu 12 Monatsmeldungen pro BEP-ID vorliegen können. Daran anschließend folgt die Bilanzinformation in einem dafür konstruierten Monat „99“. Der Vorteil dieser Datensatz-Struktur ist vor allem die übersichtliche Handhabung, die eine individuelle Zuordnung von Monats- und Jahresinformationen möglich macht.

Der Datensatz enthält aufeinanderfolgend die nach ihrer Funktion geordneten Variablen: Identifikationsvariablen, Bilanz- und GUV- sowie ifo-Variablen. Zu den Identifikationsvariablen zählen neben der BEP-ID, dem Jahr und dem Monat auch Angaben zu Branchencodes, Beschäftigtengrößenklassen, Bundesland, Börsennotierung, Rechtsform etc.¹³ Die Variablenliste enthält neben einer Übersicht zu allen Variablen, deren Bedeutung und Besonderheiten auch die Fragen der einzelnen ifo Konjunkturumfragen inklusive Erhebungszeitraum und -rhythmus. Die Auswertung des Datensatzes (deskriptive Statistiken, Befüllung ausgewählter Variablen etc.) kann am EBDC eingesehen und kopiert werden.

EBDC Business Investment Panel

Durch den halbjährlichen Turnus der ifo Investitionsumfrage (IU) ergibt sich für das EBDC Business Investment Panel eine etwas andere Struktur als beim EBDC Business Expectations Panel. Der Datensatz enthält keine Monatsangaben mehr, dafür aber eine Variable „season“, die angibt, ob die Daten aus der Frühjahrs- (1) oder Herbst- (2) Erhebung der IU bzw. aus einer Bilanz (99) stammen.

Der Datensatz ist nach BIP-ID (bip_company_id), Jahr (year), Erhebung (season) sortiert. Er liegt ebenfalls im Long-Format vor, d.h. jede Meldung ist über diese drei Variablen auch identifizierbar. Zusätzliche Identifikationsvariablen eines Unternehmens sind verschiedene Branchencodes, Beschäftigtengrößenklassen, Bundesland, Börsennotierung, Rechtsform etc.¹⁴ Die einzelnen Spalten des Panels enthalten wiederum die nach ihrer Funktion geordneten

¹³ Aus Gründen der Anonymisierung wurde die Bundesland-Angabe bei großen Unternehmen (> 10.000 Beschäftigte) gelöscht.

¹⁴ Wie zuvor wurde aus Gründen der Anonymisierung die Bundesland-Angabe bei großen Unternehmen (> 10.000 Beschäftigte) gelöscht.

Variablen: Identifikationsvariablen, Bilanz- und GUV- sowie ifo IU-Variablen. Die Variablenliste für das EBDC Business Investment Panel enthält neben einer Übersicht zu allen Variablen, deren Bedeutung und Besonderheiten die Fragen der einzelnen IU-Erhebungen inklusive Erhebungszeitraum und -rhythmus.

EBDC Business Innovation Panel

Die bereits erwähnte ifo Innovationsumfrage (INNO) bezieht sich, ebenso wie der ifo KU VG, auf einzelne Produkte eines Unternehmens. Zudem wird für die jährliche Innovationsumfrage¹⁵ ein Teil der Unternehmen aus der ifo Konjunkturumfrage für das Verarbeitende Gewerbe herangezogen, so dass diese beiden Datenquellen zusammen das EBDC Business Innovation Panel bilden. Abgefragt werden im INNO sowohl Innovationsaktivitäten und -ziele als auch Innovationsimpulse und -hemmnisse. Dabei werden Produkt- und Prozessinnovationen angesprochen, zudem gibt es wechselnde Sonderfragen in einzelnen Jahren. Die Innovationsumfrage wird als jährlicher Datensatz mit der Monatsangabe: „98“ an die Konjunkturumfrage Verarbeitendes Gewerbe gehängt, ebenso wie die Bilanzdaten mit der Monatsangabe „99“.

Erweiterungen

Alle EBDC Business Panels werden regelmäßig aktualisiert, wobei aus Anonymitätsgründen ein einjähriger Zeitabstand zum aktuellen Rand eingehalten wird. Die Aktualisierung bezieht sich generell sowohl auf die Zeitdimension als auch auf die Basis der enthaltenen Unternehmen. So werden auch zusätzliche, in die jeweiligen Adressdatenbanken neu aufgenommene Firmen jeweils per Record Linkage in das bestehende Panel integriert. Weitere Forschungsmöglichkeiten bieten Verknüpfungen der EBDC Business Panels sowie der ifo Panels untereinander. Die Auswertung aller Datensätze (deskriptive Statistiken, Befüllung ausgewählter Variablen etc.) kann am EBDC eingesehen und kopiert werden.

