

Methodenbericht zur Modellbasierten Datenergänzung in der Leistungs- und Strukturstatistik

Gerlinde Dinges, Martin Haitzmann, Sabine Zach
Statistik Austria, Wien

Abstract: EU-harmonized Structural Business Statistics, as introduced in Austria in 1997, are one of the main sources for assessing the structure and the development of the domestic economy at national and regional level. Owing to national economic actors' requests, these statistics have been designed as a so-called cut-off survey. Small enterprises – about 89% of the relevant population – are disburdened from the legal obligation to report statistically relevant data. Nonetheless, according to EU requests, STATISTICS AUSTRIA has to ensure that the transmitted data correspond to the structure of the whole population of the economic activities covered by the statistics. The aim of this article is to inform the reader, the user of the data, about the methodological background of the estimation process in use as well as on its current implementation.

Zusammenfassung: Die 1997 eingeführte EU-harmonisierte Leistungs- und Strukturstatistik stellt eine der zentralen Informationsquellen zur Beurteilung der strukturellen Entwicklung österreichischer Unternehmen auf nationaler und regionaler Ebene dar. Dem Anliegen der Wirtschaft entsprechend wird diese Statistik als Konzentrationsstichprobe geführt. Klein- und Kleinstunternehmen – rund 89% der Grundgesamtheit – sind dadurch von statistischen Verpflichtungen weitestgehend befreit. Gemäß EU-Vorgaben ist jedoch zu gewährleisten, dass die übermittelten Daten die Struktur der Grundgesamtheit aller statistischen Einheiten der erfassten Wirtschaftsbereiche widerspiegeln. Der folgende Beitrag ist der methodischen Konzeption des dafür angewandten Modells zur Datenergänzung und dessen Umsetzung in der Praxis gewidmet. Zur praktischen Veranschaulichung dienen ausgewählte Wirtschaftszweige und Merkmale der Leistungs- und Strukturstatistik 2008.

Keywords: Bewusste Auswahlverfahren, Cut-Off-Survey, Verwaltungsdaten, Konzentrationsprinzip, Wirtschaftsstatistik.

1 Einleitung

Die von STATISTIK AUSTRIA jährlich erstellte Leistungs- und Strukturstatistik besteht in dieser Form seit dem Berichtsjahr 1997 und wird nach einem Konzept durchgeführt, das den Vorgaben der EU-Verordnung über die strukturelle Unternehmensstatistik in den Produktions-, Handels- und Dienstleistungsbereichen entspricht. Leistungs- und Strukturstatistiken ermöglichen eine ausführliche Analyse der Unternehmens- und Beschäftigtenstruktur, der Leistung von Unternehmen und Betrieben sowie der regionalen Verteilung von Betrieben und Arbeitsstätten. Sie erlauben wichtige Aussagen über

die Struktur, Tätigkeit und Wettbewerbsfähigkeit der Unternehmen auf nationaler und regionaler Ebene und im internationalen Vergleich und bilden daher eine wichtige Entscheidungsgrundlage für Politik und Wirtschaft. Darüber hinaus liefern sie wichtige Basisdaten für die Volkswirtschaftliche und Regionale Gesamtrechnung, für Wirtschaftsprognosen und Marktforschung.

Bis zum Berichtsjahr 2001 wurde zur Erstellung der österreichischen Leistungs- und Strukturstatistik eine geschichtete Stichprobenerhebung mit freier Hochrechnung durchgeführt. Seit dem Berichtsjahr 2002 wird die Primärerhebung dem Anliegen der Wirtschaft entsprechend als Vollerhebung mit variablen Abschneidegrenzen geführt. Diese Erhebungsform wird auch als “Konzentrationsstichprobe” bezeichnet. Auskunftspflicht besteht nur bei Überschreitung normierter Schwellenwerte (Beschäftigten- und/oder Umsatzschwelle) unter Berücksichtigung eines bestimmten Repräsentanzgrades (Mindestumfassung des Gesamtumsatzes einer Branche). Klein- und Kleinstunternehmen sind dadurch von statistischen Verpflichtungen weitestgehend ausgenommen. Dem nationalen Ziel einer maximalen Respondentenentlastung stehen aber EU-Vorgaben gegenüber, gemäß derer zu gewährleisten ist, dass die übermittelten Daten die Struktur der Grundgesamtheit aller statistischen Einheiten der erfassten Wirtschaftsbereiche widerspiegeln. Die vorgegebenen rechtlichen und daraus resultierenden konzeptiven Rahmenbedingungen lassen keine konventionelle Hochrechnung zu, da bei einer “Konzentrationsstichprobe” das Auswahlverfahren kein repräsentatives Sample liefert. Deshalb wurde ein modellbasierter Ansatz zur Datenergänzung entwickelt, welcher den Informationsvorrat der Primärerhebung ausschöpft und zusätzlich Informationen aus Verwaltungsquellen effizient nutzt.

Ein ähnlicher Weg zur Darstellung der Grundgesamtheit wird seit dem Berichtsjahr 2008 auch in den monatlichen Konjunkturstatistiken im Produzierenden Bereich durch Einbindung von primärstatistischen Ergebnissen, die direkt beim auskunftspflichtigen Unternehmen erhoben werden, und sekundärstatistischen Ergebnissen, die aus Verwaltungsquellen oder anderen statistischen Erhebungen verfügbar sind, beschritten.

Im vorliegenden Methodenbericht wird die Konzeption des Modells zur Datenergänzung in der Leistungs- und Strukturstatistik und dessen Umsetzung in der Praxis beschrieben.

2 Die Erhebung

Die EU-Verordnung über die strukturelle Unternehmensstatistik (EG Nr. 295/2008 vom 11. März 2008) bildet die Rechtsgrundlage, um die Struktur, Tätigkeit, Wettbewerbsfähigkeit und Leistung der Unternehmen auf regionaler, nationaler und internationaler Ebene anhand harmonisierter Daten der Mitgliedstaaten der Europäischen Union vergleichen zu können. Die Leistungs- und Strukturstatistik-Verordnung (BGBl. II Nr. 428/2003 idF BGBl. II Nr. 266/2009 vom 20. August 2009) bildet seit dem Berichtsjahr 2002 die nationale Rechtsgrundlage für die Erstellung von Leistungs- und Strukturstatistiken im Unternehmensbereich. Darin wurde im Besonderen zur Entlastung der Respondenten eine Vollerhebung mit normativ vordefinierten Abschneidegrenzen und gleichzeitig die vermehrte Verwendung von Verwaltungsquellen, die Anwendung statistischer Modellberechnungen

sowie die Nutzung von Synergien zwischen Erhebungen im Sinne des Bundesstatistikgesetzes 2000 (BGBl. I Nr. 163/1999 zuletzt geändert durch BGBl. I Nr. 125/2009 vom 16. Dezember 2009) festgelegt.

Grundsätzlich erstreckt sich die Leistungs- und Strukturhebung auf Unternehmen und Arbeitsgemeinschaften die gemäß ÖNACE 2008 (siehe Zeller and Karner, 2009) dem Produzierenden Bereich oder dem Handel- und Dienstleistungsbereich zuzuordnen sind und die Tätigkeit selbständig, regelmäßig und in der Absicht zur Erzielung eines Ertrags oder sonstigen wirtschaftlichen Vorteils ausüben. Dabei besteht Auskunftspflicht

- im Produzierenden Bereich (Abschnitte B bis F der ÖNACE 2008) für alle Unternehmen, die am Stichtag (dem 30. September des der Berichtsperiode vorangegangenen Kalenderjahres) 20 und mehr Beschäftigte hatten. Darüber hinaus muss die Erhebungsmasse mindestens 90% des Gesamtumsatzes in jedem der Wirtschaftszweige gemäß den Abteilungen 05 bis 43 der ÖNACE 2008 erfassen. Wird dieses Repräsentanzkriterium mit Hilfe der voll erhobenen Schicht nicht erreicht, so besteht Auskunftspflicht auch über statistische Einheiten mit weniger als 20 Beschäftigten, die in den zwölf Monaten, die dem Stichtag vorangegangen sind, in Summe einen Umsatz (exkl. Umsatzsteuer) von mindestens einer Million Euro erzielten;
- im Handels- und Dienstleistungsbereich (Abschnitte G bis N sowie Abteilung 95 der ÖNACE 2008) für alle Unternehmen, die für das Berichtsjahr eine bestimmte Umsatzschwelle (branchenspezifisch entweder 300 Tsd., 850 Tsd., 1.8 Mio oder 3 Mio Euro) überschritten haben. In einigen Wirtschaftszweigen besteht darüber hinaus auch Auskunftspflicht über statistische Einheiten, welche die Umsatzschwelle nicht überschreiten, wenn eine bestimmte Beschäftigtenschwelle (branchenspezifisch, 10 oder 20 Beschäftigte) überschritten wird.

