

**Holger Alda**

**Betriebliche Arbeitsnachfrage und Beschäftigung – Methodenpapier**

**soeb-Arbeitspapier 2008-6**

Forschungsverbund Berichterstattung zur sozioökonomischen Entwicklung der  
Bundesrepublik Deutschland: Arbeit und Lebensweisen  
Internet: <http://www.soeb.de>  
Koordination: Soziologisches Forschungsinstitut (SOFI)  
Friedländer Weg 31  
D-37085 Göttingen  
Projektleitung: Dr. Peter Bartelheimer

## Inhaltsverzeichnis

1. Beschreibung der Lohnverteilungsmaße .....	3
2. Konstruktion und Eigenschaften der Variablen für die Betriebstypisierung .....	4
3. Vorgehen bei der Korrespondenz- und Clusteranalyse.....	6
4. Regressionsanalyse.....	8
5. Ostdeutsche Betriebsprofile .....	12
Tab. 5.1: Gruppenmittelwerte der Betriebstypen des Niedriglohnsegments .....	12
Tab. 5.2: Gruppenmittelwerte der Betriebstypen des mittleren Segments.....	13
Tab. 5.3: Gruppenmittelwerte des Betriebstyps im Hochlohnsegment.....	14
Übersicht 5.1: Ausprägungen der sozioökonomischen Profile in den ostdeutschen Betriebstypen des Jahres 2004 .....	16
5.1 Abdeckungsgrad der Typen über ostdeutsche Betriebe und Beschäftigte .....	16
Tab. 5.5: Abdeckungsgrad der ostdeutschen Typen im Jahr 2004 über Betriebe, Erwerbstätige und sozialversicherungspflichtig Beschäftigte (SvB).....	17
6. Anhänge .....	17
6.1 Anhang A1: Regressionsergebnisse und Regressionsdiagnostik für Westdeutschland. 17	
6.2. Anhang A2: Vorhergesagte Wahrscheinlichkeiten für die Segmente der Betriebstypen in Abhängigkeit von ihrem Exportanteil, sofern sie einen ausländischen Mehrheitseigentümer haben .....	21
6.3 Anhang A3: Entwicklung der Anzahl aller Erwerbstätigen und der sozialversicherungspflichtigen Beschäftigung auf der Firmenebene in Deutschland im Zeitraum 1993 bis 2004.....	22
7. Literatur.....	23

## **0. Aufbau dieses Arbeitspapiers**

Dieses Arbeitspapier dokumentiert die im zweiten Bericht zur sozioökonomischen Entwicklung in Deutschland für das Kapitel 13 „Betriebliche Arbeitsnachfrage und Beschäftigung“ vorgenommenen Datenaufbereitungsschritte, das methodische Vorgehen und empirische Ergebnisse für im Haupttext genannte, aber nicht abgedruckte Tabellen und Abbildungen. Die Datenaufbereitungsschritte werden angesprochen, weil die Arbeit gerade mit den Prozessdaten des IAB voraussetzungsvoll ist und in einigen Fällen Setzungen erforderlich sind, deren Ausgestaltung die daraus hervorgehenden empirischen Ergebnisse beeinflussen. Außerdem wurden aus den Prozessdaten einige relative Konstrukte generiert, etwa die „besten“ und „schlechten“ Verdienste.

Anschließend geht das Arbeitspapier kurz auf die Konstruktionsprinzipien einzelner Variablen ein, sofern dies für die im dritten Abschnitt beschriebene Vorgehensweise bei den Faktoren- und Clusteranalysen eine gewisse Bedeutung hat. Die letztlich aus der Clusteranalyse hervorgehende Betriebstypologie ist eine kategoriale Variable mit zehn Ausprägungen. Für die im Haupttext durchgeführten Regressionen werden die entsprechenden modelltheoretischen Annahmen im vierten Abschnitt dieses Arbeitspapiers besprochen.

Im fünften Abschnitt befinden sich Tabellen zur ostdeutschen Betriebstypologie, auf die der Haupttext an mindestens einer Stelle Bezug nimmt. Demgegenüber enthält der letzte Abschnitt (der Anhang) alle sonstigen Dokumentationen, die für die Aussagen des Haupttextes oder dieses Arbeitspapiers benötigt werden.

## **1. Beschreibung der Lohnverteilungsmaße**

Um zu sehen, inwiefern sich individuelle Lohnungleichheiten in betrieblichen Strukturen widerspiegeln, wurden unter anderem auf der Aggregatebene die besten und schlechtesten Verdienste als oberstes und unterstes Quintil der Lohnverteilung in der Gesamtwirtschaft definiert und anschließend die entsprechend klassifizierten Personen als Anteile auf der Betriebsebene berechnet.

Lohnzahlungen können nicht nur aufgrund von beobachteten beruflichen Qualifikationen und weiterer Beschäftigtencharakteristika erfolgen. Auch informell erworbene (oder mit der Datenbasis nicht messbare und beruflich verwertbare Zertifikate) Kenntnisse und Kompetenzen sowie Aspekte von Macht, die mit den unbeobachteten Lohneffekten gemessen werden, sind eine relevante Komponente zur Erklärung der Einkommensungleichheiten auf dem Arbeitsmarkt. Da bei der Datenlage nicht genau geklärt werden kann, worauf genau diese Lohneffekte

te basieren, kommt ein Verteilungsmaß zur Messung einer möglichen Betriebssegregation zur Anwendung: analog zu den besten und schlechtesten Verdiensten werden zunächst die Personen identifiziert, die aufgrund der Verteilung in der Gesamtwirtschaft – also ohne Rücksicht auf die betriebliche Ebene – im obersten oder untersten Quintil der Verteilung der unbeobachteten Lohneffekte liegen. Mit der Bestimmung des Anteils der jeweiligen Personen auf der Betriebsebene wird dann gemessen, ob und in welchem Ausmaß Betriebe derartige Lohnaufschläge zahlen, also ob sie lediglich die „klassischen“ Bildungs- und Qualifikationszertifikate entlohnen (unterstes Quintil der unbeobachteten Lohneffekte) oder ob sie nennenswerte zusätzliche Lohnzahlungen für weniger formelle Kenntnisse und Kompetenzen leisten (oberstes Quintil der unbeobachteten Lohneffekte).

## **2. Konstruktion und Eigenschaften der Variablen für die Betriebstypisierung**

Löhne und Gehälter der sozialversicherungspflichtig Beschäftigten werden in den IAB-Personendaten als Bruttomonatsverdienste tagesgenau erfasst. Die Lohnangaben sind an der unteren und oberen Beitragsbemessungsgrenze zensiert. Da auf das LIAB-Querschnittmodell zurückgegriffen wird, können Lohnänderungen nur für Beschäftigte berechnet werden, die in einem IAB-Betriebspanelbetrieb arbeiten, der in den beiden Jahren 2003 und 2004 an der Befragung teilgenommen hat. Mit Ausnahme der Lohnänderung zum Vorjahr werden für alle anderen Lohn- und Gehaltsvariablen inklusive der Streuungsmaße nur Vollzeitbeschäftigte mit Sozialversicherungspflicht (ohne Auszubildende) berücksichtigt. Bei der Deflationierung der Löhne wird der Verbraucherpreisindex des Statistischen Bundesamtes mit dem Ausgangsjahr 2000 (=100) zugrunde gelegt. Bei einem nominalen Bruttomonatslohn von 2000,- Euro bedeutet ein unverändertes Einkommen in beiden Jahren demnach einen realen Lohnverlust von etwa 24,- Euro monatlich. Auf der Betriebsebene wird mit der Standardabweichung der Lohnänderung für einzelne Beschäftigte berechnet, ob es sich um „betriebsübliche“ Lohnerhöhung handelt oder ob einzelne Personen(gruppen) diesbezüglich besser abschneiden als andere.

Bei den Lohnänderungen werden aus dem jeweiligen Betrieb zwischen den Jahren 2003 und 2004 ausgeschiedene und im gleichen Zeitraum neu eingestellte MitarbeiterInnen von der Berechnung (und der der entsprechenden Standardabweichung) ausgeschlossen. Der Wechsel der Arbeitszeit von 2003 auf 2004 (von Voll- in Teilzeit oder umgekehrt) führt ebenfalls zum Ausschluss von den jeweiligen Berechnungen.

