

## **Die Verknüpfungsqualität der LIAB-Daten**

*Holger Alda*



## Inhaltsverzeichnis

1. Einleitung in die Problemstellung .....	3
2. Die Klassifizierung der Verknüpfungsqualität anhand von Toleranzintervallen .....	3
3. Ausfall- und Inkonsistenzanalyse .....	8
4. Zusammenfassung .....	10
Anhangstabellen .....	11

## Datenverfügbarkeit

Die in diesem Methodenreport beschriebenen Daten beziehen sich auf alle der Fachöffentlichkeit zugänglichen LIAB-Datensätze. Nähere Informationen dazu erhalten Sie auf der Internetseite: <http://fdz.iab.de/> unter der Rubrik „Integrierte Betriebs- und Personendaten“.

---

## 1. Einleitung in die Problemstellung

Der Beitrag klassifiziert die Verknüpfungsqualität der fragebogenbasierten Daten des IAB-Betriebspanels und der prozessproduzierten Daten der Beschäftigten-Leistungsempfänger-Historik-Datei. Diese aufeinander bezogenen Daten sind die Linked-Employer-Employee-Daten des IAB, LIAB. Im FDZ-Datenreport Nr. 1 werden die zur Verfügung stehenden Daten und Versionen beschrieben. Die Kenntnis dieses Reports erleichtert das Verständnis des Beitrags. Bei der Datenverknüpfungen des IAB-Betriebspanels mit der Beschäftigten-Leistungsempfänger-Historik können Probleme auftreten. Einige Betriebe beispielsweise können den Personendaten nicht zugeordnet werden. Bei anderen tauchen Konsistenzprobleme auf. Auf die Gründe der Inkonsistenzen kann dieser Beitrag nicht ausführlich eingehen. Ausfall- und Inkonsistenzdeterminanten werden im Verlauf des Beitrags exemplarisch für die Jahre 2000 und 2001 bestimmt.

Ein geeignetes Maß und damit Prüfgröße für die Übereinstimmung der Daten des IAB-Betriebspanels und den LIAB-Personendatenbanken ist die Gesamtanzahl an sozialversicherungspflichtig Beschäftigten. Die Angabe liegt als Befragungsdatum und als Aggregat der Personendaten auf die Betriebsebene vor. Zu beachten ist dabei, dass die Stichprobenziehung des IAB-Betriebspanels ein Jahr *vor* der Feldphase aus der Betriebsdatei des IAB erfolgt. Die aus dieser Datei ermittelte Soll-Zahl der sozialversicherungspflichtig Beschäftigten in Betrieben des Panels sollte also demnach mit den Ergebnissen aus den aggregierten Personendaten für den Betrieb auf Basis des *Vorjahres* übereinstimmen. Beide Angaben, sowohl die aktuelle Anzahl der sozialversicherungspflichtig Beschäftigten wie auch die des Vorjahres, ist in den IAB-Betriebspaneldaten enthalten.

## 2. Die Klassifizierung der Verknüpfungsqualität anhand von Toleranzintervallen

Ein bereits angesprochenes Problem ist, dass für manche Betriebe trotz vorliegender Betriebsnummer keine entsprechenden Personendaten gefunden werden. Für diese Ausfallgruppe der nicht gefundenen Betriebe und für solche, wo die Differenz der Fragebogenangaben zur Anzahl der sozialversicherungspflichtig Beschäftigten zu der in der Personendatenbank erheblich ist, erfolgt im nächsten Abschnitt eine Ausfallanalyse für die Querschnittsfälle der Jahre 2000 und 2001.

---

Bei dem Vergleich der Befragungsdaten des IAB-Betriebspanels mit den prozessproduzierten Daten der Beschäftigten-Leistungsempfänger-Historik sind Aufstockungsbetriebe für die aus dem IAB-Betriebspanel ausscheidenden Betriebe ein Problem. Die Sollzahl der sozialversicherungspflichtig Beschäftigten stimmt mit der Beschäftigtenzahl im Vorjahr überein, weil dort die Stichprobenziehung aus der Betriebsdatei des IAB erfolgt. Die befragten Betriebe geben zwar ihre Vorjahresanzahl an sozialversicherungspflichtig Beschäftigten an, aber es stehen keine entsprechenden Personendaten aus dem Vorjahr zur Verfügung, sofern der Betrieb erstmalig an der Befragung teilnimmt.