Weitere Informationen zu Erhebungskonzept und Auskunftspflicht findet der interessierte Leser in der *Standarddokumentation zur Leistungs- und Strukturstatistik* (2007a, 2007b).

3 Weshalb die Erhebung modellbasiert ergänzen?

3.1 Konzeptionelle Vorgaben

Das Erhebungskonzept der Leistungs- und Strukturhebung wird auch als Cut-Off-Census (Vollerhebung aller Einheiten, die eine bestimmte Schwelle überschreiten) oder "Konzentrationsstichprobe" bezeichnet, da weniger (ge)wichtige Information nicht erfasst wird. Der bewusste Daten-Cut-Off liefert Informationen, die keinem repräsentativen Sample im Sinne der Stichprobentheorie entsprechen.

In der Wirtschaftsstatistik sind Konzentrationserhebungen gängige Praxis (siehe z.B. Elisson and Elvers, 2001). Grundidee dieser Methode ist es, die hohe Konzentration der Verteilung der Merkmalssummen wirtschaftsstatistischer Daten zu nutzen und anhand möglichst weniger erhobener Einheiten eine möglichst hohe primärstatistische Abdeckung jener Merkmale zu erreichen, die im Mittelpunkt des Interesses stehen.

Abbildung 1 zeigt, dass im Berichtsjahr 2008 nur Daten für 11% der rund 300.000 Unternehmen der Grundgesamtheit erhoben werden mussten, um bereits 69% der Gesamtbeschäftigten und 88% der Umsatzerlöse primärstatistisch zu erfassen.

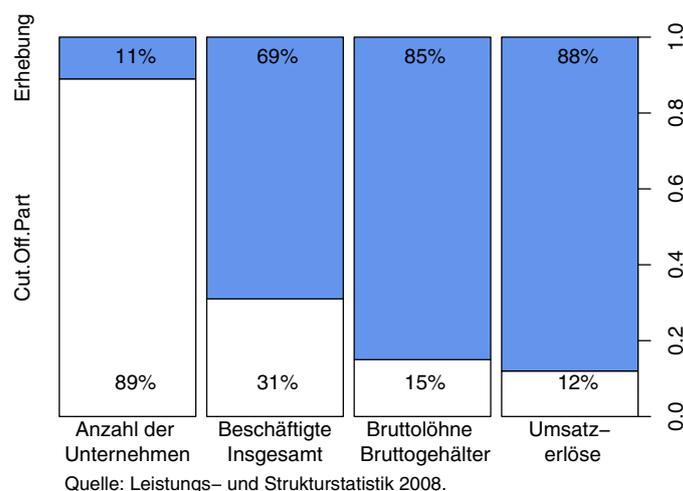


Abbildung 1: Merkmalskonzentration in der Grundgesamtheit – Primärabdeckung ausgewählter Unternehmensmerkmale.

3.2 Gesetzliche Vorgaben

Das Bundesstatistikgesetz 2000 enthält die Verpflichtung, Klein- und Kleinstunternehmen in höchstmöglicher Weise von der Auskunftspflicht auszunehmen. Gemäß nationaler Leistungs- und Strukturstatistik-Verordnung sind Unternehmen nur dann meldepflichtig, wenn diese die für ihren Wirtschaftszweig gesetzlich vorgegebenen Meldeschwellen überschreiten (vgl. Abschnitt 2). Dem Einsatz klassischer Hochrechnungsverfahren wird dadurch eine konzeptionelle und methodische Grenze gesetzt. Gemäß den europäischen Bestimmungen können jedoch die Mitgliedstaaten die erforderlichen Daten nach dem Grundsatz der verwaltungstechnischen Vereinfachung durch eine Kombination verschiedener Quellen – verbindliche Erhebungen oder andere Quellen, die in Bezug auf Genauigkeit und Qualität zumindest gleichwertig sind, sowie statistische Schätzverfahren – beschaffen (Subsidiaritätsprinzip).

4 Modellbasierte Datenergänzung

Durch die modellbasierte Datenergänzung (MDE) wird der primärstatistisch erhobene Datenkörper – wie in Abbildung 2 skizziert – um den ursprünglichen Cut-Off-Part ergänzt. Die Struktur der Grundgesamtheit kann durch die MDE als Kombination aus Primärdaten, Sekundärdaten und Modellrechnung ohne grundsätzlichen Informationsverlust dargestellt werden. Die Datenerstellung erfolgt dabei auf Mikroebene. Für jede statistische Einheit (Unternehmen) der Grundgesamtheit steht somit ein vollständiger Datensatz aller Merkmale der Leistungs- und Strukturstatistik für weitere Ergebnisdarstellungen zur Verfügung.

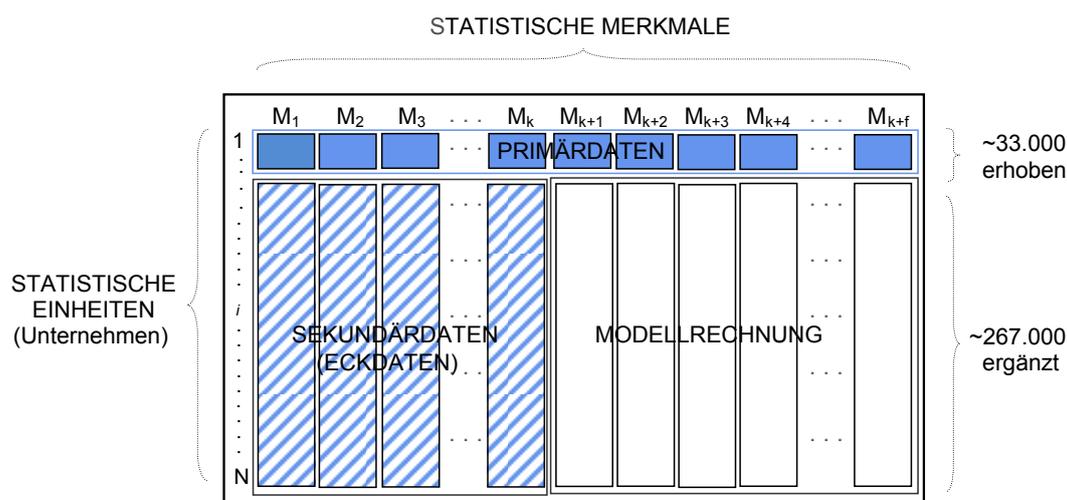


Abbildung 2: Datenkörper der Mikrodaten der Leistungs- und Strukturstatistik.

4.1 Schritte zur Mikrodatenerstellung

Als Basis für die Mikrodatenerstellung dient das statistische Unternehmensregister der STATISTIK AUSTRIA. Das Unternehmensregister erfasst alle wirtschaftsstatistisch relevanten Einheiten mit mindestens einem unselbständig Beschäftigten oder mehr als 10.000 Euro Jahresumsatz sowie Einheiten des Staates und Non-Profit Organisationen. Eine statistische Einheit ist genau dann der Grundgesamtheit der Leistungs- und Strukturstatistik zugeordnet, wenn sie gemäß Unternehmensregister wirtschaftlich aktiv ist und eine Haupttätigkeit im Sinne einer Unterklasse der ÖNACE 2008 ausübt, welche dem Erfassungsbereich (Produktion sowie Handel und Dienstleistungen) zuzurechnen ist. Bezogen auf das Berichtsjahr 2008 umfasst die Grundgesamtheit $N \approx 300.000$ Unternehmen (vgl. auch Abbildung 2). Für die etwa 267.000 zu ergänzenden statistischen Einheiten können sogenannte Eckdaten (*wirtschaftliche Aktivität, Anzahl der unselbständig Beschäftigten, Umsatzerlöse, Löhne und Gehälter*) unter Einbindung von Verwaltungsquellen erstellt werden. Alle sonstigen Merkmale eines Unternehmens werden mithilfe der Eckdaten und aktuellen primärstatistischen Zusammenhängen erstellt. Sowohl beschäftigtenbezogene Hauptmerkmale (*Sozialleistungen, Arbeitsstunden, ...*) als auch umsatzbezogene Hauptmerkmale (*Waren- und Dienstleistungskäufe, Erträge, Lagerbestand, ...*) werden dabei über ein robustes lineares Regressionsmodell berechnet (siehe Abschnitt 4.1.4). Die tiefere Aufgliederung der Hauptmerkmale zur Erstellung der Detailmerkmale erfolgt in weiterer Folge mittels Anteilsrechnung (siehe Abschnitt 4.1.5). Vorrangiges Ziel der MDE ist die Erzeugung von Mikrodaten, die unter den gegebenen gesetzlichen Vorgaben ein Optimum an Qualität des Datenkörpers gewährleisten und aufgrund ihrer Vollständigkeit eine beliebige Ergebnisdarstellung der Grundgesamtheit erlauben.