Alle Lohnangaben sind in Euro. Zur besseren Einordnung werden Durchschnittslöhne als *Bruttomonatslöhne* in den Ergebnistabellen ausgewiesen. Anteile beziehen sich immer auf die Personalstruktur zum 30. Juni 2004, Ein- und Austrittsraten werden als Kettenindizes zum Vorjahr berechnet<sup>1</sup>. Die Betriebszugehörigkeitsdauer wird in Jahren gemessen. Für den Anteil der Berufe an der Gesamtbeschäftigung wird jedes im Betrieb vorkommende Berufsbild (auf der Dreistellerebene nach BA-Klassifikation 1982) einmal gezählt und anschließend durch die Gesamtanzahl der Beschäftigten im Betrieb geteilt (und der Wert mit einhundert multipliziert). Ein Wert von 100 ergibt sich, wenn alle Beschäftigten innerhalb eines Betriebs einen anderen Beruf ausüben. Werte nahe Null ergeben sich, wenn alle Beschäftigten innerhalb eines Betriebs den gleichen Beruf ausüben.

Informelle Qualifikationen werden über den unbeobachtbaren Lohnbonus gemessen, der sich aus einer Panelregression (Zeitraum 2000 bis 2004 bei jährlicher Beobachtungsstruktur der Daten) einer erweiterten Mincer'schen Lohngleichung ergibt. Es handelt sich also um den fixed effect  $\theta_i$ .  $\theta_i$  ist für jede in Vollzeit beschäftigte Person ein individueller Wert. Er liegt nur für Personen vor, die in mindestens zwei Jahren in einem IAB-Betriebspanelbetrieb der Jahre 2000 bis 2004 jeweils in Vollzeit gearbeitet haben. Einen besonders hohen Lohnbonus für informelle Qualifikationen erzielen alle Beschäftigten, die im obersten, einen besonders geringen alle die im untersten Quintil der  $\theta_i$ -Verteilung liegen. Dies bedeutet nicht zwangsläufig, dass Personen mit einem Lohnbonus für informelle Qualifikationen immer besser bezahlt werden, sondern nur, dass sie neben den Erträgen für formelle Qualifikationen – in den IAB-Personendaten sind das beispielsweise das Bildungsniveau in Verbindung mit der beruflichen (Erst-)Ausbildung, der ausgeübte Beruf oder die Betriebszugehörigkeitsdauer – auch Lohnerträge erzielen, die aufgrund der Wirkung unbeobachteter Variablen zustande kommen.

Der individuelle Lohnbonus für informelle Qualifikationen ist also im Kontrast zu den Lohnzahlungen zu sehen, den Beschäftigte für ihre formelle Qualifikation und weitere (mit IAB-Daten) beobachtbare Merkmale erzielen. Eine vor noch nicht allzu langer Zeit neu in den Beschäftigungsbetrieb eingetretene Person mit einem hohen Lohnbonus für informelle Qualifikationen kann als un- und angelernte Person in einem wenig anspruchsvollen Berufsbild also real weniger verdienen als eine Person mit einem geringen (im Extremfall: negativen) Lohnbonus für informelle Qualifikationen, die aber bereits seit geraumer Zeit mit einem Hochschulabschluss in einem anspruchsvollen Berufsbild im Beschäftigungsbetrieb arbeitet.

---

<sup>1</sup> Formel:  $2 * E_t / (N_t + N_{t-1})$ . E ist die Anzahl der Ein- oder Austritte zwischen dem 30. Juni 2003 und dem 30. Juni 2004, N ist die Gesamtbeschäftigtenzahl im Betrieb, t ist der 30. Juni 2004 und t-1 der 30. Juni 2003.

### 3. Vorgehen bei der Korrespondenz- und Clusteranalyse

Die am Ende des zweiten Abschnitts angesprochenen Einzelaspekte der sozioökonomischen Profile von Betrieben werden mit den LIAB-Querschnittdaten operationalisiert und gehen anschließend zunächst in eine Faktorenanalyse ein, um die empirisch relevanten Facetten einer jeden Dimension zu bestimmen. Da Faktoren- und Clusteranalysen auf der Messung n-dimensionaler räumlicher Distanzen basieren, wurde bei der Konstruktion der Variablen – trotz späterer z-Standardisierung – bereits im Vorfeld auf ein möglichst einheitliches Wertespektrum geachtet. Alle Variablen gehen anschließend z-standardisiert (Mittelwert eins; Standardabweichung null) in die Faktoren- und Clusteranalysen ein, werden aber unstandardisiert, also in ihrer „ursprünglichen“ Form, in den Ergebnistabellen der Typisierung ausgewiesen.

Die Ausweisung und Interpretation der Ergebnisse erfolgt außerdem ausschließlich auf der Betriebsebene. Das ist erwähnenswert, weil es bei der Ausweisung betrieblicher Anteilswerte immer auch Betriebsgrößenklasseneffekte gibt. Jede Einheit (hier: Betriebsanteile) geht (ohne weiteres) in eine beliebige Mittelwertberechnung mit dem Gewicht von eins ein, und zwar unabhängig davon, ob der Betrieb zehn oder tausend Beschäftigte hat. Demnach hat eine Person bei einer beliebigen Anteilsbildung in dem einen Betrieb ein Gewicht von zehn Prozent, im anderen nur von einem Promille. Wenn sich also Merkmale wie etwa Niedriglohnbeschäftigung insbesondere in kleineren Betrieben beobachten lassen, bekommen sie bei der Mittelwertausweisung ein viel größeres Gewicht, als wenn sich bestimmte Charakteristika von Beschäftigten in Großbetrieben konzentrieren.

Auf die Darstellung von Einzelheiten beim Vorgehen und zu den Ergebnissen der Faktorenanalyse wird aus Platzgründen verzichtet. In der Faktorenanalyse wurden die relevanten Facetten einer jeden Dimension ermittelt und jede dieser Facetten wird mit mindestens einem Variablenvertreter in der Betriebstypisierung berücksichtigt. Einzelheiten und Messergebnisse sind auf Nachfrage erhältlich. Es wurde eine Hauptkomponentenanalyse mit anschließender Rotation durchgeführt, bei der einzelne Items miteinander korreliert sein können (orthogonales Quartimin-Verfahren).

Zur Findung der „optimalen“ Clusteranzahl und einer entsprechenden Clusterzentrenbeschreibung werden für die Clusterung zu den Betriebstypen zwei aufeinander aufbauende Verfahren angewendet. Im ersten Schritt bestimmt das Ward-Verfahren die „optimale“ Clusteranzahl, die sich anhand des Duda-Hard-Index (Duda/ Hard, 1973) und des Calinski-

Harabasz-F-Wertes (Calinski/ Harabasz, 1974) ergibt<sup>2</sup>. Die Anwendung des Ward-Verfahren hat einige Implikationen. Als hierarchisch-agglomeratives Verfahren fasst es sukzessive Objekte zusammen, bis zum Schluss nur noch ein Cluster übrig ist. Aus den Zwischenschritten der Agglomeration ergibt sich dann - beurteilt nach den eben diskutierten Teststatistiken – die am besten geeignete Clusteranzahl. Beim Ward-Verfahren erfolgt die Vereinigung der Cluster (Objekte) nach dem Kriterium der geringsten Binnenvarianz (Bacher, 1996: Kapitel 3; Backhaus, 2006: Kapitel 8). Einmal vereinigte Objekte werden im weiteren Verlauf der Agglomeration nicht wieder getrennt. Außerdem verändern die Clusterzentren (Zentroide) bei den einzelnen Vereinigungsschritten ihre Lage im Raum. Mit der Anwendung des K-Means-Verfahren kann eine nachträgliche Umsortierung der Objekte abhängig von den jeweiligen Zentroiden erfolgen. Es ist deswegen kompatibel zum Ward-Verfahren (Rüb / Werner, 2007: 14).

Im zweiten Schritt der Clusteranalyse wurden die über das Ward-Verfahren ermittelten zehn Clusterzentren als Startwerte für die K-Means-Methode eingesetzt. Im Gegensatz zum Ward-Verfahren erfolgt im K-Means-Verfahren in einem iterativen Prozess eine ständige Neuordnung der zu typisierenden Betriebe, bis die Varianz – bei vorgegebener Clusteranzahl – minimal ist. Beide „Startwerte“ - die Anzahl der Cluster und ihre Zentroide – stammen aus dem Ward-Verfahren, wobei im K-Means-Verfahren die Zentroide ihren jeweiligen Schwerpunkt ändern können.