Für die Klassifizierung der Verknüpfungsqualität wird dann aus beiden Datenquellen (Befragungsdaten, prozessproduzierte Daten) die Angabe des aktuellen Befragungsjahres verwendet. Bei Längsschnittbetrieben hingegen werden die Information der Vorjahresbeschäftigung aus dem Organisationsfile des IAB-Betriebspanels und die der entsprechenden Vorjahresscheibe miteinander verglichen. Nach diesen Vergleichsverfahren wurden zusätzlich Längsschnittbetriebe als innerhalb der Toleranzgrenzen liegend klassifiziert, wenn sie aufgrund des Vorjahresvergleichs zwar (knapp) außerhalb, beim Vergleich der aktuellen Anzahl der sozialversicherungspflichtig Beschäftigten aus dem Fragebogen und der auf die Betriebsebene aggregierten Anzahl der Beobachtungen in der aktuellen Jahresscheibe innerhalb der definierten Toleranzgrenzen liegen. Aufgrund von Erinnerungslücken und daraus resultierenden Ungenauigkeiten bei der Angabe für einen vergangenen Stichtag werden (kleinere) Abweichungen bei den Angaben aus beiden Datenquellen zugelassen<sup>1</sup> und als prozentuale Abweichungen bestimmt. Bei einem solchen Vorgehen sollten je nach Betriebsgröße unterschiedliche Toleranzintervalle gelten, weil ein Ein- oder Austritt von Beschäftigten bezogen auf die Gesamtbelegschaft in kleineren Betrieben eine höhere prozentuale Veränderung hervorruft als in größeren.

Der Referenzzeitpunkt für die Fragen und der Beginn der Feldphase des IAB-Betriebspanels ist meist der 30. Juni eines jeden Befragungsjahres<sup>2</sup>. Die Feldphase der persönlich-mündlichen Betriebsinterviews endet etwa im September. Das genaue Antwortverhalten der Betriebe könnte sich auf den aktuellen Befragungszeitpunkt beziehen und nicht (exakt) auf den 30. Juni. Die kurzfristigen betrieblichen Personalschwankun-

---

<sup>1</sup> Zwar bezieht sich die Abfrage immer auf den 30. Juni, allerdings können natürlich nicht alle Interviews zeitnah und damit quasi gleichzeitig durchgeführt werden.

<sup>2</sup> Das gilt für alle Personalstrukturfragen, insbesondere die Anzahl der sozialversicherungspflichtig Beschäftigten.

---

gen sollten demnach anteilig nicht größer sein als die zugelassenen prozentualen Toleranzen für die Abweichung der Zahl der sozialversicherungspflichtig Beschäftigten laut IAB-Betriebspanel und Personendatenbank.

Die Tabelle 1 zeigt das Ergebnis einer monatlichen betrieblichen Fluktuationsanalyse. Sie wird hier nur für westdeutsche Betriebe ab fünf Beschäftigte vom Juni 1998 bis zum Juni 1999 durchgeführt<sup>3</sup>.

**Tabelle 1: Zusammenfassende Statistiken für die monatliche prozentuale sozialversicherungspflichtige Beschäftigungsveränderung in westdeutschen Betrieben ab fünf Beschäftigte von Juni 1998 bis Juni 1999 (gepoolte Daten)**

	Betriebe mit ....											
	bis zu 20 Beschäftigten ..... Monate später				20 bis 49 Beschäftigten ..... Monate später				50 bis 99 Beschäftigten ..... Monate später			
	einen	drei	sechs	elf	einen	drei	sechs	elf	einen	drei	sechs	elf
<b>Mittelwert</b>	-1,37	-2,53	-4,07	-6,79	-0,002	0,24	0,45	-0,39	-0,05	0,08	0,17	-1,78
<b>Standardabweichung</b>	20,73	31,75	40,94	47,19	5,86	10,86	16,64	22,53	5,66	7,32	10,43	15,94
<b>5-Prozent-Perzentil</b>	-16,67	-33,33	-50,00	-66,66	-6,89	-10,71	-14,71	-18,18	-4,68	-8,00	-12,00	-24,68
<b>95- Prozent-Perzentil</b>	14,28	25,00	33,33	35,29	7,41	12,5	17,02	21,43	4,91	8,33	13,33	18,84
<b>Anzahl Beobachtungen</b>	11564	9403	6237	1032	6234	5116	3394	567	3882	3182	2118	342
	Betriebe mit ....											
	100 bis 500 Beschäftigten ..... Monate später				mehr als 500 Beschäftigten ..... Monate später				alle Betriebe insgesamt ..... Monate später			
	einen	drei	sechs	elf	einen	drei	sechs	elf	einen	drei	sechs	elf
<b>Mittelwert</b>	-0,12	-0,17	-0,24	-1,17	-0,1	-0,17	-0,21	-1,17	-0,48	-0,81	-1,27	-2,86
<b>Standardabweichung</b>	5,66	10,69	14,16	20,04	2,36	5,21	7,39	12,24	12,34	19,39	25,41	30,68
<b>5-Prozent-Perzentil</b>	-3,34	-6,22	-9,73	-15,00	-2,22	-4,06	-6,58	-9,65	-6,89	-14,28	-22,22	-33,33
<b>95- Prozent-Perzentil</b>	3,54	6,69	10,18	13,34	2,17	4,02	6,29	8,61	6,45	14,28	20,00	25,00
<b>Anzahl Beobachtungen</b>	9190	7520	5013	835	5839	4771	3173	527	36709	29992	19935	3303

Quelle: Beschäftigten-Leistungsempfängerhistorik-Datei des Instituts für Arbeitsmarkt- und Berufsforschung; monatliche betriebliche Fluktuationsraten für etwa 3300 westdeutsche IAB-Betriebspanelbetriebe mit mindestens fünf Beschäftigten, ungewichtete Werte