4.1.1 Erstellung der Eckdaten aus Verwaltungsquellen

Über die Verknüpfungen der Unternehmen des statistischen Unternehmensregisters (siehe Haslinger, 2004) zu den vorliegenden Umsatzsteuermeldungen der Finanzbehörde so-

wie den Beschäftigtenmeldungen des Hauptverbands der österreichischen Sozialversicherungsträger (HV) werden für alle nicht erhobenen und folglich zu schätzenden Einheiten zuerst die zugehörigen Eckdaten erstellt. Als Eckdaten werden jene k Merkmale bezeichnet, die sekundärstatistisch aus Verwaltungsquellen übernommen werden können und an die gebunden die Schätzung der restlichen Merkmale erfolgt. Beginnend mit der Zuweisung der jeweiligen unternehmensspezifischen Merkmale (z.B. *Wirtschaftsaktivität, Rechtsform, ...*), die aus dem Unternehmensregister für jede statistische Einheit zur Verfügung stehen, erfolgt danach die Einbindung der *unselbständig Beschäftigten* des HV, gegliedert nach Geschlecht und Qualifikationen (*Arbeiter, Angestellte, Lehrlinge*). Anschließend werden die Werte der Jahresumsatzsteuermeldungen bzw. die monatlichen Umsatzsteuervoranmeldungen der Finanzbehörde übernommen.

Hauptverbands- und Umsatzsteuerdaten werden bereits seit dem Berichtsjahr 2002 zur Erstellung der Eckdaten herangezogen. Beginnend mit dem Berichtsjahr 2008 (Ergebnisse erstmals publiziert im Juli 2010) können die bisherigen Eckdaten durch eine weitere sekundärstatistische Quelle wesentlich verstärkt werden. Im Zuge des MEETS¹-Projekts "*Feasibility study of implementing wage tax data in Structural Business Statistics*" wurde die Einbindung von Lohnzetteldaten zur Erstellung der *Bruttolöhne und -gehälter* umfassend geprüft. Die Lohnzetteldaten stammen aus der Lohnsteuerdatenbank des Bundesministeriums für Finanzen. Die Einbindung der Lohnzetteldaten für nicht meldepflichtige Einheiten kann in einigen Wirtschaftsteilbereichen als wesentliche Maßnahme zur Verbesserung der Datenqualität betrachtet werden, da anhand von Konzentrationsdaten für inhomogene Branchen mit sehr divergierender Verdienststruktur und hohem Teilzeitanteil durch reine Modellrechnung kaum unverzerrte Schätzer für die Datenergänzung gewonnen werden können (vgl. dazu die durch Box-Plots visualisierten Verteilungen der Pro-Kopf-Bruttoverdienste der ÖNACE = 68209 (*Vermietung, Verpachtung*) in Abbildung 3).

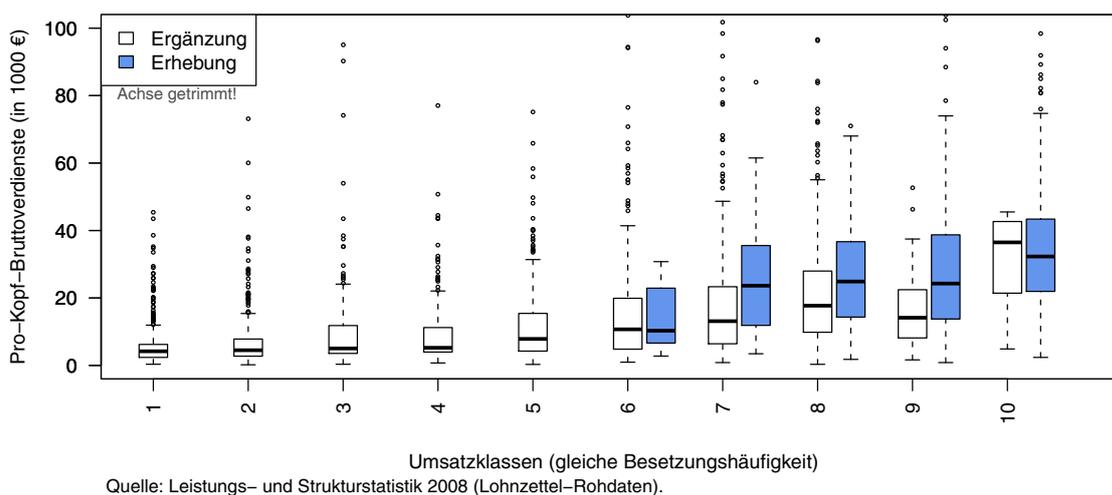


Abbildung 3: Strukturunterschiede innerhalb einer Branche – Verteilung der Pro-Kopf-Bruttoverdienste für 3.093 Unternehmen der ÖNACE = 68209 (*Vermietung, Verpachtung*).

¹MEETS – Modernisation of European Enterprise and Trade Statistics. Projects on modernisation of Business Statistics.

4.1.2 Imputation fehlender Eckdaten

Bei fehlender Beschäftigtenmeldung vom HV kann aufgrund des hohen Verknüpfungsgrades im Unternehmensregister davon ausgegangen werden, dass das Unternehmen keine unselbständig Beschäftigten hat. Nicht (rechtzeitig) vorliegende Jahresumsatzsteuermeldungen können für $\approx 15\%$ der Unternehmen durch aggregierte Umsatzsteuervoranmeldungen ersetzt werden, wobei fehlende Monatsmeldungen unter Berücksichtigung der individuellen Unternehmensentwicklung und der zugehörigen Branchenentwicklung imputiert werden. Unterjährig fehlende Werte treten zum einen auf, da von einem Teil der Unternehmen keine fristgerechte Erstattung der Umsatzsteuervoranmeldung bei der Finanzbehörde erfolgt. Zum anderen bedingen Ausnahmen von der monatlichen Umsatzsteuervoranmeldepflicht für kleinere Unternehmen, eine höhere Ausfallwahrscheinlichkeit in Abhängigkeit von der Unternehmensgröße. Für die weiteren Berechnungen gilt:

Die Grundgesamtheit ist unterteilt in L Branchen B_1, \dots, B_L ; $B = \bigcup_{j=1}^L B_j$, $B_j \cap B_r = \emptyset$.

- $j = 1, \dots, L$... laufender Branchenindex
- $i = 1, \dots, N$... Index der statistischen Einheiten in der Grundgesamtheit
- $x_{i,m}$... Umsatzsteuervoranmeldung einer stat. Einheit i im Monat m
- $M_i := \{m : m \in \{1, 2, \dots, 12\}, x_{i,m} \text{ liegt für Unternehmen } i \text{ im Monat } m \text{ vor}\}$
- $I_j := \{i : \text{Unternehmen mit Index } i \text{ ist in Branche } j \text{ der Grundgesamtheit}\}$
- $I_j = E_j \oplus S_j$, mit $E_j := \{i \in I_j : \text{erhoben}\}$ und $S_j := \{i \in I_j : \text{nicht erhoben}\}$.

Dem branchenspezifischen mittleren Umsatzanteil des Berichtsmonats m entspricht:

$$\bar{p}_{j,m} = \sum_{i \in S'_j} \frac{x_{i,m}}{\sum_{t=1}^{12} x_{i,t}} \left(\sum_{i \in S'_j} 1 \right)^{-1} \quad \text{mit } S'_j := S_j \setminus \{i : |M_i| < 12\}. \quad (1)$$

Dem imputierten Umsatz einer Unternehmenseinheit i für den Berichtsmonat m entspricht:

$$\hat{x}_{i,m} = \begin{cases} (\sum_{t=1}^{12} x_{i,t} \mathbf{I}_{i,t} / \sum_{t=1}^{12} \bar{p}_{j,t} \mathbf{I}_{i,t}) \bar{p}_{j,m} & \text{für } |M_i| \geq 6 \\ \text{Vorjahreswert oder Branchensubstitut} & \text{für } |M_i| < 6, \end{cases} \quad (2)$$

mit $\mathbf{I}_{i,t} = 1$ wenn $x_{i,t}$ vorliegt und $\mathbf{I}_{i,t} = 0$ sonst.