Das liegt daran, dass im K-Means-Verfahren die Ähnlichkeit zwischen Objekten mit der Fehlerquadratsumme (ESS = error sum of squares) berechnet wird. Der Zentroid eines Cluster C mit 1 ... k Objekten ergibt sich dann aus Gleichung (1).

$$(1) \text{ESS}(C_k) = \sum_{i \in C_k} \sum_{f=1}^F (x_{if} - x_f(k))^2 \text{ mit der Bedingung } \text{ESS} = \sum_{k=1}^K \text{ESS}(C_k) \rightarrow \min .$$

Gleichung (1) bedeutet unter anderem, dass jedes Mal, wenn ein Objekt den Cluster wechselt, eine komplette Neuberechnung der Clusterzentroiden erfolgt. Damit ist der Rechenaufwand für das K-Means-Verfahren zwar wesentlich größer als bei hierarchisch-agglomerativen Verfahren, aber die aufgezählten Vorzüge machen diesen Nachteil mehr als wett, zumal es bei Fallzahlen von maximal 6,000 Einzelobjekten auf modernen Rechenanlagen keine Kapazitätsprobleme gibt.

---

<sup>2</sup> Beim Duda-Hard-Index wird nach lokalen Minima der Pseudo-T-Squared-Teststatistik gesucht, die mit lokalen Maxima im Je(2) / Je(1)-Quotienten einhergehen. Beim Calinski-Harabasz-F-Wert zeigen lokale Minima gut geeignete Clusterzahlen an (Fynch, 2005: 93).

Diese Vorgehensweise liefert beispielsweise für die westdeutsche Betriebslandschaft im Jahr 2004 das folgende Ergebnis. Der F-Wert von Calinski / Harabasz gibt in beiden Landesteilen nur mäßige Informationen über eine optimale Clusteranzahl in den jeweiligen Landesdatensätzen. Aus den F-Werten lässt sich lediglich ablesen, dass die optimale Clusteranzahl zwischen zwei und 14 liegen sollte. Aufschlussreicher ist ein Blick auf den Duda-Hart-Index. Nach dem  $Je(2) / Je(1)$  – Kriterium (lokales Maximum) und der Pseudo-T-Squared-Statistik (lokales Minimum) existieren gute Lösungen für vier und zehn Cluster, sechs und zwölf Cluster werden ebenfalls von der Teststatistik relativ gut bewertet, aber etwas schlechter als die beiden erstgenannten. Die Wahl der Cluster fiel deswegen auf zehn in beiden Landesteilen. Der Wert zehn ist vor diesem Hintergrund eine gut geeignete Anzahl, weil der  $Je(2) / Je(1)$  – Index nach dem Ward-Verfahren dort ein lokales Maximum, und die Pseudo-T<sup>2</sup>-Statistik ein lokales Minimum hat. Wie auch die Regressionsanalysen und entsprechende statistische Tests bestätigen, ist die Zehn-Clusterlösung robust<sup>3</sup>. Die Beschreibung der Typen im Haupttext erfolgt anhand der Ausweisung der jeweiligen Clusterzentren, d.h. anhand der multivariaten Mittelwerte der typisierenden Variablen nach der Anwendung des wie eben angewendeten K-Means-Verfahrens.

#### 4. Regressionsanalyse

Um alle Betriebstypen einer simultanen Analyse zu unterziehen, wird ein multinomiales Logit-Modell (MNLM) angewendet<sup>4</sup>. Bei einem einfachen Logit-Modell für eine binär codierte Variable wird mit

$$(2) \quad \Omega(x) = \frac{\Pr(y=1|x)}{\Pr(y=0|x)} = \frac{\Pr(y=1|x)}{1 - \Pr(y=1|x)}$$

die Wahrscheinlichkeit ausgedrückt, wie oft etwas passiert ( $y = 1$ ) relativ dazu, wie oft es nicht passiert ( $y = 0$ ).  $X$  ist ein Set an unabhängigen Variablen. Bei Logarithmierung von  $\Omega$  ergibt sich für die so genannten *Logits* das Werteintervall minus bis plus  $\infty$ , woraus die lineare Beziehung  $\ln(\Omega) = x \cdot \beta (+ \varepsilon)$  abgeleitet wird. Das MNLM ist ein Regressionsmodell, in dem die abhängige kategoriale Variable mehr als zwei Ausprägungen hat. Dies hat zur Folge, dass

---

<sup>3</sup> Alle Ergebnisse inklusive der entsprechenden Cluster-Dendrogramme werden auf Anfrage gerne zur Verfügung gestellt.

<sup>4</sup> Da eine *eindeutige* Rangordnung der Betriebstypen als zu voraussetzungsvoll erscheint, wurde der Ansatz des Ordered-Logit-Modells nicht weiter verfolgt.



eine Basiskategorie  $b$  definiert wird. Die Koeffizienten des MNLM ermitteln dann die (signifikanten) Abweichungen der Kovariaten der einzelnen Betriebstypen von dieser Basiskategorie, denn die Wahrscheinlichkeit, einen anderen Betriebstyp  $m$  als die Basiskategorie in Abhängigkeit von Kovariaten  $x$  anzutreffen, ist das Verhältnis der Logits der beiden Typen zueinander, formell

$$(3) \quad \ln \Omega_{m|b} = \ln \frac{\Pr(y = m | x)}{\Pr(y = b | x)} = x\beta_{m|b} .$$

Da  $\ln \Omega_{b|b}(x) = \ln 1 = 0$  ist, sind die Einflüsse der unabhängigen Variablen bei der Basiskategorie ebenfalls notwendigerweise null. Die Wahrscheinlichkeit den Betriebstyp  $m$  anzutreffen, lässt sich bei  $J = 1 \dots 9$  anderen Betriebstypen als relationales System ihrer relativen Abweichungen von der Basiskategorie beschreiben.

$$(4) \quad \Pr(y = m | x) = \frac{\exp(x\beta_{m|b})}{\sum_{j=1}^J \exp(x\beta_{j|b})} .$$

Der Wechsel der Basiskategorie wirkt sich demnach nicht auf die vorhergesagten Wahrscheinlichkeiten der  $m$  Ausprägungen aus, aber sie werden mit einem Wechsel unterschiedlich parametrisiert. Insbesondere bei einer hohen Anzahl an Regressoren besteht deshalb eine gewisse Gefahr, sich in „nervtötenden“ (tedious) Vergleichen zu verlieren (vgl. Long/ Freese, 2001: Abschnitt 6.6) oder von der Komplexität der Ergebnisse überwältigt zu werden (ebd: 171). Eine Vielzahl von Regressoren ist im vorliegenden Fall aber empfehlenswert, denn es wird eine Querschnittsschätzung durchgeführt, in der die Einflüsse unbeobachteter Variablen nicht berücksichtigt werden. Da der Fokus auf den Wirkungen der Exportvariable und der IuK-Investitionen liegen wird, wird auf die Ergebnisse für die Vielzahl an weiteren Kontrollvariablen nicht näher eingegangen. Letztere sind notwendig, um die Einflüsse der interessierenden Variablen im Rahmen der methodischen Möglichkeiten einer Querschnittsschätzung möglichst unverzerrt abzubilden. Einzelheiten zu den in das Regressionsmodell aufgenommenen befinden sich im Anhang (Tabelle 6.1).

In Gleichung (4) wird angenommen, dass die Wahrscheinlichkeit, das Auftreten einer Ausprägung  $m$  zu beobachten, nicht von der Wahrscheinlichkeit des Auftretens einer (oder mehrerer) anderen Ausprägung  $n$  der abhängigen Variable beeinflusst wird, sondern sich in den Wahrscheinlichkeitsbeziehungen des Vektors  $x$  für die unabhängigen Variablen (relativ zur Basiskategorie) manifestiert. Diese so genannte IIA-Hypothese (IIA: independance of irrelevant alternatives) wird überprüft, indem getestet wird, ob sich durch das Hinzufügen oder die Wegnahme von einzelnen Ausprägungen der abhängigen Variable die Logits für unabhängige Variablen ändern. Dann ist die Annahme

$$(5) \quad \frac{\Pr(y = m | x)}{\Pr(y = n | x)} = \exp(x [\beta_{m|b} - \beta_{n|b}])$$

verletzt, was sich nach Hausman/McFadden (1984) mit einem Hausman-Test überprüfen lässt<sup>5</sup>. Einen approximativen Likelihood-Ratio-Test schlagen McFadden/ Tye/ Train (1978) vor, der durch Small/Hsiao (1985) noch einmal verbessert wurde<sup>6</sup>. Die Signifikanz der Effekte einzelner Variablen des Vektors X werden mit Wald-Tests überprüft, zum einen, inwiefern die unabhängigen Variablen überhaupt einen Effekt auf die Wahl des Betriebstyps haben<sup>7</sup> und zum anderen, inwiefern die unabhängigen Variablen (als Gruppe) in der Lage sind, zwischen zwei Ausprägungen der abhängigen Variable zu differenzieren. Sofern – wie im vorgelegten Fall - in diesen Tests keine Verletzungen der entsprechenden Annahmen vorliegen, sagt das MNLM mit einer geschätzten Wahrscheinlichkeit  $\hat{P}$  die geschätzten Effekte  $\hat{\beta}$  für die einzelnen Faktoren aus dem wettbewerbsrelevanten Umfeld vorher:

$$(6) \quad \hat{P}(y = m | x) = \frac{\exp(x\hat{\beta}_{m|j})}{\sum_{j=1}^J \exp(x\hat{\beta}_{j|j})}$$