Die Berechnung und Ausweisung der Ergebnisse ist für jede Spalte gleich. Für den nächsten Monat bzw. drei (sechs, elf) Monate (=i) später wurde die anteilige Beschäftigungsdifferenz als  $[(E_t - E_{t-i})/E_t] * 100$  mit E als Gesamtanzahl der sozialversicherungspflichtig Beschäftigten, t in Monaten und  $i=[1,3,6,11]$  berechnet<sup>4</sup>. Die ermittelte Differenz der Angaben wird aus der Anzahl der sozialversicherungspflichtig Beschäftigten laut IAB-Betriebspanel minus der aus der Beschäftigten-Leistungsempfänger-Historik

<sup>3</sup> Es gibt derzeit keine Anhaltspunkte, dass das ausgewiesene Ergebnis bei anderen Grundgesamtheiten und/oder Zeiträumen erheblich anders ausfällt.

<sup>4</sup> Die Ergebnisse nach einer Formel  $[(2*(E_t - E_{t-i})) / (E_t + E_{t-i})] * 100$  sind ähnlich. Die in der Tabelle 1 angewendete Formel produziert tendenziell höhere Werte.

berechnet und dann in Relation zu der Angabe aus dem IAB-Betriebspanel prozentuiert. Basierend auf den Ergebnissen der Tabelle 1 sind die Personalschwankungen an anteilige Beschäftigtendifferenz innerhalb von drei Monaten in den kleinsten Betrieben nur sehr selten größer als 25 Prozent. In den größten Betrieben liegen 90 Prozent der Beobachtungen in einem Abweichungsbereich von maximal 15 Prozent. Daher werden folgende Toleranzwerte für die Differenz der betrieblichen Fragebogenangaben und dem entsprechenden Ergebnis der Aggregation der Personendaten auf die Betriebs-ebene bei der späteren Berechnung einer Qualitätskontrollvariablen zugelassen:

- |  |            |
|--|------------|
| (1) Betriebe bis fünf Beschäftigte:          | 40 Prozent |
| (2) Betriebe mit fünf bis 19 Beschäftigten : | 30 Prozent |
| (3) Betriebe mit 20 bis 100 Beschäftigten:   | 20 Prozent |
| (4) Betriebe mit mehr als 100 Beschäftigten: | 10 Prozent |

Mit den Toleranzgrenzen werden betriebliche Personalschwankungen zu mehr als neunzig Prozent eingeschlossen. Solche Schwankungen der Mitarbeiterzahl können demnach bei der vorgenommenen Definition als Quelle für nicht deckungsgleiche Angaben aus dem IAB-Betriebspanel und der Beschäftigten-Leistungsempfänger-Historik ausgeschlossen werden. Wenn Betriebe nicht im Toleranzintervall liegen, kann die gleichzeitige Verwendung von Personen- und IAB-Betriebspaneldaten problematisch sein<sup>5</sup>. Die prozentuale Abweichung ist in einer zusätzlichen Variable dokumentiert, die auf Wunsch externen Forschern bei ihrem Gastaufenthalt im IAB zur Verfügung gestellt wird<sup>6</sup>.

Bei der nachfolgenden Qualitätsklassifikation in Tabelle 2 wird diese Information über die prozentuale Abweichung beispielsweise genutzt, um Betriebe mit Abweichungen von mehr als 100 Prozent gesondert auszuweisen. Die Ausweisung der Ergebnisse erfolgt für die gültigen Querschnittsfälle der jeweiligen Jahre.

---

<sup>5</sup> Wenn ein Betrieb beispielsweise Fragebogenangaben für 1000 Beschäftigte macht (z.B. Investitionssumme), kann die simultane Nutzung von und Verknüpfung mit nur 100 vorliegenden Personendatensätzen Irritationen auslösen.

<sup>6</sup> Die Variable eignet sich zur Filterführung.

---

**Tabelle 2: Klassifizierung der Verknüpfungsqualität von IAB-Betriebspaneldaten und den entsprechenden Daten aus der Beschäftigten-Leistungshistorikdatei des IAB (alle Befragungsjahre 1993 bis 2001)**

**a) Westdeutschland**

<b>absolute Anzahl westdeutsche Betriebe</b>	<b>1993</b>	<b>1994</b>	<b>1995</b>	<b>1996</b>	<b>1997</b>	<b>1998</b>	<b>1999</b>	<b>2000</b>	<b>2001</b>
im Toleranzbereich	-	2665	2433	2369	2333	2345	2493	2342	4079
nach Fragebogenangaben im Toleranzbereich	2769	357	547	556	567	846	775	2968	2503
keine Verknüpfung mit Individualdaten möglich	76	55	117	147	222	178	155	292	290
nicht im Toleranzbereich	1270	982	879	812	866	817	842	2192	2172
nicht im Toleranzbereich, Kleinstbetrieb*	111	50	59	113	51	67	74	223	404
Abweichung mehr als 100 Prozent	39	13	61	32	63	36	88	399	400
<b>insgesamt</b>	<b>4265</b>	<b>4166</b>	<b>4096</b>	<b>4029</b>	<b>4102</b>	<b>4289</b>	<b>4427</b>	<b>8416</b>	<b>9848</b>
<b>anteilig westdeutsche Betriebe</b>	<b>1993</b>	<b>1994</b>	<b>1995</b>	<b>1996</b>	<b>1997</b>	<b>1998</b>	<b>1999</b>	<b>2000</b>	<b>2001</b>
im Toleranzbereich	-	64%	59%	59%	57%	55%	56%	28%	41%
nach Fragebogenangaben im Toleranzbereich	65%	9%	13%	14%	14%	20%	18%	35%	25%
keine Verknüpfung mit Individualdaten möglich	2%	1%	3%	4%	5%	4%	4%	3%	3%
nicht im Toleranzbereich	30%	24%	21%	20%	21%	19%	19%	26%	22%
nicht im Toleranzbereich, Kleinstbetrieb*	3%	1%	1%	3%	1%	2%	2%	3%	4%
Abweichung mehr als 100 Prozent	1%	0%	1%	1%	2%	1%	2%	5%	4%
<b>insgesamt</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