Tabelle 1: Anteil fehlender Monatsmeldungen bei Erstellung des Jahreswertes aus aggregierten Umsatzsteuervoranmeldungen (Auszug aus den 14 ÖNACE-Abschnitten).

ÖNACE-Abschnitt	Juli 2007 – Juni 2009	Min.	Max.
Herstellung von Waren		1.1%	2.3%
Bau		1.2%	2.4%
Handel; Instandhaltung und ...		1.2%	2.4%
Erbringung von Finanzdienstleistungen ...		0.9%	3.8%

Sparklines: grau schattierte Balken entsprechen den mittleren 50% einer Beobachtungsreihe

Der Anteil der Unternehmen, für die bei Übernahme der aggregierten Umsatzsteuervoranmeldungen eine unterjährige Imputation fehlender Monatsmeldungen erfolgen muss, ist nicht nur abhängig von der Unternehmensgröße, sondern weist auch eine deutlich erkennbare saisonale Komponente auf (siehe Tabelle 1).

Liegen für ein Unternehmen im Berichtsjahr weniger als sechs Umsatzsteuervoranmeldungen vor, wird eine fehlende Jahresumsatzsteuermeldung anhand der Vorjahresmeldung des Unternehmens (ein über branchenspezifische Umsatzentwicklungen fortgeschriebener Wert aus Umsatzsteuererklärung oder Leistungs- und Strukturdaten) oder, gewichtet über die Beschäftigten des Unternehmens, anhand eines Branchensubstituts imputiert ($\approx 10\%$ der Fälle). Die Berechnung branchenspezifischer Schätzer zur Imputation fehlender Werte erfolgt über die schwerpunktmäßige wirtschaftliche Aktivität einer statistischen Einheit unter Anwendung eines robusten linearen Regressionsmodells (mehr dazu in Abschnitt 4.1.4).

Die Information aus den Lohnzetteldaten wird über die HV-Beschäftigten einer statistischen Einheit eingebunden, wobei die Beschäftigten (gegliedert nach Qualifikation) als Hilfsvariablen dienen. Bei etwa 9% der Unternehmen ist keine direkte Übernahme der Bruttolöhne und -gehälter möglich, da die Beschäftigtenstruktur der beiden Quellen divergiert. Für diese Einheiten wird das unternehmensspezifische Lohn-/Gehaltsvolumen entweder entsprechend der jeweiligen HV-Struktur gewichtet oder die Bruttolöhne und -gehälter über die HV-Beschäftigten des Unternehmens anhand branchenspezifischer medianer Jahreslöhne/-gehälter imputiert.

Nach Erstellung und Vervollständigung der Eckdaten erfolgt die modellbasierte Berechnung der restlichen beschäftigten- und umsatzbezogenen Merkmale (*Arbeitsstunden, Waren- und Dienstleistungskäufe, Bruttoinvestitionen, Lagerbestand, etc.*).

4.1.3 Auswahl der Modellbasis

Um den nicht erhobenen statistischen Einheiten möglichst ähnliche erhobene Einheiten zugrunde zu legen, erfolgt eine iterative Auswahl der zur Schätzung erforderlichen Modellbasis innerhalb von Umsatzklassen und Wirtschaftstätigkeit. Die Auswahl der Strukturspender erfolgt dabei *bottom-up*, beginnend auf der tiefst möglichen ÖNACE-Gliederungsebene, welche bei der Wirtschaftstätigkeit den rund 540 ÖNACE 2008-Unterklassen der Leistungs- und Strukturstatistik entspricht. Aufgrund der systematischen Stichprobenauswahl mittels Cut-Off-Schwellen werden zudem nur die kleinsten erhobenen Unternehmen einer Branche ausgewählt. Diese werden über das unterste Dezil der primärstatistischen Umsatzverteilung einer Branche bestimmt, welches von mindestens 30 statistischen Einheiten nicht überschritten wird. Für die Klassenauswahl der Modellbasis gilt:

- x_i ... Jahresumsatz einer statistischen Einheit i
- d_j^p ... $p\%$ -Dezil des Jahresumsatzes x der Branche j der Erhebung
- d_j^α ... kleinstes $p\%$ -Dezil einer Branche j mit ausreichender Besetzung, wobei $\alpha = \min\{p \in \{10, 20, \dots, 70\} : (p/100)|E_j| \geq 30\}$
- $E_j^\alpha := \{i \in E_j : x_i < d_j^\alpha\}$... Indexmenge der Unternehmen in der Modellbasis der Branche j
- $e_j^\alpha = |E_j^\alpha|$... Anzahl der Unternehmen in der Modellbasis der Branche j .

Bei nicht ausreichender Primärbesetzung bis zum 7. Dezil erfolgt die Auswahl auf einer übergeordneten ÖNACE-Ebene. Dadurch sollen branchenspezifische Strukturunterschiede der verschiedenen Wirtschaftszweige als auch mögliche unternehmensgrößenspezifische Strukturunterschiede innerhalb einer Branche Berücksichtigung finden.

Schon am einfachen Beispiel der durchschnittlichen *Anteile der Waren- und Dienstleistungskäufe (Vorleistungen) am Unternehmensumsatz* in Abbildung 4 ist ersichtlich, dass sich die durch Box-Plots visualisierten Verteilungen der ÖNACE 2008-Abteilungen² deutlich unterscheiden, und zusätzlich zur Unternehmensgröße auch die Berücksichtigung der branchenspezifischen Strukturunterschiede (also der Wirtschaftstätigkeit einer Einheit) unbedingt erforderlich ist.

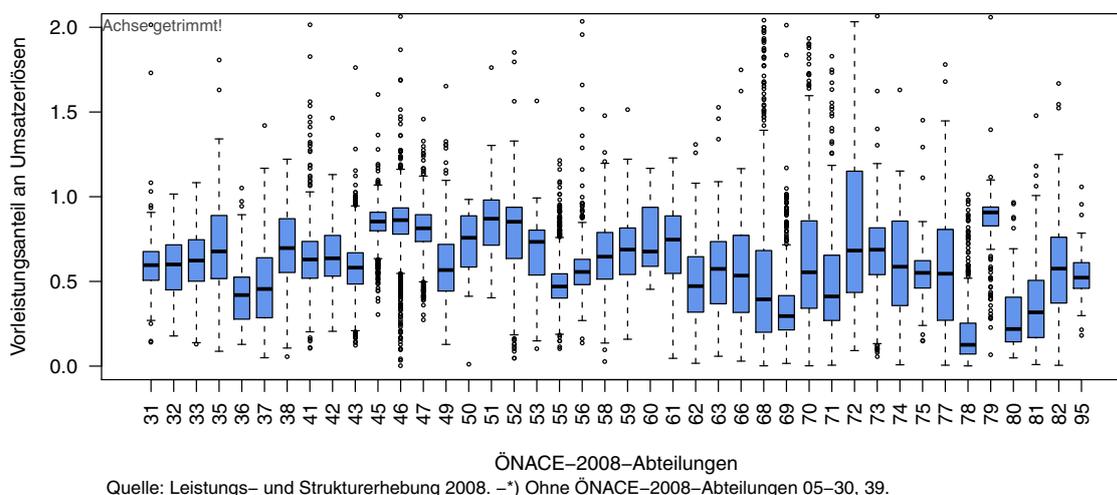


Abbildung 4: Strukturunterschiede zwischen den Branchen – Verteilungen der Anteile der Waren- und Dienstleistungskäufe der Unternehmen der ÖNACE 2008-Abteilungen.