Mit der Variation von x variiert auch  $\hat{P}_m$ . Gegeben konstante Werte aller anderen Variablen von X lässt sich die *Faktoränderung* für eine einzelne x-Variable berechnen<sup>8</sup>. Die Faktoränderung gibt an, wie sich  $\hat{P}$  ändert, wenn die entsprechende x-Variable um eine Standardabweichung erhöht wird. Der marginale Effekt wird mit

$$(7) \quad \frac{\partial \Pr(y = m | x)}{\partial x_i} = \Pr(y = m | x) [\beta_{k, m|j} - \sum_{j=1}^J \beta_{k, j|j} \Pr(y = j | x)]$$

berechnet. Beim Samplemittelwert etwa geben die Änderungen der  $\beta$ -Koeffizienten Auskunft darüber, wie sich die deutsche Betriebslandschaft – unter der Annahme, es gäbe nur durchschnittliche Betriebe bezüglich der x-Variablen - in Bezug auf ihre sozioökonomischen Profile (Betriebstypen) wandelt, wenn die betrieblichen Exportanteile bzw. die IuK-Investitionen zu- oder abnehmen. Für die Ausstattungsmerkmale von Betrieben, die mit binären Variable-

<sup>5</sup> Der Hausman-Test ist dreistufig. Im ersten Schritt werden alle J Ausprägungen in einem Modell geschätzt. Im zweiten Schritt wird ein restringiertes Modell geschätzt, bei der eine oder mehrere Ausprägungen der abhängigen Variable nicht berücksichtigt werden. Im dritten Schritt wird  $\hat{\beta}_F^*$  als Unterset von  $\hat{\beta}_R$  definiert. Die Teststatistik von Hausman / McFadden ist mit  $\hat{V}$  als vorhergesagte Varianz  $H = (\hat{\beta}_R - \hat{\beta}_F^*)' [\hat{V}(\hat{\beta}_R) - \hat{V}(\hat{\beta}_F^*)]^{-1} (\hat{\beta}_R - \hat{\beta}_F^*)$ . H ist asymptotisch  $\chi^2$  verteilt. H hat so viele Freiheitsgrade, wie es Spalten in  $\hat{\beta}_R$  gibt, bei denen die IIA wahr ist. Signifikante Werte für H zeigen also die Verletzung der IIA-Annahme an.

<sup>6</sup> Vereinfacht gesagt wird im Small-Hsiao-Test das Sample in zwei Zufallsstichproben gleicher Größe unterteilt. Anschließend wird ein zweites restringiertes Sample erzeugt, in dem alle Fälle mit einem bestimmten und auszuwählendem Wert in der abhängigen Variable gelöscht werden. Anschließend werden beide Verteilungen miteinander verglichen (vgl. Long/Freese, 2001: 189). Die Anzahl der Freiheitsgrade bestimmt sich im Small-Hsiao-Test durch die Anzahl der unabhängigen Variablen (K) und beträgt K+1.

<sup>7</sup> Es wird geprüft, inwiefern die J-1 Koeffizienten *simultan* gleich null sind.

<sup>8</sup> In der Literatur werden Faktoränderungen auch als Effektkoeffizienten bezeichnet.

nausprägungen beschrieben werden - hier ob Betriebe einen ausländischen Mehrheitseigentümer haben bzw. ihren Mitarbeitern die Möglichkeit zum E-Learning bieten – wird der marginale Effekt über die diskrete Änderung beschrieben:

$$(8) \quad \frac{\Delta \Pr(y = m | x)}{\Delta x_k} = \Pr(y = m | x, x_k = x_E) - \Pr(y = m | x, x_k = x_S).$$

Gleichung (8) berechnet demnach J Koeffizienten für die diskrete Veränderung einer jeden binären Variable.  $\beta_{x_k | b}$  ist null. Die allgemeine Form der Faktoränderungen lautet

$$(9) \quad \frac{\Omega_{m | n}(x, x_k + \delta)}{\Omega_{m | n}(x, x_k)} = e^{\beta_{k, m | n} \delta}.$$

Wenn  $x_k$  eine binäre Variable ist, dann verändern sich die Logits des Betriebstyps  $m$  gegenüber denen vom Typ  $n$  um einen geschätzten Faktor von  $\exp(\beta_{k, m | n})$ . Für alle anderen Variablentypen wird erwartet, dass sich mit  $\delta = s_{x_k}$  - also einer Änderung um eine Standardabweichung von  $x_k$  - die Logits für  $m$  gegenüber  $n$  um den Faktor  $\exp(\beta_{k, m | n} \times s_k)$  ändern (Long/Freese, 2001: 203).

Inhaltlich bestimmt das MNLM, welche Korrelationen es zwischen der Wahrscheinlichkeit des Beobachtens einzelner sozioökonomischer Betriebstypen und den unabhängigen  $x$ -Variablen gibt.

Kausale Aussagen lassen sich demnach aus der hier beschriebenen Methodologie nicht ableiten. Mit den Vorhersagewerten des MNLM kann abgeschätzt werden, wie sich die deutsche Betriebslandschaft hinsichtlich der Verteilung der sozioökonomischen Profile ändert, etwa wenn die deutsche Wirtschaft (noch) mehr exportiert bzw. Betriebe ihre IuK-Investitionen intensivieren. Aber auch diese Ergebnisse sind nicht kausal interpretierbar, weil lediglich die (Veränderung der) Kovarianzmatrix der unabhängigen Variablen ausgenutzt wird, um daraus die vorhergesagten Prozentanteile der einzelnen sozioökonomischen Betriebstypen abzuleiten. Die Genauigkeit dieser Vorhersage hängt von der Güte des zugrunde gelegten Regressionsmodells ab. Im Anhang sind am Ende der Tabelle 6.1 Gütekriterien zur Beurteilung von MNLM für die in diesem Beitrag aufgestellten Regressionsgleichungen aufgelistet.

## 5. Ostdeutsche Betriebsprofile

**Tab. 5.1: Gruppenmittelwerte der Betriebstypen des Niedriglohnsegments**

<b>Löhne und Gehälter</b>	<b>Ia</b>	<b>Ib</b>	<b>Ic</b>	<b>Id</b>
durchschnittlicher Monatslohn (in Euro)	1253	1329	1279	1511
Standardabweichung <i>Tageslohn</i>	13,5	11,6	10,6	14,8
Quotient 90- zu 50-Perzentil	1,45	1,39	1,38	1,43
Anteil beste Verdienste	2,2	1,7	0,5	3,6
Anteil schlechteste Verdienste	79,3	75,3	82,3	50,2
Anteil hoher Lohnbonus für Weiterqualifizierung	2,6	2,2	0,6	3,7
Anteil niedriger Lohnbonus für Weiterqualifizierung	61,4	65,5	72,5	44,8
durchschnittliche Monatslohnänderung zum Vorjahr	-65,4	-18,1	-22,5	-44,0
Standardabweichung der <i>Tageslohn</i> änderung	4,9	3,8	3,2	5,9
<b>Beschäftigungsstabilität</b>				
durchschnittl. Betriebszugehörigkeitsdauer in Jahren	2,3	5,4	6,6	5,1
Anteil Kernbelegschaft	15,2	45,3	57,5	44,2
Austrittsrate	47,7	19,8	16,5	35,1
Austrittsrate schlechteste Verdienste	46,3	17,9	17,4	52,8
Austrittsrate beste Verdienste	3,2	1,1	0,4	9,5
Eintrittsrate	76,6	25,3	19,4	26,6
Eintrittsrate schlechteste Verdienste	84,5	28,7	21,6	52,0
Eintrittsrate beste Verdienste	1,3	0,3	0,2	5,1
<b>Qualifikation und Erwerbsformen</b>				
Anzahl Berufe/ Gesamtbeschäftigtenzahl	27,7	31,8	37,8	26,7
Anteil Hochschulabsolventen	7,5	2,6	3,5	18,2
Anteil Teilzeitbeschäftigte	36,6	24,2	15,0	2,2
Anteil befristet Beschäftigte	75,5	10,7	3,6	11,4
Anteil Auszubildende	1,1	3,3	7,7	62,4
Anteil sonstige Beschäftigte	6,6	19,4	4,4	0,9
Anteil un-/ angelernte Arbeitskräfte	19,5	57,5	5,4	1,1
Anteil Facharbeiter	44,4	11,7	56,3	19,4

Quelle: LIAB-Querschnittmodell 2004, Betriebe ab fünf Beschäftigte.