\* mit weniger als fünf Beschäftigten

**b) Ostdeutschland**

<b>absolute Anzahl ostdeutsche Betriebe</b>	<b>1996</b>	<b>1997</b>	<b>1998</b>	<b>1999</b>	<b>2000</b>	<b>2001</b>
im Toleranzbereich	-	2882	2999	3153	3494	3397
nach Fragebogenangaben im Toleranzbereich	2740	883	922	1124	817	1036
keine Verknüpfung mit Individualdaten möglich	103	151	128	161	167	193
nicht im Toleranzbereich	1260	705	709	733	765	759
nicht im Toleranzbereich, Kleinstbetrieb*	147	30	56	71	171	191
Abweichung mehr als 100 Prozent	63	97	90	93	101	112
<b>insgesamt</b>	<b>4313</b>	<b>4748</b>	<b>4904</b>	<b>5335</b>	<b>5515</b>	<b>5688</b>
<b>anteilig ostdeutsche Betriebe</b>	<b>1996</b>	<b>1997</b>	<b>1998</b>	<b>1999</b>	<b>2000</b>	<b>2001</b>
im Toleranzbereich	-	61%	61%	59%	63%	60%
nach Fragebogenangaben im Toleranzbereich	64%	19%	19%	21%	15%	18%
keine Verknüpfung mit Individualdaten möglich	2%	3%	3%	3%	3%	3%
nicht im Toleranzbereich	29%	15%	14%	14%	14%	13%
nicht im Toleranzbereich, Kleinstbetrieb*	3%	1%	1%	1%	3%	3%
Abweichung mehr als 100 Prozent	1%	2%	2%	2%	2%	2%
<b>insgesamt</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

\* mit weniger als fünf Beschäftigten

Quelle: Personendaten LIAB-Querschnittsmodell, Version 1 und IAB-Betriebspanel 1993 bis 2001, Organisationsdatenfile 2002

Auch die Betriebe mit bis zu fünf Beschäftigten werden gesondert ausgewiesen, weil bei ihnen nur ein Beschäftigungsein- oder -austritt zu einer wenigstens 20-prozentigen Beschäftigungsveränderung führt. Damit wird ein Verlassen des Toleranzbereichs

wahrscheinlicher und die hier beschriebenen Einordnungskriterien könnten bei den kleinsten Betrieben problematisch sein. Wie die Tabelle 2 zeigt, sind diese Betriebe aber für das Gesamtergebnis nicht sonderlich bedeutsam. Mit jährlichen Schwankungen liegen durchschnittlich etwa drei von vier IAB-Betriebspanelbetrieben in dem definierten Toleranzintervall, sofern dazu die ersten beiden ausgewiesenen Kategorien gezählt werden. Die fehlende Zuordnung zu den Personendaten ist ein eher marginales Problem und, wie weiterführende Analysen zeigen, fast ausschließlich auf kleinste Betriebe zurückzuführen. Die letzten drei Kategorien beschreiben Betriebe, die nicht in den definierten Toleranzintervallen liegen.

Eine numerische Variable mit den in Tabelle 2 ausgewiesenen Informationen und eine weitere, die bei Betrieben außerhalb des Toleranzbereichs über die festgestellte prozentuale Abweichung informiert, ist auf Anfrage im FDZ verfügbar. Der entsprechende Datensatz enthält die systemfreien Betriebsidentifikatoren der Querschnittsfälle 1993 bis 2001 aus dem IAB-Betriebspanel, eine Jahresinformation sowie die beiden interessierenden Variablen der Klassifizierungsqualität und der prozentualen Abweichung. Die hier benannten Informationen sind also kompatibel abrufbar für alle Versionen der LIAB-Datenmodelle. Der Datensatz umfasst insgesamt mehr als 78 Tsd. Beobachtungsjahre der Betriebe.

Im nächsten Abschnitt werden zwei Ausprägungen der Klassifizierung aus Tabelle 2 je mit Betrieben verglichen, die innerhalb des Toleranzbereichs liegen. Es wird untersucht, wie sich sowohl in den Personendaten nicht gefundene Betriebe als auch Betriebe außerhalb der zugelassenen Toleranzintervalle von denen *innerhalb* der Toleranzintervalle unterscheiden.