Zum Zeitpunkt der Durchführung der MDE ($t + 15$ Monate nach der Stichprobenziehung) weist ein gewisser Teil der meldepflichtigen Einheiten einen Unternehmensumsatz auf, welcher unter der branchenspezifischen Meldeschwelle liegt und vice versa. Aufgrund dieser Fluktuation an der Meldeschwelle kann den größeren zu schätzenden Einheiten innerhalb einer Branche eine vergleichbare Spenderbasis zugrunde gelegt werden. Je näher die zu schätzenden Einheiten an den Meldeschwellen liegen, umso größer ist einerseits ihr Gewicht (im Sinne der “Bedeutung” der Einheit, nicht im Sinne eines Hochrechnungsgewichtes), aber umso ähnlicher sind sie andererseits auch den Strukturspendern in der Modellbasis. Mit wachsendem Abstand zur Meldeschwelle verringert sich zugleich auch das Gewicht der zu schätzenden Einheiten und somit deren Einfluss auf die Merkmalssummen eines Wirtschaftsbereichs. Dieser Eigenschaft der zu schätzenden Einheiten in Kombination mit der grundsätzlich hohen primärstatistischen Merkmalsabdeckung in der Mehrzahl der Wirtschaftsbereiche der Leistungs- und Strukturstatistik sollten den Erwartungen nach qualitativ hochwertige Ergebnisse hinsichtlich zeitlicher Vergleichbarkeit und Kohärenz mit anderen wirtschaftsstatistischen Projekten (insbesondere der Konjunk-

²Titel und Erläuterungen zur in Abbildung 4 verwendeten Codierung der ÖNACE 2008-Abteilungen sind verfügbar unter http://www.statistik.at/web_de/klassifikationen/.

turstatistik) entsprechen – auch unter der Annahme, dass innerhalb eines Wirtschaftsbereichs unternehmensgrößenabhängige Strukturunterschiede vorliegen.

Als problematisch muss bei der MDE jedoch das Vorliegen starker branchenspezifischer Strukturunterschiede gesehen werden, wenn gleichzeitig eine Konstellation aus geringer Primärbesetzung und schwacher Primärabdeckung auftritt. Zu unterscheiden ist dabei die Problematik der schwachen Primärabdeckung der Grundgesamtheit, bei der der Anteil der ergänzten Datenmasse an den Merkmalssummen eines Wirtschaftsbereiches hoch ist, von jener der geringen Primärbesetzung, bei der in der Erhebung zu wenig Strukturspender für eine branchenspezifische Modellbildung vorliegen. Wenn aufgrund zu geringer Primärbesetzung die Parameterschätzung zur modellbasierten Ergänzung einer ÖNACE 2008-Unterkategorie auf der übergeordneten ÖNACE 2008-Kategorie (die mehrere Unterkategorien aufgrund ihrer gleichartigen Wirtschaftstätigkeit zusammenfasst) erfolgen muss und die zugehörigen Unterkategorien beispielsweise ein deutlich höheres Lohnniveau aufweisen als die zu ergänzende, sind aufgrund “gemischter Strukturen” für die betroffene Branche auch mit der vorgenommenen Modellbasisauswahl deutlich verzerrte Schätzer zu erwarten. Ein gewisser systematischer Effekt kann in diesen Fällen bei der Datenergänzung kaum vermieden werden (siehe dazu auch Abschnitt 4.1.7).

4.1.4 Erstellung der Hauptmerkmale

Wirtschaftsstatistische Zusammenhänge von Mikrodaten können häufig schon durch ein einfaches lineares Modell gut beschrieben werden. Da wirtschaftsstatistische Daten jedoch stets mit Ausreißern behaftet sind – im vorliegenden Fall Beobachtungen, die nicht dem linearen Muster der Datenmehrheit folgen – ist der Einsatz einer robusten Methode zur Gewährleistung einer stabilen Qualität der Modellanpassung unerlässlich (siehe Huber, 1981; Hampel, 1995). Basierend auf der iterativ bestimmten Modellbasis werden bei der MDE deshalb alle Hauptmerkmale branchenspezifisch mit Hilfe eines robusten linearen Regressionsmodells berechnet.

In der Literatur werden verschiedenste robuste Techniken empfohlen. Zur Schätzung der Hauptmerkmale der Leistungs- und Strukturstatistik wird die Least Trimmed Squares Regression (LTS-Regression) angewandt. Dieses Verfahren ist höchst robust, hat aber gegenüber bspw. LMS-Schätzern den Vorteil einer wesentlich stabileren Zielfunktion, besserer Effizienz und Konvergenz (Rousseeuw and Van Driessen, 2006). Der bei Anwendung einer LTS-Technik deutlich höhere Rechenaufwand wird durch den Einsatz geeigneter Approximationsalgorithmen obsolet. Dies ist auch ein wesentlicher Grund, warum anderen robusten Regressionsmethoden wie etwa MM-Regression (siehe Yohai, Stahel and Zamar, 1991) nicht der Vorzug gegeben wurde.

Um den Einfluss additiver Effekte bei der Mikrodatenerstellung zu vermeiden, wird grundsätzlich ein lineares Modell ohne Intercept angewandt. Für das Modell $y_i = x_{i1}\beta_{j1} + \dots + x_{ip}\beta_{jp} + \epsilon_i$, mit ϵ_i als Störterm und $i \in E_j^\alpha$ ist die Zielfunktion in der LTS-Regression folgendermaßen definiert (Rousseeuw, 1984):

$$\hat{\beta}_{e_j^\alpha}^{LTS} = \arg \min_{\beta_j} \sum_{k=1}^{h_j} r_{(i)}^2. \quad (3)$$

Die Summe der h_j kleinsten quadratischen Residuen ist dabei zu minimieren, wobei

die quadrierten Residuen $r_i^2 = (y_i - \hat{y}_i)^2$ in geordneter Form wie folgt definiert sind:

$$r_{(1)}^2 \leq r_{(2)}^2 \leq \dots \leq r_{(e_j^\alpha)}^2. \tag{4}$$

Der Umfang h_j ist ein Subset aus den gegebenen e_j^α Beobachtungen der Branche j , wobei h_j definiert ist als

$$h_j \in \left[\frac{e_j^\alpha + 1}{2}, e_j^\alpha \right]. \tag{5}$$

Die unter (3) bis (5) formal beschriebene LTS-Schätzung kann vereinfacht als iteratives Verfahren erklärt werden, wie es in Abbildung 5 am Beispiel einer der 540 ÖNACE 2008-Unterklassen der Modellbasis demonstriert wird. Wird der Umfang des Subsets beispielsweise auf $h_j = (e_j^\alpha + 1)/2$ gesetzt, so werden wiederholt zufällig 21 Beobachtungen (dunkle Punkte) aus den $e_j^\alpha = 40$ Beobachtungen ausgewählt und basierend darauf die Least Squares (LS) Regression und die Summe der 21 kleinsten quadratischen Residuen berechnet. Ziel ist es, jene Teilmenge h_j aus den e_j^α Beobachtungen zu finden, bei der die Summe der h_j kleinsten quadratischen Residuen minimiert wird. Dem LTS-Schätzer entspricht dann die LS-Anpassung an diese h_j Punkte.

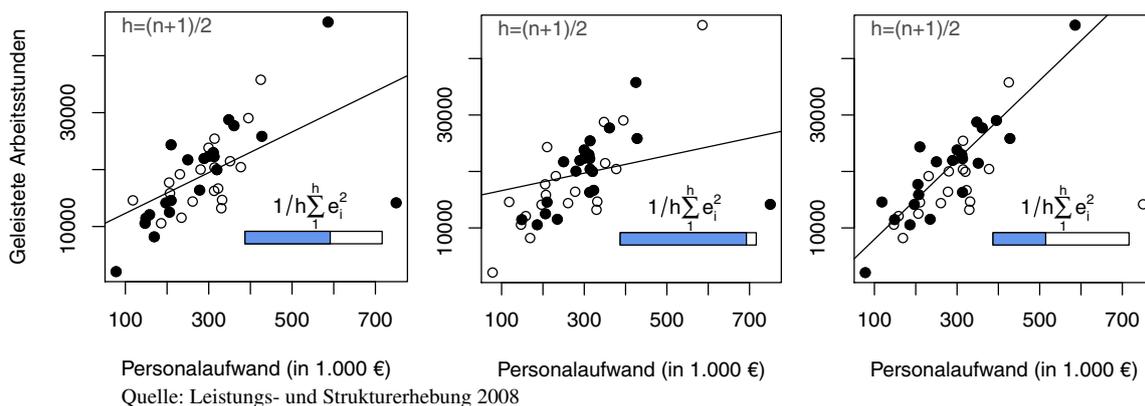


Abbildung 5: LTS-Technik – zufällig gewählte Subsets (Greedy Algorithmus).