Anmerkungen zur Konstruktion und Interpretation einzelner Variablen befinden sich in Abschnitt 2: Konstruktion und Eigenschaften der Variablen für die Betriebstypisierung.

**Tab. 5.2: Gruppenmittelwerte der Betriebstypen des mittleren Segments**

<b>Löhne und Gehälter</b>	<b>IIIa</b>	<b>IIIb</b>
durchschnittlicher Monatslohn (in Euro)	1889	2403
Standardabweichung <i>Tageslohn</i>	14,1	30,6
Quotient 90- zu 50-Perzentil	1,31	1,68
Anteil beste Verdienste	4,2	23,7
Anteil schlechteste Verdienste	20,2	14,6
Anteil hoher Lohnbonus für Weiterqualifizierung	4,0	24,1
Anteil niedriger Lohnbonus für Weiterqualifizierung	12,6	8,8
durchschnittliche Monatslohnänderung zum Vorjahr	-3,4	18,5
Standardabweichung der <i>Tageslohn</i> änderung	4,6	11,8
<b>Beschäftigungsstabilität</b>		
durchschnittl. Betriebszugehörigkeitsdauer in Jahren	7,9	7,2
Anteil Kernbelegschaft	70,0	63,3
Austrittsrate	11,7	13,2
Austrittsrate schlechteste Verdienste	13,1	13,4
Austrittsrate beste Verdienste	2,5	10,7
Eintrittsrate	11,9	14,2
Eintrittsrate schlechteste Verdienste	17,2	49,5
Eintrittsrate beste Verdienste	0,6	4,5
<b>Qualifikation und Erwerbsformen</b>		
Anzahl Berufe/ Gesamtbeschäftigtenzahl	35,5	34,0
Anteil HochschulabsolventInnen	6,3	20,2
Anteil Teilzeitbeschäftigte	16,2	14,5
Anteil befristet Beschäftigte	5,3	4,2
Anteil Auszubildende	4,7	3,4
Anteil sonstige Beschäftigte	3,5	3,9
Anteil un-/ angelernte Arbeitskräfte	5,5	7,0
Anteil FacharbeiterInnen	46,4	17,3

Quelle: LIAB-Querschnittmodell 2004, Betriebe ab fünf Beschäftigte, gewichtete Werte.

Anmerkungen zur Konstruktion und Interpretation einzelner Variablen befinden sich in Abschnitt 2: Konstruktion und Eigenschaften der Variablen für die Betriebstypisierung.

**Tab. 5.3: Gruppenmittelwerte des Betriebstyps im Hochlohnsegment**

<b>Löhne und Gehälter</b>	<b>I Ib</b>
durchschnittlicher Monatslohn (in Euro)	3423
Standardabweichung <i>Tageslohn</i>	32,8
Quotient 90- zu 50-Perzentil	1,45
Anteil beste Verdienste	69,9
Anteil schlechteste Verdienste	2,8
Anteil hoher Lohnbonus für Weiterqualifizierung	69,3
Anteil niedriger Lohnbonus für Weiterqualifizierung	1,7
durchschnittliche Monatslohnänderung zum Vorjahr	29,6
Standardabweichung der <i>Tageslohn</i> änderung	20,3
<b>Beschäftigungsstabilität</b>	
durchschnittl. Betriebszugehörigkeitsdauer in Jahren	7,2
Anteil Kernbelegschaft	59,4
Austrittsrate	13,1
Austrittsrate schlechteste Verdienste	11,2
Austrittsrate beste Verdienste	8,4
Eintrittsrate	9,7
Eintrittsrate schlechteste Verdienste	11,1
Eintrittsrate beste Verdienste	5,9
<b>Qualifikation und Erwerbsformen</b>	
Anzahl Berufe/ Gesamtbeschäftigtenzahl	25,1
Anteil Hochschulabsolventen	49,0
Anteil Teilzeitbeschäftigte	19,3
Anteil befristet Beschäftigte	5,9
Anteil Auszubildende	1,6
Anteil sonstige Beschäftigte	2,3
Anteil un-/ angelernte Arbeitskräfte	2,4
Anteil Facharbeiter	6,3

Quelle: LIAB-Querschnittmodell 2004, Betriebe ab fünf Beschäftigte.

Anmerkungen zur Konstruktion und Interpretation einzelner Variablen befinden sich in Abschnitt 2: Konstruktion und Eigenschaften der Variablen für die Betriebstypisierung.

**Tab. 5.4: Gruppenmittelwerte der Betriebe im Umbruch**

<b>Löhne und Gehälter</b>	IVa	IVb	IVc
durchschnittlicher Monatslohn (in Euro)	1606	2150	2228
Standardabweichung <i>Tageslohn</i>	12,4	16,3	18,8
Quotient 90- zu 50-Perzentil	1,35	1,45	1,36
Anteil beste Verdienste	1,2	28,3	13,3
Anteil schlechteste Verdienste	46,3	31,1	7,7
Anteil hoher Lohnbonus für Weiterqualifizierung	2,3	29,7	12,7
Anteil niedriger Lohnbonus für Weiterqualifizierung	38,3	22,1	3,9
durchschnittliche Monatslohnänderung zum Vorjahr	-25,4	30,4	18,5
Standardabweichung der <i>Tageslohn</i> änderung	4,9	7,0	7,6
<b>Beschäftigungsstabilität</b>			
durchschnittl. Betriebszugehörigkeitsdauer in Jahren	6,6	5,3	7,4
Anteil Kernbelegschaft	58,5	38,3	67,1
Austrittsrate	63,2	17,2	41,4
Austrittsrate schlechteste Verdienste	58,6	23,5	159,2
Austrittsrate beste Verdienste	186,1	13,2	7,6
Eintrittsrate	12,5	43,8	12,0
Eintrittsrate schlechteste Verdienste	17,6	34,9	13,2
Eintrittsrate beste Verdienste	1,0	179,0	0,9
<b>Qualifikation und Erwerbsformen</b>			
Anzahl Berufe/ Gesamtbeschäftigtenzahl	37,4	29,4	36,4
Anteil HochschulabsolventInnen	7,4	19,9	10,2
Anteil Teilzeitbeschäftigte	11,7	26,8	17,1
Anteil befristet Beschäftigte	5,3	5,0	5,8
Anteil Auszubildende	7,0	6,9	3,2
Anteil sonstige Beschäftigte	7,6	4,5	3,5
Anteil un-/ angelernte Arbeitskräfte	9,9	8,2	8,4
Anteil FacharbeiterInnen	37,3	57,8	36,2

Quelle: LIAB-Querschnittmodell 2004, Betriebe ab fünf Beschäftigte.

Anmerkungen zur Konstruktion und Interpretation einzelner Variablen befinden sich in Abschnitt 2: Konstruktion und Eigenschaften der Variablen für die Betriebstypisierung.

**Übersicht 5.1: Ausprägungen der sozioökonomischen Profile in den ostdeutschen Betriebstypen des Jahres 2004**

		<b>Löhne und Gehälter</b>	<b>Beschäftigungsstabilität</b>	<b>Qualifikationsprofil</b>
<b>I Niedriglohnsegment</b>				
	Ia	--	--	--
	Ib	-	-	-
	Ic	-	-	o
	Id	-	-	o
<b>II Hochlohnsegment</b>				
	IIb	++	+	++
<b>III Betriebe mit mittlerem Profil</b>				
	IIIa	o	++	o
	IIIb	o	+	o
<b>IV Betriebe im Umbruch</b>				
	IVa	o	o	o
	IVb	o	-	o
	IVc	o	+	o

Legende: o : durchschnittlich ausgeprägt    - : schwach (niedrig) ausgeprägt  
+ : stark (hoch) ausgeprägt.

Ein doppeltes Zeichen kennzeichnet den höchsten (niedrigsten) Gruppenmittelwert aller Betriebstypen.