### **3. Ausfall- und Inkonsistenzanalyse**

Die beiden Zielkategorien „nicht gefundene Betriebe“ und „Betrieb außerhalb des Toleranzbereichs“ werden mit Betrieben innerhalb des Toleranzintervalls verglichen. Betriebe der Zielkategorien werden je mit einer „Eins“, Betriebe innerhalb des Toleranzbereichs mit einer „Null“ kodiert. Die Analyse beschränkt sich auf Querschnittsfälle und die Jahre 2000 und 2001 (gepoolte Daten). Als Prediktoren für den Ausfall bzw. die Abweichung werden je die folgende Informationen aus dem IAB-Betriebspanel verwendet:

---



- Betriebsalter (drei Dummies für junge (nach 1995 gegründet), mittlere (zwischen 1990 und 1994) und alte Betriebe (vor 1990 gegründet))
- Rechtsform (sechs Dummies entsprechend den Kategorien des Fragebogens)
- Kammerzugehörigkeit (vier Dummies)
- Eigentumsverhältnisse (vier Dummies)
- Stellung im Unternehmensverbund (vier Dummies)
- siedlungsstruktureller Typ (drei Dummies städtisches Kern- und Randgebiet sowie ländliche Regionen)
- Betriebsgrößen- und Wirtschaftszweigdummies

In der Anhangstabelle 1 befindet sich die zusammenfassende Statistik für die aufgenommenen Variablen und Referenzgruppen. Geschätzt wurden Probit-Regressionen. Die Anhangstabellen weisen den jeweiligen marginalen Effekt aus. Es gibt derzeit keine Hinweise darauf, dass die Ergebnisse durch prinzipiell beobachtete, aber durch das Poolen der Daten unbeobachtbare betriebliche Heterogenitäten drastisch verzerrt sind<sup>7</sup>. Die Ergebnisse sind denen in den Anhangstabellen 2a und 2b ausgewiesenen sehr ähnlich. Allerdings ist die Hypothese unkorrelierter Beobachtungen auf der Betriebsebene für die Probit-Schätzung lt. Anhang 2b (Betriebe außerhalb der Toleranzintervalle) abzulehnen. Betriebe liegen also eher systematisch außerhalb des Toleranzintervalls.

Da fast ausschließlich Dummies für die Regressionen verwendet werden, ist für jeden Dummy-Set der Wert des jeweiligen (partiellen) Likelihood-Ratio-Tests dokumentiert. Bei fettgedruckten Werten ist die Nullhypothese (der Dummy-Set verbessert nicht die Vorhersagekraft des Modells) abzulehnen (Zahl in Klammern: Anzahl der Freiheitsgrade). Nicht gefundene Betriebe (Schätzung Anhang 2a) sind meist kleinere Betriebe<sup>8</sup>. Ihre Stellung im Unternehmensverbund ist signifikant häufiger die einer Filiale. Auch ein Eigentümerwechsel hat einen positiven Effekt auf diese Ausfallgruppe. Betrieben, die von Personen mit Vollhaftung geführt werden, sind signifikant häufiger keine Personendaten zuzuordnen. Die Kammerzugehörigkeit hingegen senkt das „Risiko“, keine Entsprechung in den Personendaten zu haben. Relativ bedeutsam sind In- und Out-

---

<sup>7</sup> Signifikanzniveaus und  $\chi^2$ -Tests werden nicht entscheidend beeinflusst, wenn von einem Modell mit gepoolten Daten auf ein Panelregressionsmodell übergegangen wird.

sourcing-Aktivitäten (Ausnahme Ausgründung von Betriebsteilen) und die Betriebsgröße. Vor allem kleinere Betriebe haben häufiger keine Entsprechung in den Personendaten.

Betriebe, die außerhalb des wie in Abschnitt 2 definierten Toleranzintervalls liegen (Schätzung Anhang 2b), unterscheiden sich von Betrieben innerhalb dieser Toleranzen in folgenden Charakteristika. Umgekehrt zur Schätzung 2a wirkt die Betriebsgröße. Am häufigsten liegen die größten Betriebe außerhalb der Toleranzgrenzen. Die In- und Outsourcing-Aktivitäten haben einen eigenständigen signifikant positiven Einfluss. Auch das Betriebsalter kann darüber entscheiden, ob ein bestimmter Betrieb innerhalb oder außerhalb des Toleranzintervalls liegt. Bei jungen Betrieben sind Abweichungen bei der Zahl der sozialversicherungspflichtig Beschäftigten lt. Fragebogenangabe und Personendaten am wahrscheinlichsten. Zuordnungsprobleme treten auch auf, wenn es sich nicht um einen Einzelbetrieb handelt, sondern er Teil eines Unternehmens ist. Die Eigentumsverhältnisse hingegen beeinflussen die Ausfallwahrscheinlichkeit lt. Anhang 2b nicht signifikant. Betriebe in ländlichen Regionen haben häufiger als solche in städtischen Gebieten größere Diskrepanzen zwischen Fragebogenangaben und Personendatenbankergebnissen.