Diese Technik ist extrem rechenintensiv, da theoretisch der “ganze Raum” abgesucht werden müsste, um das Minimum zu finden. Deshalb werden spezielle Algorithmen eingesetzt, die die Lösung für größere Datensätze approximieren (Rousseeuw and Van Driessen, 2006).

Mit der LTS-Methode werden stets $e_j^\alpha - h_j$ Beobachtungen ausgeschlossen, wodurch relevante Information verloren gehen kann. Mit wachsendem Subset h_j gewinnt der Schätzer zwar mehr Information aus den Daten, verliert aber an Robustheit und vice versa. Wird $h_j = 0.75e_j^\alpha$ gewählt, so gilt dies als guter Kompromiss zwischen Robustheit und Effizienz. Es ist jedoch kaum zu erwarten, dass in der Modellbasis alle Wirtschaftsbereiche und Merkmalszusammenhänge bis zu 25% ausreißerbehaftet sind. Andererseits ist die tatsächliche Ausreißeranzahl im Vorhinein nicht bekannt, d.h. auch nicht exakt über h_j bestimmbar. Deshalb wird für die MDE eine gewichtete LS-Regression angewandt, die in einem zusätzlichen Schritt – basierend auf dem besten LTS-Subset und der zugehörigen LTS-Funktion – nur noch Beobachtungen mit großen Residuen ausschließt. Basierend

auf dem neu gesetzten Subset h'_j wird der finale Schätzer berechnet (im Regelfall gilt: $h'_j > h_j$). Für die Beispieldaten in Abbildung 5 gilt $h'_j = 39$, d.h. alle bis auf eine der Beobachtungen werden in die finale Berechnung aufgenommen.

Mit Hilfe der so berechneten branchenspezifischen Regressionsparameter erfolgt, gebunden an die zuvor erstellten Eckdaten der 267.000 zu schätzenden statistischen Einheiten, die Berechnung der Hauptmerkmale wie *Waren- und Dienstleistungskäufe, Lagerbestand, Investitionen*, etc.

4.1.5 Erstellung der Detailmerkmale

Bei der Erstellung der Detailmerkmale der Leistungs- und Strukturstatistik wird ein Hauptmerkmal y nach D Anteilen aufgegliedert.

$$y = (y_1, \dots, y_D)^t, \quad y_v \geq 0, \quad v = 1, \dots, D, \quad \sum_{v=1}^D \frac{y_v}{y} = 1. \quad (6)$$

Es handelt sich folglich um Kompositionsdaten, für deren Analyse in der Literatur der log-ratio Ansatz empfohlen wird (siehe Hron, Templ and Filzmoser, 2010). Kompositionsdaten sind Daten, die als Teile eines Ganzen nur relative Informationen beinhalten, wodurch nur die Verhältnisse zwischen den Werten der Variablen relevant sind. Strukturelle Nullen stellen jedoch ein besonderes Problem bei der Verarbeitung von Kompositionsdaten dar (siehe Pawlowsky-Glahn and Egozcue, 2007), weshalb für die Leistungs- und Strukturstatistik keine entsprechende Transformation implementiert werden konnte, denn unterschiedliche Wirtschaftsstrukturen in Abhängigkeit von der wirtschaftlichen Aktivität bedingen in allen Wirtschaftsbereichen der Leistungs- und Strukturserhebung Nullwerte bei einzelnen Detailmerkmalen.

Da es sich bei Wirtschaftsdaten in der Regel um log-normalverteilte Daten handelt und im vorliegenden Fall - aufgrund systematisch fehlender Daten - nicht das Gewicht der statistischen Einheiten entsprechend ihrer Unternehmensgröße, sondern das Gewicht der kleinsten Einheiten, entsprechend ihrer höheren Anzahl, in die Berechnung eingehen soll, erfolgt die Aufgliederung der Hauptmerkmale – z.B. des *Lagerbestandes* in die Komponenten *Brenn- und Treibstoffe, Roh- und Hilfsstoffe, Handelswaren, unfertige Erzeugnisse, fertige Erzeugnisse* – über branchenspezifisch gemittelte Anteile:

$$\hat{y}_{i,v} = \frac{1}{e_j^\alpha} \sum_{s \in E_j^\alpha} \frac{y_{s,v}}{\sum_{r=1}^D y_{s,r}} \hat{y}_i. \quad (7)$$

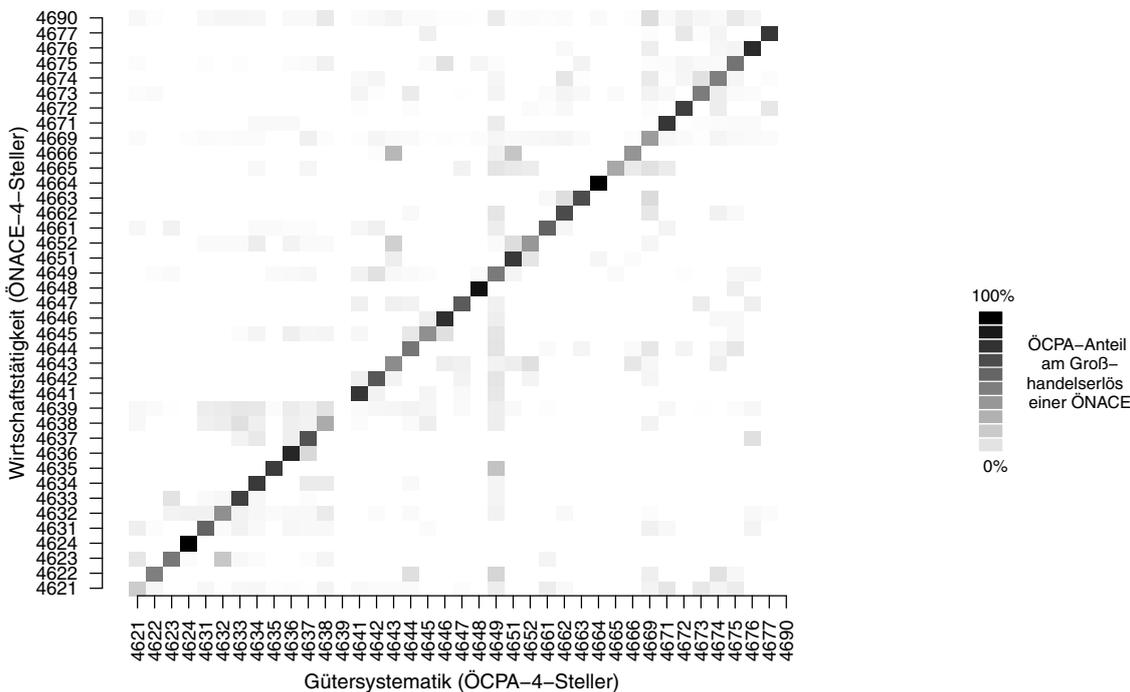
Darstellungsmerkmale (wie z.B. *Produktionswert* oder *Bruttowertschöpfung*) werden anschließend wie bei den primärstatistisch erhobenen Einheiten aus den Detailmerkmalen berechnet. Eine vollständige Auflistung aller Haupt-, Detail- und Darstellungsmerkmale ist der *Standarddokumentation der Leistungs- und Strukturstatistik* (2007) zu entnehmen.

4.1.6 Gütergliederung

Für die Ergänzung des primärstatistischen Datenkörpers ist in ausgewählten Handels- und Dienstleistungsbereichen auch die Aufgliederung der Umsatzerlöse nach Güterklassen,

also nach den vorgegebenen Tätigkeitsbereichen der Grundsystematik der Güter (siehe dazu *Grundsystematik der Güter, ÖCPA 2008, 2010*), erforderlich, um neben der Ergebnisdarstellung nach dem Aktivitätsansatz (der schwerpunktmäßigen wirtschaftlichen Aktivität eines Unternehmens) auch eine Ergebnisdarstellung nach dem Güteransatz (der Darstellung nach allen gleichartig klassifizierten Güter) zu ermöglichen.

Grundsätzlich ist eine plausible ÖCPA-Zuordnung über die wirtschaftliche Aktivität einer statistischen Einheit möglich. Die Matrixdarstellung nach der Systematik der Wirtschaftstätigkeiten, der ÖNACE 2008, und der statistischen Güterklassifikation in Verbindung mit den Wirtschaftszweigen, der ÖCPA 2008, weist generell eine hohe Konzentration des relevanten Umsatzvolumens an der Hauptachse auf.