**5.5 Abdeckungsgrad der Typen über ostdeutsche Betriebe und Beschäftigte**

Die nachfolgende Tabelle 5.5 für Ostdeutschland basiert auf den gleichen Konstruktionsprinzipien wie die im Haupttext für Westdeutschland. Dabei ist erkennbar, dass das ostdeutsche Niedriglohnsegment in etwa doppelt so groß ist wie in Westdeutschland und spiegelbildlich das Hochlohnsegment um etwa den gleichen Prozentsatz kleiner. Interessanterweise deckt das mittlere Segment (IIIa/IIIb) etwa gleich viele Beschäftigte wie in Westdeutschland bei geringeren betrieblichen Anteilen an gesamt ab, d.h. die Betriebe des mittlere Segments (IIIa/IIIb) sind in Ostdeutschland größer als in Westdeutschland. Ein weiterer Befund ist, dass sich ostdeutsche Betriebe öfter in einem für D1 bis D3 maßgeblichen Umbruch befinden, allerdings im Gegensatz zu Westdeutschland mit deutlichem Übergewicht bei den Aufsteigern (IVb) bzw. den Upgradern (IVc).



**Tab. 5.5: Abdeckungsgrad der ostdeutschen Typen im Jahr 2004 über Betriebe, Erwerbstätige und sozialversicherungspflichtig Beschäftigte (SvB)**

	Anteil an gesamt (Betriebe)	Anzahl in Tsd.	Anteil an allen Erwerbstätigen	Anzahl in Tsd.	Anteil an allen SvB	Anzahl in Tsd.
<b>Niedriglohnsegment</b>	<b>50,4</b>	<b>55</b>	<b>36,3</b>	<b>1270</b>	<b>35,3</b>	<b>1107</b>
Ia	3,3	4	4,5	157	4,8	150
Ib	9,4	10	9,5	334	8,0	250
Ic	36,8	40	19,6	686	19,6	615
Id	0,8	1	2,7	93	2,9	92
<b>Hochlohnsegment</b>	<b>6,5</b>	<b>7</b>	<b>8,8</b>	<b>308</b>	<b>9,1</b>	<b>284</b>
IIb	6,5	7	8,8	308	9,1	284
<b>mittleres Segment</b>	<b>36,9</b>	<b>40</b>	<b>46,8</b>	<b>1635</b>	<b>47,8</b>	<b>1497</b>
IIIa	26,3	29	24,5	857	24,6	771
IIIb	10,5	11	22,3	778	23,2	727
<b>Betriebe im Umbruch</b>	<b>6,3</b>	<b>7</b>	<b>8,0</b>	<b>281</b>	<b>7,9</b>	<b>246</b>
IVa	1,3	1	1,0	35	1,0	30
IVb	1,5	2	2,0	70	2,0	62
IVc	3,5	4	5,1	176	4,9	154
<b>insgesamt</b>	<b>100</b>	<b>108</b>	<b>100</b>	<b>3494</b>	<b>100</b>	<b>3134</b>

Quelle: LIAB-Querschnittmodell für das Jahr 2004, Betriebe ab fünf Beschäftigte, die in den Jahren 2003 und 2004 an der IAB-Betriebspanelbefragung teilgenommen haben, Querschnittgewichtete Werte.

## 6. Anhänge

### 6.1 Regressionsergebnisse<sup>9</sup> und Regressionsdiagnostik für Westdeutschland

[Basiskategorie IIIc (größere Betriebe mit hohen Facharbeiteranteilen, einem mittleren Lohnniveau, hoher Lohnspreizung, langen Betriebszugehörigkeitsdauern und geringen Fluktuationsraten)]

<sup>9</sup> Mindestens auf dem Fünf-Prozentsniveau signifikante Koeffizienten sind fettgedruckt. Gestrichelte Linien legen einen Variablenblock fest, der ohne Angabe einer Referenzgröße kollinear wäre.

Organisationswandel	Niedriglohnbetriebe			Hochlohnbetriebe	
	Ia	Ib	Ic	IIa	IIb
Outsourcing	-0,343	-0,657	-0,382	<b>-0,485</b>	-0,430
Insourcing	<b>1,368</b>	0,814	<b>2,063</b>	<b>1,177</b>	<b>1,326</b>
Verlagerung Ausland	-0,723	-1,020	-0,623	-0,521	-0,594
Produktionsstruktur geändert	<b>-0,680</b>	0,010	0,095	0,024	-0,131
Aufgabenstruktur geändert	0,354	-0,149	-0,130	0,228	0,319
<b>technologischer Fortschritt</b>					
Forschung und Entwicklung	0,280	<b>-0,588</b>	<b>-0,570</b>	<b>0,441</b>	<b>0,429</b>
Investitionssumme (log)*	-0,229	<b>-0,252</b>	<b>-0,344</b>	0,008	<b>-0,080</b>
Investitionssumme in IuK (log)*	0,187	<b>-0,160</b>	<b>-0,105</b>	<b>0,095</b>	<b>0,378</b>
<b>Institutionelle Regulierung</b>					
Flächentarif	-0,194	<b>-0,426</b>	-0,283	<b>-0,511</b>	<b>-0,830</b>
Haustarif	0,116	0,112	-0,700	<b>-1,051</b>	-0,024
übertarifliche Bezahlung	-0,482	-0,274	-0,213	0,154	0,270
Gleichstellungsvereinbarungen	0,454	0,319	0,381	<b>0,434</b>	<b>0,496</b>
Betriebsrat	<b>-1,133</b>	<b>-1,053</b>	<b>-0,665</b>	-0,043	-0,056
<b>Internationalisierung</b>					
ausländisches Eigentum	-0,313	-0,348	-0,908	<b>0,676</b>	<b>1,281</b>
betrieblicher Exportanteil in %*	-0,022	0,005	-0,001	<b>0,012</b>	<b>0,013</b>
<b>Weiterbildung</b>					
externe Kurse	<b>-1,608</b>	<b>-1,260</b>	<b>-0,564</b>	-0,205	0,316
interne Kurse	<b>-1,192</b>	<b>-1,085</b>	<b>-0,739</b>	-0,008	0,439
Jobrotation	-1,019	-0,237	-0,361	0,061	0,275
selbstgesteuertes Lernen	<b>-1,360</b>	<b>-0,720</b>	<b>-1,086</b>	0,233	<b>0,795</b>
sonstige	<b>-0,857</b>	-1,040	-0,511	0,042	0,606
E-Learning (ja=1)	-0,178	<b>-0,731</b>	-0,267	0,103	<b>0,781</b>
<b>Personalstruktur</b>					
Anteil weiblicher Beschäftigter	<b>0,066</b>	<b>0,089</b>	<b>0,061</b>	<b>0,067</b>	<b>0,057</b>
Anteil unter 25-Jährige	<b>0,078</b>	<b>0,041</b>	<b>0,080</b>	<b>-0,036</b>	<b>-0,170</b>
Anteil 25- bis unter 35-Jährige	<b>0,046</b>	0,012	<b>0,036</b>	<b>0,019</b>	0,006
Anteil 45- bis unter 55-Jährige	-0,007	0,005	-0,004	-0,007	0,003
Anteil 55- bis 65-Jährige	-0,005	-0,011	<b>-0,031</b>	-0,012	-0,002
<b>Operierungsmerkmale</b>					
gute Ertragslage	-0,021	-0,111	-0,352	<b>0,413</b>	<b>0,648</b>
Arbeitszeit in Stunden	<b>0,241</b>	<b>0,183</b>	<b>0,159</b>	0,013	-0,042
<b>Ausstattungsmerkmale</b>					
junger Betrieb (nach 1990)	<b>-1,159</b>	<b>-0,944</b>	<b>-1,655</b>	-0,349	<b>-0,734</b>
Zentrale/Hauptverwaltung	-0,361	-0,252	0,130	<b>0,577</b>	<b>1,041</b>
Niederlassung/Filiale	0,098	0,142	0,364	0,002	<b>0,688</b>
Mittelinstanz	-0,509	-0,759	0,139	<b>-0,997</b>	-0,505
Einzelunternehmen	0,333	<b>0,656</b>	<b>0,792</b>	<b>-1,163</b>	<b>-2,618</b>
Personengesellschaft	0,336	0,193	0,225	-0,344	-0,469
Kapitalgesellschaft	-1,867	-0,359	<b>-1,547</b>	<b>0,439</b>	<b>1,202</b>
Körperschaft, Verein	<b>-3,035</b>	-1,458	<b>-2,839</b>	<b>-1,106</b>	<b>-2,115</b>
sonstige Rechtsform	<b>1,124</b>	-0,063	-0,041	<b>1,352</b>	<b>1,519</b>
öffentliches Eigentum	-0,585	<b>-2,108</b>	<b>-1,093</b>	-0,332	-0,323
städtische Randgebiete	<b>0,606</b>	<b>0,426</b>	0,324	-0,217	<b>-0,953</b>
ländliche Gebiete	0,469	0,114	0,095	<b>-0,679</b>	<b>-1,794</b>
weitere Kontrollvariablen	9 Branchen-, 7 Betriebsgrößenklassen-, 10 Bundesland-Dummies				