#### **4. Zusammenfassung**

Der Beitrag beschäftigt sich mit Verknüpfungsproblemen der Daten des IAB-Betriebspanels mit den prozessproduzierten Daten der Beschäftigten-Leistungsempfänger-Historik in den LIAB-Versionen, die über das Forschungsdatenzentrum der BA im IAB externen Forschern zugänglich sind (vgl. FDZ-Datenreporte Nr. 2 bis 4). Für die Jahre 2000 und 2001 wurde analysiert, welche Betriebe in der Beschäftigten-Leistungsempfänger-Historik nicht gefunden werden und bei welchen Betrieben die Abweichungen zwischen den Beschäftigtenangaben aus beiden Datenquellen erheblich voneinander abweichen. Kenntnisse über die Verknüpfungsqualität der Daten sollen die Arbeit mit den LIAB-Daten vor Ort schon im Vorfeld erleichtern. Außerdem stellt das Forschungsdatenzentrum auch LIAB-Versionen bereit, die Verknüpfungsprobleme im Vorhinein vermeiden, indem nur Betriebe ausgewählt werden,

---

<sup>8</sup> Die Fallzahl für diese Schätzung ist wesentlich geringer als in Anhang 2b, weil die je andere Ausfallgruppe (für die Schätzung 2a Betriebe, die nicht im Toleranzbereich liegen) aus der Analyse ausgeschlossen wird.

---

die innerhalb des Toleranzintervalls liegen (wie beispielsweise die erste Version des LIAB-Längsschnittmodells) und damit explizit Bezug auf diesen Beitrag nehmen.

## Anhangstabellen

**Anhangstabelle 1: Zusammenfassende Statistiken für die Probit-Regressionen der Anhangstabellen 2a und 2b**

<i>Variablen</i>	<i>Beobach- tungen</i>	<i>Mittelwert</i>	<i>Standard- Abweichung</i>
Betrieb existiert lt. BLH nicht (ja=1)	21239	0,05	0,209
Betrieb außerhalb Toleranz (ja=1)	28495	0,28	0,450
<b>Betriebsgröße</b>			
bis 20 Beschäftigte	29467	0,42	0,492
20 bis 49 Beschäftigte	29467	0,18	0,383
50 bis 99 Beschäftigte	29467	0,11	0,314
100 bis 499 Beschäftigte	29467	0,20	0,403
500 und mehr Beschäftigte (Referenz)	29467	0,09	0,289
Anzahl sonstiger Beschäftigter (log)	29467	0,58	1,073
<b>Outsourcing-Aktivitäten</b>			
Schließung von Betriebsteilen	29358	0,02	0,134
Auslagerung von Betriebsteilen	29358	0,02	0,153
Ausgründung von Betriebsteilen	29358	0,01	0,096
keine Outsourcing-Aktivität (Referenz)	29358	0,95	0,221
Insourcing-Aktivität (ja=1)	29160	0,04	0,182
Betriebsstandort Westdeutschland (ja=1)	29467	0,62	0,485
<b>Betriebsalter</b>			
alter Betrieb (vor 1990)	29239	0,81	0,477
Betrieb mittleren Alters (1990 - 1996)	29239	0,14	0,350
junger Betrieb (1997 u. später) (Referenz)	29239	0,05	0,221
<b>Status der Gründung</b>			
Neugründung (Referenz)	29467	0,93	0,443
Eigentümerwechsel	29467	0,03	0,172
Ausgründung aus bestehendem Betrieb	29467	0,04	0,197
<b>Kammerzugehörigkeit</b>			
keine Kammerzugehörigkeit (Referenz)	29040	0,24	0,42
Mitglied Handelskammer	29040	0,16	0,371

(Fortsetzung nächste Seite)

<i>Variablen</i>	<i>Beobach- tungen</i>	<i>Mittelwert</i>	<i>Standard- Abweichung</i>
Mitglied Industrie- und Handelskammer	29040	0,53	0,498
Mitglied andere Kammer	29040	0,07	0,259
<b>Betriebsstellung im Unternehmen</b>			
Einzelunternehmen (Referenz)	28867	0,70	0,424
Zentrale	28867	0,09	0,296
Filiale	28867	0,17	0,378
Mittelinstanz	28867	0,04	0,191
westdeutsches Eigentum (Referenz)	28763	0,65	0,498
ostdeutsches Eigentum	28763	0,22	0,416
ausländisches Eigentum	28763	0,05	0,216
kein Mehrheitseigentümer	28763	0,06	0,229
unbekannter Eigentümer	28763	0,02	0,138
<b>Rechtsform</b>			
Person mit Vollhaftung (Referenz)	29096	0,23	0,27
Personengesellschaft	29096	0,08	0,27
GmbH	29096	0,44	0,495
Kapitalgesellschaft	29096	0,05	0,209
Körperschaft u.ä.	29096	0,13	0,338
Sonstige (Vereine, Genossenschaften u.ä.)	29096	0,07	0,256
<b>siedlungsstruktureller Typ</b>			
ländliches Gebiet (Referenz)	29467	0,52	0,499
städtisches Randgebiet	29467	0,23	0,416
städtisches Kerngebiet	29467	0,25	0,434
andere Betriebseinheit befragt (ja=1)	29467	0,03	0,167
Jahr 2000 (ja=1)	29467	0,47	0,499
<b>Branchendummies</b>			
Verarbeitendes Gewerbe I (Referenz) (Nahrung/Grundstoff/Agrar)	29467	0,17	0,377
Verarbeitendes Gewerbe II (Investitionsgüter, Bau)	29467	0,23	0,424
Handel/Verkehr/Nachrichten	29467	0,17	0,379
Finanzdienstleistungen	29467	0,03	0,161
unternehmensnahe Dienstleistungen	29467	0,10	0,296
sonstige Dienstleistungen	29467	0,20	0,397
öffentlicher Sektor	29467	0,10	0,297