Quelle: Leistungs- und Strukturhebung 2008

Abbildung 6: Vergleichsmatrix nach ÖNACE 2008-Klassen und ÖCPA-2008-Klassen für den Großhandel (ohne Handel mit Kraftfahrzeugen und Krafträdern).

Der Matrixplot in Abbildung 6 veranschaulicht die Gliederung der Erlöse aus Großhandel der Primärdaten nach Wirtschaftsbereichen und Gütergruppen. An der Intensität der Farbe ist die Konzentration des Großhandelsvolumens eines Wirtschaftsbereichs in den verschiedenen Gütergruppen ersichtlich. In Zeile sechs der Vergleichsmatrix ist beispielsweise die Konzentration der ≈ 1.4 Mrd. Euro Großhandelserlöse der ÖNACE 2008-Klasse 4632 (*Großhandel mit Fleisch und Fleischwaren*) veranschaulicht. Für die ÖNACE 2008-Klasse 4632 liegt die Konzentration des charakteristischen Erlösvolumens (ÖCPA-2008-Klasse 4632) bei rund 96%, die restlichen 4% entfallen auf nicht charakteristische Produkte (ÖCPA-2008-Klasse \neq 4632).

Für die zu ergänzenden statistischen Einheiten erfolgt die prozentuelle Aufgliederung der verschiedenen Umsatzarten eines Unternehmens auf die relevanten Gütercodes wiederum mittels Anteilsrechnung (siehe Abschnitt 4.1.5), wobei jedoch die Modellbasis-

auswahl für die Schätzung auf Güterebene ausschließlich auf Ebene der ÖNACE 2008-Unterklassen erfolgt. Bei nicht ausreichender primärstatistischer Besetzung wird beim Güteransatz nicht auf eine übergeordnete Wirtschaftsebene gewechselt, statt dessen werden innerhalb der betreffenden ÖNACE 2008-Unterkategorie schrittweise auch jene großen erhobenen Einheiten in die Modellbasis aufgenommen, die beim Aktivitätsansatz nicht berücksichtigt werden. Die Einbindung beschränkt sich dadurch im Wesentlichen auf Gütercodes, die dem Kriterium des “wirtschaftlichen Ursprungs” folgen (deren Codierung auf Ebene der 4–Steller mit der betreffenden ÖNACE 2008-Kategorie übereinstimmt). Da die Schätzung des Besetzungsgewichtes eines Gütercodes für die vorgesehenen statistischen Zwecke nicht unbedingt erforderlich ist und nur aggregierte Komponenten der verschiedenen Umsatzarten (*Großhandel, Einzelhandel, ...*) publiziert werden müssen, vereinfacht sich auch das Berechnungsmodell auf Güterebene sehr stark – im Gegensatz zum eher komplexen Modell, welches in der Konjunkturstatistik des Produzierenden Bereiches zum Einsatz kommt (mehr dazu in Dinges and Haitzmann, 2009).

4.1.7 Modellbedingte Effekte

Durch das Zusammenwirken verschiedener Abstrahierungen können modellbedingte Effekte auftreten, die jedoch bei ausreichender Primärabdeckung zumeist vernachlässigbar sind. Zu unterscheiden sind im Wesentlichen drei Haupteffekte:

- Fehler durch Modellbildung mittels z.B. linearer Regression (Modellierungsfehler)
- Fehler durch die Übertragung der basierend auf den Primärdaten gewonnenen Modellparameter, auf die nicht im Modell beobachtbaren kleineren statistischen Einheiten (systematischer Fehler aufgrund struktureller Unterschiede innerhalb einer Wirtschaftsunterklasse)
- Unschärfe durch die Berechnung der Modellparameter auf übergeordneten Wirtschaftsebenen bei zu geringer Primärbesetzung (systematischer Fehler aufgrund struktureller Unterschiede zwischen den Wirtschaftsunterklassen).

In primärstatistisch schwach besetzten, nicht homogenen Wirtschaftsteilbereichen ist folglich von stärkeren modellbedingten Effekten auszugehen, die nur durch ausreichende Primärabdeckung oder qualifiziertes Expertenrating³ abgefangen werden können.

Zur Quantifizierung der Effekte wäre die Durchführung einer Erhebung basierend auf einer kleinen Kontrollstichprobe notwendig, anhand derer die Differenzen der modellhaften Berechnung zur Realität abgeschätzt werden können. Die Realisierung einer solchen ex-Post Erhebung ist aufgrund der fehlenden gesetzlichen Grundlagen jedoch derzeit nicht absehbar.

4.2 Meldeausfälle

Etwa 5-6% der meldepflichtigen Unternehmen kommen im Rahmen der Leistungs- und Strukturhebung ihrer gesetzlichen Meldepflicht nicht nach und fallen bei der Erstel-

³Beim Expertenrating handelt es sich um Einschätzungen und Modifizierungen einzelner Werte und Parameter durch Fachexperten der Wirtschaftsstatistik aufgrund ihrer Kenntnisse branchenspezifischer Eigenheiten und Kriterien, die beim Vorliegen bekannter systematischer Abweichungsmechanismen vorgenommen werden.

lung der Grundgesamtheit in den zu ergänzenden Datenbereich. Grundsätzlich können auch Meldeausfälle mithilfe des in Abschnitt 4.1 beschriebenen Schätzmodells imputiert werden, insbesondere wenn es sich um Unternehmen nahe der Meldeschwelle handelt. Bei größeren Unternehmen kann jedoch davon ausgegangen werden, dass die Einbindung der historischen Struktur einer statistischen Einheit mittels Cold-Deck-Technik zu besseren Imputationsergebnissen führt als modellbasierte Verfahren oder die Anwendung von Hot-Deck-Imputation (siehe McKnight, McKnight, Sidani and Figueredo, 2007). Diese Annahme liegt darin begründet, dass in der Wirtschaftsstatistik mit ihrer Dominanz von Einzelfällen gerade bei größeren Unternehmen das Auffinden geeigneter Strukturspender aus der aktuellen Datenmatrix oftmals unmöglich ist (siehe auch Richter, 2002). Da in der Leistungs- und Strukturhebung für größere Meldeausfälle in der Regel Daten aus früheren Erhebungen vorliegen, kann die Mikrodatenerstellung für diese Unternehmen durch Fortschreibung der unternehmensspezifischen Vorjahresstruktur unter Einbindung der für das Berichtsjahr sekundärstatistisch verfügbaren Eckdaten aus Verwaltungsquellen erfolgen. Für den Produzierenden Bereich können darüber hinaus bestimmte Erlösbestandteile, die Beschäftigten, die Verdienste, die Arbeitsstunden sowie die Vollzeitäquivalente aufgrund der vorhandenen Synergien aus den Ergebnissen der Konjunkturstatistik im Produzierenden Bereich übernommen werden (zur Konjunkturerhebung siehe *Standarddokumentation zur Konjunkturstatistik im Produzierenden Bereich*, 2007).

5 Datenqualität

Die im vorliegenden Beitrag beschriebene Methode der Datenergänzung in der Leistungs- und Strukturstatistik illustriert die Vielzahl an Faktoren, welche bei der Beurteilung der Datenqualität zu berücksichtigen sind. Die Beachtung verschiedenster Qualitätsfaktoren ist in den konzeptionellen Vorgaben begründet, die keine konventionelle Hochrechnung zulassen. Dennoch wird *cut-off-sampling* bei wirtschaftsstatistischen Daten mit starker Merkmalskonzentration nicht nur als kosteneffiziente Methode zur Respondentenentlastung, sondern, wenn Regressorvariablen etwa aus Verwaltungsquellen verfügbar sind, oftmals auch als die genauere Methode betrachtet (siehe Knau, 2007).