Organisationswandel	mittleres Lohnprofil		Betriebe im Umbruch	
	IIIa	IIIb	IVa	IVb
Outsourcing	<b>-1,047</b>	<b>-0,504</b>	0,608	0,249
Insourcing	-0,387	0,524	0,413	<b>2,209</b>
Verlagerung Ausland	0,215	0,176	-1,417	<b>-36,070</b>
Produktionsstruktur geändert	0,035	-0,197	-0,166	-0,238
Aufgabenstruktur geändert	<b>-0,449</b>	-0,064	-0,221	0,334
<b>technologischer Fortschritt</b>				
Forschung und Entwicklung	<b>-0,773</b>	-0,236	-0,352	0,196
Investitionssumme (log)*	<b>-0,246</b>	-0,046	-0,051	-0,132
Investitionssumme in IuK (log)*	<b>-0,124</b>	0,001	-0,151	0,048
<b>Institutionelle Regulierung</b>				
Flächentarif	0,102	<b>-0,435</b>	0,030	-0,504
Haustarif	-0,175	<b>-0,534</b>	0,462	0,299
übertarifliche Bezahlung	0,036	-0,049	0,073	0,423
Gleichstellungsvereinbarungen	0,242	<b>0,380</b>	0,245	0,132
Betriebsrat	<b>-0,801</b>	-0,027	-0,491	-0,058
<b>Globalisierungsindikatoren</b>				
ausländisches Eigentum	<b>-2,986</b>	-0,005	0,542	-1,384
betrieblicher Exportanteil in %*	<b>-0,017</b>	0,009	0,011	-0,006
<b>Weiterbildung</b>				
externe Kurse	-0,303	-0,207	-0,403	-0,366
interne Kurse	<b>-0,636</b>	-0,088	-0,542	<b>-0,999</b>
Jobrotation	-0,331	0,053	-0,373	-0,184
selbstgesteuertes Lernen	<b>-0,839</b>	0,214	<b>-1,215</b>	-1,517
sonstige	<b>-0,741</b>	0,023	<b>-0,919</b>	<b>-1,381</b>
E-Learning (ja=1)	-0,096	-0,162	0,397	-0,647
<b>Personalstruktur</b>				
Frauenanteil	<b>0,056</b>	<b>0,070</b>	<b>0,056</b>	0,066
Anteil unter 25-Jährige	<b>0,040</b>	-0,013	<b>0,049</b>	<b>0,074</b>
Anteil 25- bis unter 35-Jährige	<b>0,019</b>	0,007	<b>0,043</b>	<b>0,043</b>
Anteil 45- bis unter 55-Jährige	0,003	0,003	0,013	0,016
Anteil 55- bis 65-Jährige	-0,014	-0,010	<b>0,029</b>	0,029
<b>Operierungsmerkmale</b>				
gute Ertragslage	<b>-0,601</b>	-0,174	-0,223	-0,279
Arbeitszeit in Stunden	<b>0,122</b>	<b>0,095</b>	0,073	<b>0,146</b>
<b>Ausstattungsmerkmale</b>				
junger Betrieb (nach 1990)	-0,152	0,003	<b>-0,627</b>	<b>-0,893</b>
Zentrale/Hauptverwaltung	0,046	0,249	0,528	0,726
Niederlassung/Filiale	-0,394	0,230	-0,188	0,873
Mittelinstanz	-0,112	-0,365	-1,812	-0,302
Einzelunternehmen	<b>1,015</b>	0,210	-0,364	-1,098
Personengesellschaft	<b>0,689</b>	0,378	-0,465	0,160
Kapitalgesellschaft	-0,808	-0,016	-0,151	0,954
Körperschaft, Verein	<b>-2,185</b>	-0,559	<b>-2,878</b>	<b>-3,288</b>
sonstige Rechtsform	-0,514	<b>1,140</b>	0,075	0,324
öffentliches Eigentum	0,200	-0,209	-0,882	-0,396
städtische Randgebiete	<b>0,582</b>	0,004	-0,267	0,119
ländliche Gebiete	<b>0,510</b>	-0,118	-0,169	0,239
weitere Kontrollvariablen	9 Branchen,- 7 Betriebsgrößenklassen-, 10 Bundesland-Dummies			

## Regressionsdiagnostik

Kriterium	alle	Privat-	Kriterium	alle	Privat-
	Betriebe (N=5031)	wirtschaft (N=2534)		Betriebe (N=5031)	wirtschaft (N=2534)
<b>Mc Fadden</b> R <sup>2</sup>	0,318	0,339	<b>Maddala</b> R <sup>2</sup>	0,726	0,749
adjustiertes R <sup>2</sup>	0,263	0,236	<b>Cragg/ Uhler</b> R <sup>2</sup>	0,738	0,762
<b>Count</b> R <sup>2</sup>	0,483	0,491	<b>BIC</b>	- 24.153	- 8.854
adjustiertes R <sup>2</sup>	0,326	0,377	<b>BIC'</b>	- 1.828	- 587

### Erläuterungen zur Regressionsdiagnostik<sup>10</sup>:

**Mc Fadden Pseudo-R<sup>2</sup>:** ist der Likelihood-Ratio-Index des jeweiligen Regressionsmodells verglichen mit einem Modell nur mit der Konstante (dem Intercept). Da mit jeder neu aufgenommenen Variable unabhängig von ihrer Signifikanz das Pseudo-R<sup>2</sup> steigt, wird das in der adjustierten Version durch die Subtraktion der Anzahl der Parameter (nicht der unabhängigen Variablen) im Zähler für R<sup>2</sup> korrigiert.

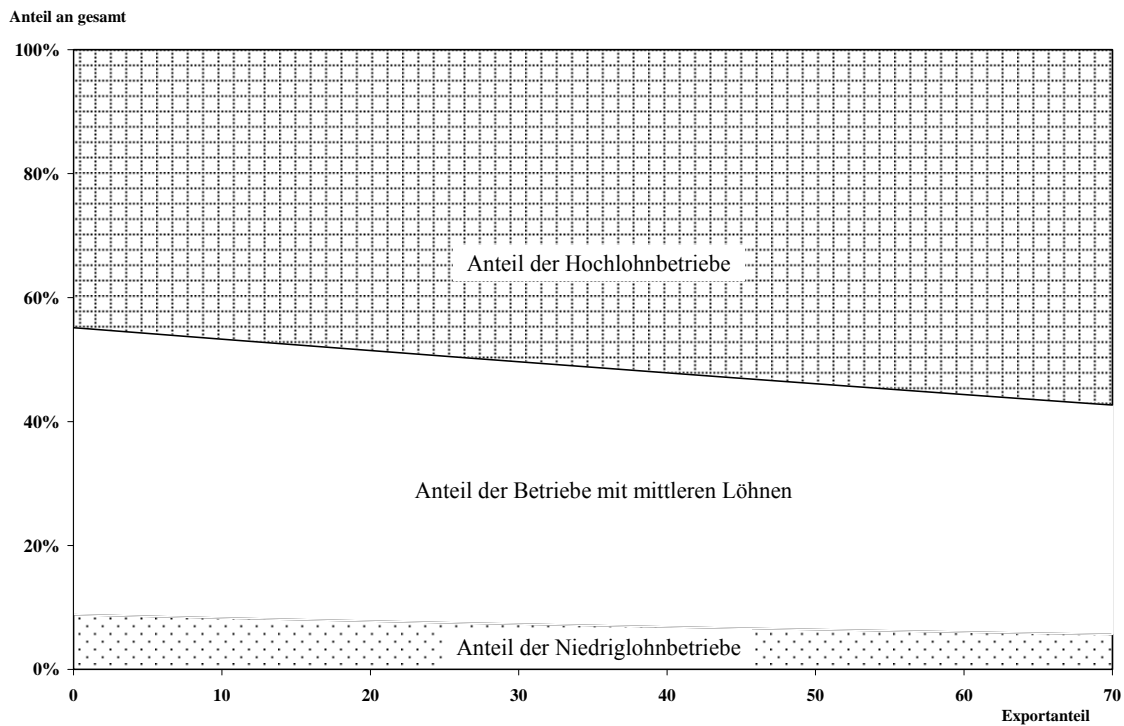
**Count R<sup>2</sup>:** gibt die Anteile an Beobachtungen für einzelne Ausprägungen der abhängigen kategorialen Variable an, die vom Modell korrekt vorhergesagt werden. In der adjustierten Version werden die korrekten Vorhersagen jenseits einer Vorhersage angegeben, wenn für jeden Fall die zahlenmäßig am stärksten auftauchende Kategorie eingesetzt wird.