Quellen: LIAB-Querschnittsmodell, Version 1 sowie Datensätze und Organisationsdatenfile des IAB-Betriebspanels für die Jahre 2000 und 2001

**Tabelle 2a: Regressionsergebnisse für Betriebe, die nicht mit Personendaten aus der BLH verknüpft werden können (=1) und Betriebe, die innerhalb der definierten Toleranzgrenzen liegen (=0) (Probit-Schätzung mit marginalen Effekten)**

Prediktoren	marginaler Effekt	Mittelwert	t-wert
<b>Betriebsgröße</b>		<b>193,24** (4)</b>	
bis 20 Beschäftigte	0,062	0,42	7,10
20 bis 49 Beschäftigte	0,016	0,21	1,91
50 bis 99 Beschäftigte	0,015	0,12	1,72
100 bis 499 Beschäftigte	-0,012	0,19	-0,16
Anzahl sonstiger Beschäftigter (log)	-0,003	0,41	-1,52
<b>Outsourcing-Aktivitäten</b>		<b>22,07** (3)</b>	
Schließung von Betriebsteilen	0,038	0,02	3,16
Auslagerung von Betriebsteilen	0,031	0,02	2,89
Ausgründung von Betriebsteilen	0,023	0,01	1,37
Insourcing-Aktivität (ja=1)	0,041	0,03	4,20
Betriebsstandort Westdeutschland (ja=1)	0,005	0,57	1,30
<b>Betriebsalter</b>		<b>0,52 (2)</b>	
alter Betrieb (vor 1990)	-0,006	0,64	-1,51
Betrieb mittleren Alters (1990 - 1996)	0,0002	0,15	0,05
<b>Status der Gründung</b>		<b>12,83** (2)</b>	
Eigentümerwechsel	0,025	0,03	3,30
Ausgründung aus bestehendem Betrieb	-0,011	0,04	-1,86
<b>Kammerzugehörigkeit</b>		<b>8,21* (3)</b>	
Mitglied Handelskammer	-0,012	0,18	-2,48
Mitglied Industrie- und Handelskammer	-0,009	0,54	-1,90
Mitglied andere Kammer	-0,015	0,07	-3,00
<b>Betriebsstellung im Unternehmen</b>		<b>38,33** (3)</b>	
Zentrale	-0,009	0,08	-1,61
Filiale	0,024	0,17	5,41
Mittelinstanz	0,017	0,03	1,66
<b>Eigentumsverhältnisse</b>		<b>3,26 (4)</b>	
ostdeutsches Eigentum	-0,005	0,25	-1,06
Ausländisches Eigentum	-0,002	0,05	-0,36
kein Mehrheitseigentümer	-0,002	0,05	-0,38
unbekannter Eigentümer	0,013	0,02	1,36

(Fortsetzung nächste Seite)

Prediktor	marginaler Effekt	Mittelwert	t-Wert
<b>Rechtsform</b>		<b>20,39** (5)</b>	
Personengesellschaft	0,008	0,08	1,50
GmbH	-0,001	0,47	-0,31
Kapitalgesellschaft	0,005	0,04	0,65
Körperschaft u.ä.	-0,017	0,11	-2,81
Sonstige (Vereine, Genossenschaften)	-0,014	0,07	-2,39
<b>siedlungsstruktureller Typ</b>		<b>11,30** (2)</b>	
städtisches Randgebiet	0,012	0,23	3,49
städtisches Kerngebiet	0,006	0,27	1,88
andere Betriebseinheit befragt (ja=1)	0,0005	0,03	0,07
Jahr 2000 (ja=1)	-0,002	0,47	-0,70
Branchendummies	<b>ja***</b>		<b>18,96** (6)</b>
Pseudo R <sup>2</sup>		0,065	
Wald-Statistik		462,73** (39)	
Anzahl Beobachtungen		20017	

Quellen: LIAB-Querschnittsmodell, Version 1 sowie Datensätze und Organisationsdatenfile des IAB-Betriebspanels für die Jahre 2000 und 2001, Erläuterungen zum Schätzansatz und den Werten befinden sich im Text, \*, \*\* bedeuten Signifikanz auf einem Niveau von  $\alpha < 0.05$  bzw. 0.01

**Tabelle 2b: Regressionsergebnisse für Betriebe, die außerhalb (=1) und Betriebe, die innerhalb der definierten Toleranzgrenzen (=0) liegen (Probit-Schätzung mit marginalen Effekten)**