Die Darstellung der Grundgesamtheit auf Mikroebene erfordert neben geeigneter mathematischer Verfahren auch das Einbeziehen von Informationen aus verschiedenen primär- und sekundärstatistischen Datenquellen. Neben der Schlüsselrolle, die dabei dem Unternehmensregister der STATISTIK AUSTRIA zukommt, sind es vor allem Verfügbarkeit und konzeptionelle Eignung der externen Verwaltungsquellen, die in die Gesamtqualität der Daten einfließen. Umfangreiche Voranalysen und Testberechnungen zeigten, dass die sekundärstatistisch ergänzten Eckdaten grundsätzlich als "voll erhoben" betrachtet werden können und auch regional detaillierte Auswertungen erlauben, mit qualitativ hochwertigen Ergebnissen. Die Güte der sonstigen beschäftigten- und umsatzbezogenen Merkmale hängt in erster Linie davon ab, inwieweit ausreichende primärstatistische Informationen zur Modellbildung vorliegen, aber auch davon, inwieweit Eckdaten und gewähltes Schätzmodell zur Erklärung eines Merkmals beitragen. Wie in Abschnitt 4.1.7 beschrieben, treten bei einer MDE durch das Zusammenwirken verschiedener Abstrahierungen modellbedingte Effekte auf. Modellierungsfehler können bei einzelnen Merk-

malen eines Wirtschaftsbereiches wie etwa bei der Schätzung der *Bruttoinvestitionen* größeres Gewicht haben. Systematische Fehler können sich in nicht homogenen Branchen, beispielsweise Branchen in denen die Leistungen auf Provisionsbasis erbracht werden, auf die Ergebnisse eines Wirtschaftsbereiches stärker auswirken. Überall dort, wo das eigentliche Ziel einer MDE umgesetzt werden kann – nämlich die modellbasierte Ergänzung eines vorwiegend primärstatistisch erfassten Datenkörpers – kann jedoch von hoher Datenqualität ausgegangen werden. Diese Annahme ist einerseits darin begründet, dass in diesem Fall für die Datenergänzung in der Regel ausreichend primärstatistische Informationen zur Modellentwicklung vorliegen, und andererseits, dass selbst in Wirtschaftsbereichen mit deutlicher Strukturabweichung mögliche Modelleffekte mit wachsender Primärabdeckung gegen Null tendieren.

Die MDE erbringt für die Mehrzahl von Hauptaggregaten auch ohne ressourcenintensive Arbeitsschritte zufriedenstellende Ergebnisse. In primärstatistisch schwach besetzten, nicht homogenen Wirtschaftsteilbereichen und in Bereichen mit deutlicher Strukturabweichung kann die Ergänzung jedoch nicht ausschließlich anhand automationsunterstützter Verfahren erfolgen. In Einzelfällen ist qualifiziertes Expertenrating unerlässlich.

6 Ausblick

Weitere Entwicklungen, wie die Berücksichtigung zusätzlicher sekundärstatistischer Quellen zur Verstärkung der Eckdaten und der an sie gebundenen Modellrechnung, werden im Sinne einer ständigen Qualitätsverbesserung angestrebt. Gegenwärtig verfügbare Verwaltungsquellen werden bei entsprechender Qualität allerdings schon weitestgehend genutzt.

Es sind zur Zeit jedoch methodische Arbeiten vorgesehen, um die Möglichkeit zu prüfen, dem Datennutzer zur Bewertung der Datengenauigkeit – trotz der im Sinne der Stichprobentheorie fehlenden statistischen Voraussetzungen – geeignete Kennzahlen oder Qualitätsindikatoren zur Verfügung zu stellen. Für 2011 ist die Publikation eines Fachartikels zur möglichen *Abschätzung der statistischen Unschärfe bei der modellbasierten Datenergänzung von Cut-Off-Surveys* geplant (Dinges and Kowarik, 2011).

Danksagung

Die Autoren danken *Johann Hameseder*, stv. Direktor der Direktion Unternehmen der STATISTIK AUSTRIA, für die tatkräftige Unterstützung bei der Projektumsetzung. Ein *Merci* für die gute Zusammenarbeit geht an *Christoph Gründler* und *Harald Murlasits* (Projektleitung LSE), *Leopold Milota* (Projektleitung KJP), *Martin Hirsch* und *Nina Waltner* (Analyse LSE) und *Elisabeth Gruber* (Projektleitung MEETS). Wir danken zwei anonymen Reviewern für “critical comments” und Hinweise zur Verbesserung des Manuskripts. Dank gebührt auch *Josef Richter*, Dozent an den Universitäten Linz und Innsbruck, der den Autoren in den letzten Jahren mehrfach als Berater in wirtschaftsstatistischen Fragen zur Seite stand.

Literatur

- Dinges, G. and Haitzmann, M. (2009). Modellbasierte Ergänzung der Konjunkturstatistik im Produzierenden Bereich. *Statistische Nachrichten*, 12, 1153-1166. Verfügbar unter http://www.statistik.at/web_de/downloads/methodik/kjp.pdf
- Dinges, G. and Kowarik, A. (2011). *Abschätzung der statistischen Unschärfe bei der modellbasierten Datenergänzung von Cut-Off-Surveys*. (Draft)
- Elisson, H. and Elvers, E. (2001). *Cut-off sampling and estimation. Proceedings of Statistics Canada Symposium 2001* (Bericht). Statistics Canada. Verfügbar unter <http://www.statcan.gc.ca/pub/11-522-x/2001001/session10/6259-eng.pdf>
- Grundsystematik der Güter, ÖCPA 2008*. (2010). Wien: Statistik Austria. Verfügbar unter http://www.statistik.at/web_en/publications_services/new_publications/index.html
- Hampel, F. (1995). *Wozu brauchen wir Robuste Statistik?* (Bericht Nr. 75). Swiss Federal Institute of Technology Zurich: ETH Zürich. Verfügbar unter http://stat.ethz.ch/research/publ_archive/1997
- Haslinger, A. (2004). Data Matching for the Maintenance of the Business Register of Statistics Austria. *Austrian Journal of Statistics*, 33, 55-67. Verfügbar unter <http://www.stat.tugraz.at/AJS/ausg041+2/041+2Haslinger.pdf>
- Hron, K., Templ, M. and Filzmoser, P. (2010). Imputation of missing values for compositional data using classical and robust methods. *Computational Statistics and Data Analysis*, 54, 3095-3107.
- Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.
- Knau, J. R. (2007). Cutoff Sampling and Inference. *InterStat*. Verfügbar unter <http://interstat.statjournals.net/YEAR/2007/abstracts/0704006.php>
- McKnight, P. E., McKnight, K. M., Sidani, S. and Figueredo, A. J. (2007). *Missing Data: A Gentle Introduction*. New York: The Guilford Press.
- Pawlowsky-Glahn, V. and Egozcue, J. J. (2007). *Statistische Analyse von Kompositionsdaten*. Freiberg. Verfügbar unter http://tu-freiberg.de/fakult3/geomageo/konferenz_2007_bht/Pawlowsky.pdf (58. Berg- und Hüttenmännischer Tag 2007 — Tagungsband Geologische Modellierung)
- Richter, J. (2002). *Kategorien und Grenzen der empirischen Verankerung der Wirtschaftsforschung*. Stuttgart: Lucius & Lucius.
- Rousseeuw, P. J. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association*, 79, 871-880.
- Rousseeuw, P. J. and Van Driessen, K. (2006). Computing LTS Regression for Large Data Sets. *Data Mining and Knowledge Discovery*, 12, 29-45.
- Standarddokumentation zur Konjunkturstatistik im Produzierenden Bereich*. (2007). Wien: Statistik Austria. Verfügbar unter http://www.statistik.at/web_de/wcmsprod/groups/gd/documents/stdok/022248.pdf
- Standarddokumentation zur Leistungs- und Strukturstatistik*. (2007a). Wien: Statistik Austria. Verfügbar unter http://www.statistik.at/web_de/dokumentationen/HandelDienstleistungen/index.html (Handel und Dienstleistungen)

- Standarddokumentation zur Leistungs- und Strukturstatistik.* (2007b). Wien: Statistik Austria. Verfügbar unter http://www.statistik.at/web_de/dokumentationen/ProduktionundBauwesen/index.html (Produktion und Bauwesen)
- Yohai, V., Stahel, W. A. and Zamar, R. H. (1991). *A Procedure for Robust Estimation and Inference in Linear Regression.* New York: Springer-Verlag.
- Zeller, M. and Karner, T. (2009). Systematik der Wirtschaftstätigkeiten - ÖNACE 2008. *Statistische Nachrichten*, 2, 177-183.

Adresse der Autoren:

Gerlinde Dinges
Fachbereich Methodik
Abt. Register, Klassifikationen u. Methodik
E-Mail: gerlinde.dinges@statistik.gv.at

Martin Haitzmann
Produzierender Bereich
Direktion Unternehmen
E-Mail: martin.haitzmann@statistik.gv.at

Sabine Zach
Dienstleistungen
Direktion Unternehmen
E-Mail: sabine.zach@statistik.gv.at

Statistik Austria
Guglgasse 13
A-1110 Wien
<http://www.statistik.at/>