**Maximum-Likelihood-R<sup>2</sup> von Maddala und Cragg/ Uhler R<sup>2</sup>:** weitere Messvorschläge in Analogie zum Mc Fadden Pseudo-R<sup>2</sup>. Im Cragg/ Uhler R<sup>2</sup> erfolgt eine Normierung des Wertebereichs des R<sup>2</sup> von Maddala.

**BIC und BIC':** Dieses Bayesianische Kriterium vergleicht unter anderem Regressionsmodelle in ihrer genesteten und ungenesteten Form. In der Bayesianischen Logik überprüfen die BIC- und BIC'-Werte, welches Modell eher der tatsächlich vorhandenen Datensatzstruktur zugrunde gelegt wurde. Je negativer die Werte, umso besser ist das aufgestellte Modell. Verglichen wird das Likelihood-Ratio von  $\chi^2$ , wobei die Anzahl der Freiheitsgrade der Zahl der unabhängigen Regressoren (nicht Anzahl der Parameter) entspricht.

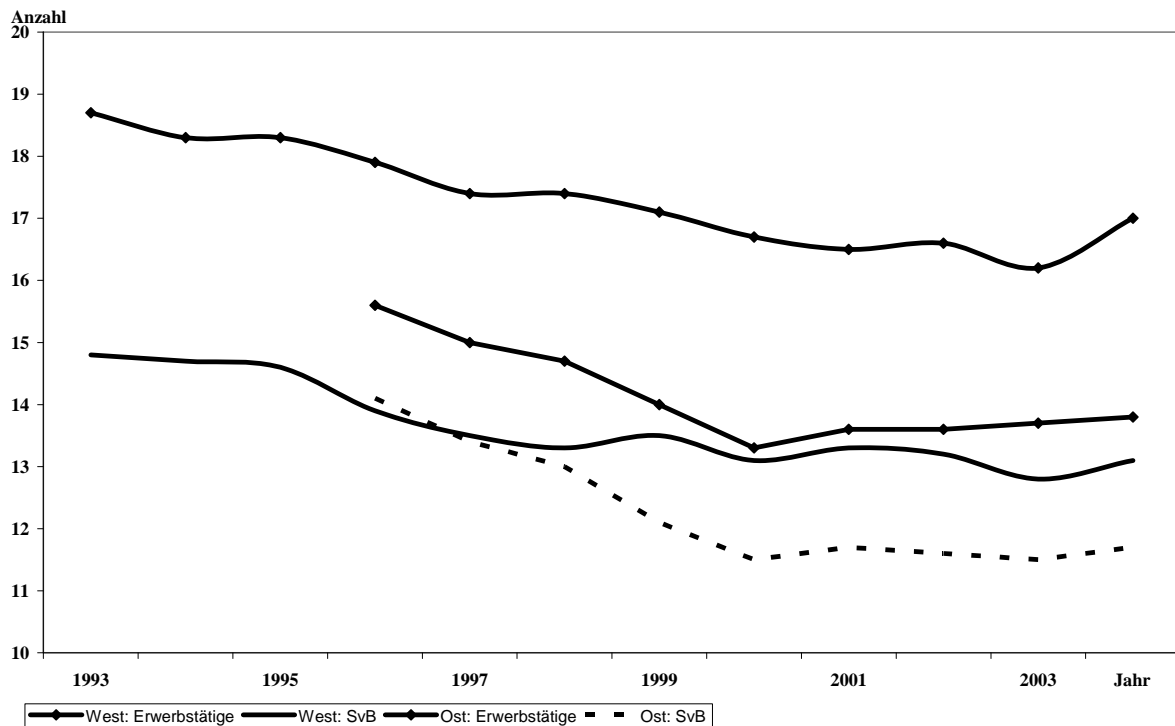
<sup>10</sup> Die Erläuterungen basieren auf den Ausführungen von Long/ Freese (2003) im Abschnitt 3.4.

## 6.2. Vorhergesagte Wahrscheinlichkeiten für die Segmente der Betriebstypen in Abhängigkeit von ihrem Exportanteil, sofern sie einen ausländischen Mehrheitseigner haben



**Quelle:** LIAB-Querschnittmodell für das Jahr 2004, Betriebe ab fünf Beschäftigte, die in den Jahren 2003 und 2004 an der IAB-Betriebspanelbefragung teilgenommen haben

### 6.3 Anhang A3: Entwicklung der Anzahl aller Erwerbstätigen und der sozialversicherungspflichtigen Beschäftigung auf der Firmenebene in Deutschland im Zeitraum 1993 bis 2004



Quelle: IAB-Betriebspanel 1993 bis 2004, eigene Berechnungen

### 6.4 : Anhang A4: Entwicklung der durchschnittlichen Beschäftigtenanzahl pro Betrieb

	Westdeutschland		Ostdeutschland	
	Erwerbstätige	SvB	Erwerbstätige	SvB
1993	18,7	14,8		
1994	18,3	14,7		
1995	18,3	14,6		
1996	17,9	13,9	15,6	14,1
1997	17,4	13,5	15	13,4
1998	17,4	13,3	14,7	13
1999	17,1	13,5	14	12,1
2000	16,7	13,1	13,3	11,5
2001	16,5	13,3	13,6	11,7
2002	16,6	13,2	13,6	11,6
2003	16,2	12,8	13,7	11,5
2004	17	13,1	13,8	11,7

Quelle: IAB-Betriebspanel 1993 bis 2004, eigene Berechnungen

## 7. Literatur

- Abowd, J. / Kramarz, F. (1999): The analysis of labor markets using matched employer-employee data. In O. Ashenfelter & D. Card (eds.) 'Handbook of Labor Economics', Vol. 3B, Elsevier, Amsterdam, chapter 40, pp. 2567–627.
- Alda, Holger; Bender, Stefan; Gartner, Hermann (2005): The linked employer-employee dataset created from the IAB establishment panel and the process-produced data of the IAB (LIAB). In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 125, H. 2, S. 327-336
- Alda, H. / Hauss, F./ Land, R./ Willisch, A. (2004): Erwerbsverläufe und sekundärer Integrationsmodus. Ergebnisse einer empirischen Untersuchung, Berliner Debatte Initial, Jg. 15, Nr.2: 70-85
- Bacher, J. (1996): Clusteranalyse – eine anwendungsorientierte Einführung, München
- Backhaus, K./ Erichson, B./ Plinke, W. / Weiber, R. (2006): Multivariate Analysemethoden – Eine anwendungsorientierte Einführung, 11. Auflage, Berlin / New York, Kapitel 8: Clusteranalyse.
- Calinski, T. / Harabsaz, K. (1974): A dendrite method for cluster analysis, Communications in Statistics 3: 1-27
- Duda, R.O. / Hart, P. E. (1973): Pattern, Classification, and Scene Analysts, John Wiley and Sons
- Fynch, H. (2005): Comparison of Distance Measures in Cluster Analyses with Dichotomous Data, Journal of Data Science 3: 85 - 100
- Hausman, J./ McFadden, D. (1984): Specification tests for the multinomial logit model, Econometrica 52 (5): 1219 - 1240
- Lazaer, E.P. / Shaw, K. J. (2008): An international comparison of the Structure of Wages. The Chicago Press
- Long, J. S./ Freese, J. (2001): Regression Models for Categorical Dependant Variables using STATA, 1<sup>st</sup> edition, Stata Corporation, Texas
- McFadden, D./ Tye, W. /Train, K. (1978): An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model. TRANSPORTATION RESEARCH RECORD: FORECASTING PASSENGER AND FREIGHT TRAVEL, No. 637, 39-46, Transportation Research Board, 1978.
- Rüb, F. / Werner, D. (2007): Typisierung von SGB II-Trägern, IAB-Forschungsbericht 1/2007, Nürnberg

- Small, K.A. / Hsiao, C. (1985): Multinomial logit specification tests, *International Economic Review* 26 (3): 619 - 627
- Statistisches Bundesamt (2007): Verbraucherpreisindizes für Deutschland, <https://www-genesis.destatis.de>
- Wagner, J. (1998): Exportverhalten und Exportdynamik in der niedersächsischen Industrie. In: Gerlach, K. / Hübler, O. / Meyer, W. (Hrsg.): *Ökonomische Analysen betrieblicher Strukturen und Entwicklungen. Das Hannoveraner Firmenpanel*, Frankfurt u.a., 278-299