Prediktoren	marginaler Effekt	Mittelwert	t-wert
<b>Betriebsgröße</b>		<b>499,05** (4)</b>	
bis 20 Beschäftigte	-0,084	0,41	-6,97
20 bis 49 Beschäftigte	0,187	0,18	-18,05
50 bis 99 Beschäftigte	-0,16	0,11	-14,82
100 bis 499 Beschäftigte	-0,1	0,21	-9,76
Anzahl sonstiger Beschäftigter (log)	0,079	0,57	26,90
<b>Outsourcing-Aktivitäten</b>		<b>40,03** (3)</b>	
Schließung von Betriebsteilen	0,054	0,02	2,50
Auslagerung von Betriebsteilen	0,067	0,02	3,43
Ausgründung von Betriebsteilen	0,098	0,01	3,04

(Fortsetzung nächste Seite)

Prediktoren	marginaler Effekt	Mittelwert	t-wert
Insourcing-Aktivität (ja=1)	0,097	0,03	6,01
Betriebsstandort Westdeutschland (ja=1)	0,092	0,61	10,47
<b>Betriebsalter</b>		<b>119,85** (2)</b>	
Alter Betrieb (vor 1990)	-0,089	0,64	-10,26
Betrieb mittleren Alters (1990 – 1996)	-0,039	0,14	-3,81
<b>Status der Gründung</b>		0,39 (2)	
Eigentümerwechsel	0,004	0,03	0,28
Ausgründung aus bestehendem Betrieb	-0,008	0,04	-0,54
<b>Kammerzugehörigkeit</b>		<b>12,83** (3)</b>	
Mitglied Handelskammer	-0,037	0,17	-3,06
Mitglied Industrie- und Handelskammer	-0,012	0,54	-1,22
Mitglied andere Kammer	-0,024	0,07	-1,88
<b>Betriebsstellung im Unternehmen</b>		<b>78,82** (3)</b>	
Zentrale	0,071	0,1	6,91
Filiale	0,035	0,16	4,15
Mittelinstanz	0,078	0,04	4,72
<b>Eigentumsverhältnisse</b>		<b>3,73 (4)</b>	
ostdeutsches Eigentum	-0,009	0,23	-0,91
ausländisches Eigentum	-0,01	0,05	-0,75
kein Mehrheitseigentümer	0,019	0,06	1,51
Unbekannter Eigentümer	0,025	0,02	1,22
<b>Rechtsform</b>		<b>219,73** (5)</b>	
Personengesellschaft	-0,067	0,08	-6,00
GmbH	-0,091	0,44	-10,48
Kapitalgesellschaft	-0,054	0,05	-3,52
Körperschaft u.ä.	-0,071	0,12	-4,81
Sonstige (Vereine, Genossenschaften )	-0,112	0,07	-8,51
<b>siedlungsstruktureller Typ</b>		<b>32,91** (2)</b>	
städtisches Randgebiet	-0,021	0,22	-2,95
städtisches Kerngebiet	-0,032	0,25	-4,62
andere Betriebseinheit befragt (ja=1)	0,018	0,03	1,13
Jahr 2000 (ja=1)	0,02	0,47	3,18
Branchendummies	ja**	<b>112,97** (6)</b>	

(Fortsetzung nächste Seite)

<b>Prediktoren</b>	<b>marginaler Effekt</b>	<b>Mittelwert</b>	<b>t-wert</b>
Pseudo R <sup>2</sup>		0,101	
Wald-Statistik		2771,64** (39)	
Anzahl Beobachtungen		26697	

Quellen: LIAB-Querschnittsmodell, Version 1 sowie Datensätze und Organisationsdatenfile des IAB-Betriebspanels für die Jahre 2000 und 2001, Erläuterungen zum Schätzansatz und den Werten befinden sich im Text, \*, \*\* bedeuten Signifikanz auf einem Niveau von  $\alpha < 0.05$  bzw. 0.01



## Impressum

### **FDZ** *Methodenreport*

Nr. 1/2005

#### **Herausgeber**

Forschungsdatenzentrum (FDZ)  
der Bundesagentur für Arbeit  
Regensburger Str. 104  
D-90478 Nürnberg

#### **Redaktion**

Dagmar Herrlinger, Annette Kohlmann

#### **Technische Herstellung**

Dagmar Herrlinger

#### **Rechte**

Nachdruck – auch auszugsweise – nur mit  
Genehmigung des FDZ gestattet

#### **Bezugsmöglichkeit**

Volltext-Download dieses Datenreports unter  
[http://doku.iab.de/fdz/reporte/2005/MR\\_1.pdf](http://doku.iab.de/fdz/reporte/2005/MR_1.pdf)

#### **FDZ im Internet**

<http://fdz.iab.de/>

#### **Rückfragen zum Inhalt an**

Holger Alda, Forschungsdatenzentrum (FDZ) der  
Bundesagentur für Arbeit (BA) im Institut für Arbeits-  
markt- und Berufsforschung (IAB), Regensburger  
Strasse 104, 90478 Nürnberg,  
E-Mail: [holger.alda@iab.de](mailto:holger.alda@iab.de)