Vergabe der DOI's

Die fertigen Panels werden mit DOI's (Digital Objects Identifier) versehen und besitzen damit einen eindeutigen Identifikator. Er besteht aus einer für das EBDC vergebenen, festen Identifikationsnummer vor dem Schrägstrich und der Abkürzung des Panels auf Englisch sowie dessen Jahreszahl nach dem Schrägstrich.¹⁶ Damit wird die Nachvollziehbarkeit der Forschung im Rahmen „guter wissenschaftlichen Praxis“ unterstützt. Die DOI-Nummer identifiziert das jeweilige Panel mittels dieser eindeutigen, numerischen Zeichenfolge und kann über einen Link zu einer URL direkt aufgelöst werden.¹⁷ Diese Seite der EBDC-Homepage enthält alle wichtigen Angaben zu der speziellen Version des Panels, was die Kommunikation über empirische Datensätze erleichtert. Jede Änderung der Panels muss als neue Version

¹⁵ Die Erhebung wurde 2017 eingestellt und endet daher mit dem Jahr 2016.

¹⁶ Bsp. Business Expectations Panel (BEP): 10.7805/ebdc-bep-2016

¹⁷ Bsp. BEP: <http://dx.doi.org/10.7805/ebdc-bep-2016>

gespeichert und mit einem neuen DOI versehen werden. Dies ermöglicht auch nach Jahren noch eine eindeutige Identifizierung und ist damit Voraussetzung für die Archivierung der Forschung im EBDC-Archiv.

EBDC-Archiv

Das EBDC-Archiv bietet Nutzern des EBDC ebenso wie Forschern der wirtschaftswissenschaftlichen Fakultäten der Ludwig-Maximilians-Universität München die Möglichkeit, ihre Forschungsarbeit langfristig zu speichern. Voraussetzungen sind die Dokumentation der Prozeduren und die Hinterlegung der Metadaten. Die Panel-DOI sowie Angaben zur Forschung und zum Forscher werden in einem Übergabeprotokoll festgehalten und in einer Metadatenbank gespeichert. Das EBDC gewährleistet die Migration der archivierten Daten bzw. Emulation der alten Software. So kann auch noch nach Jahren im Einverständnis mit dem Primärforscher eine Sekundäranalyse durch interessierte Forscher erfolgen.

Mögliche Themen für die Forschung

Die Panels des EBDC ermöglichen vielfältige Forschung zu den unterschiedlichsten Themen. So wurden die Investitionseffekte von Haushaltskonsolidierungen analysiert oder Europa als Standort für die Ausgabe von Wertpapieren betrachtet, ebenso wie der regulatorische Wettbewerb innerhalb der Eurozone. Das Business Expectations Panel war die Basis für Forschung zur Kreditversorgung deutscher Firmen und deren Einfluss auf die Firmenerwartungen. Eine andere Arbeit setzte die Kreditbeschränkungen ins Verhältnis zu den jeweiligen Wechselkursen. Ein weiteres Projekt behandelte die Auswirkungen der Bankenregulierung auf die grenzüberschreitende Kreditvergabe. Aber auch Transfers innerhalb von Familienunternehmen konnten erörtert werden. Außerdem konnte die Wirkungsweise von Maßnahmen bei finanziellen Beschränkungen von Unternehmen mit Hilfe der Investitionsumfrage bzw. des Business Investment Panels ebenso ermittelt werden wie firmeneigene Investitionsinnovationen. Des Weiteren wurde mit Unterstützung externer Panels die Wirkungsweise öffentlicher Exportgarantien untersucht. Auf diese Art und Weise eröffnet sich auch für die Zukunft ein weites Forschungsfeld.

8.2.4 Weitere Wirtschaftsdaten im EBDC - das Forschungsdatenzentrum des Bayerischen Landesamtes für Statistik im EBDC

Statistiken aus den Bereichen Steuern und Wirtschaft/Finanzen sowie Energie- bzw. Umweltstatistiken werden zusätzlich im EBDC über die Zweigstelle des Forschungsdatenzentrums des Bayerischen Landesamtes für Statistik angeboten. Das Forschungsdatenzentrum hält derzeit Mikrodaten aus insgesamt über 100 amtlichen Statistiken zur wissenschaftlichen Nutzung bereit. Die räumliche Integration in das EBDC bietet für die Forschung ein Service-Plus durch die Bündelung des Angebots an empirischen Forschungsdaten.

Abbildung 8.1: Organigramm des EBDC

