

## **2   Synthese von Wirtschaftsstrukturen mit r88-Structurama**

**Timotheus Klein, Sven Altenburg**

Grundlage jeder Modellierung ist eine Datenbasis, die regionale Strukturdaten möglichst realitätsnah und in ihrer Schärfe der jeweiligen Fragestellung angemessen repräsentieren kann. Der Aufbau einer derartigen Datenbasis für den Wirtschaftsverkehr weist vielfältige methodische und finanzielle Hürden auf. Daher ist die Suche nach synthetischen Verfahren, die den Datenbedarf möglichst weit minimieren, ein interessantes Forschungsfeld. Das Programm „r88-Structurama“ stellt einen derartigen potentiellen Ansatz dar. Der Beitrag stellt das Experiment vor, eine von der TUHH „herkömmlich“ erarbeitete Wirtschaftsstruktur der Metropolregion Hamburg exemplarisch für zwei ländliche Kreise im Süden der Region synthetisch zu reproduzieren. Dabei soll geprüft werden, ob die Wirtschaftsstrukturen mit r88-Structurama realitätsnah abgebildet werden können. Zusätzlich wird dargestellt, welche Vor- und Nachteile das Verfahren gegenüber komplexen Datenbasen hat und wo seine Grenzen liegen. Darüber hinaus werden erste Empfehlungen für die praktische Anwendung gegeben.

## 2 Synthese von Wirtschaftsstrukturen mit r88-Structurama

**Timotheus Klein, Sven Altenburg**

### 2.1 Veranlassung und Zielsetzung

Eine räumlich und inhaltlich angemessene Datenbasis ist unerlässliche Voraussetzung für eine effiziente und realistische Modellierung von Verkehrsnachfrage. Während für die Modellierung des Personenverkehrs nach jahrzehntelanger Forschung und Praxis mittlerweile ein für die meisten Planungsaufgaben hinreichender Methodenbaukasten zur Verfügung steht, wird bei der erklärenden Modellierung des Wirtschaftsverkehrs vielfach im Trüben gefischt: Eine ungleich höhere Vielfalt sowohl bei Transportvorgängen und -gegenständen wie auch bei den Strukturen, die die Nachfrage hervorrufen, fällt mit einer teilweise unterschätzten Bedeutung dieser Verkehre für kommunale Planungsaufgaben zusammen.

Umso wichtiger ist es, effiziente Methoden zu entwickeln, mit denen Transportvorgänge oder verkehrsrelevante Strukturen möglichst umfassend und realitätsnah abgebildet werden können. Da der Aufbau einer adäquaten Datenbasis für den Wirtschaftsverkehr, die räumliche Verteilung sowie Größe und Anzahl von Betrieben möglichst differenzierter Branchen beinhaltet, ist dieser oft extrem aufwändig und weist vielfältige methodische Hürden auf. Beispielhaft sei die schwierige Datenlage genannt. So existieren zu den Arbeitsplätzen zwar umfangreiche amtliche Statistiken, in der Regel beinhalten diese jedoch nur sozialversicherungspflichtig Beschäftigte und lassen aus Datenschutzgründen nur sehr begrenzte Königkeiten zu. Andere Datenquellen auf Betriebsebene (IHK, Hoppenstedt, kommerzielle Geodatenanbieter u.a.) hingegen können je nach Größe des Untersuchungsgebietes sehr kostspielig und oftmals nur eingeschränkt methodisch kombinierbar sein.

Daher ist die Suche nach weitgehend synthetischen Verfahren mit minimalem Datenbedarf ein interessantes Forschungsfeld. Die den Autoren bekannten Verfahren zur Generierung synthetischer Wirtschafts- bzw. Bevölkerungsstrukturen greifen bislang zumeist auf konkrete Stichproben zurück und extrapolieren diese anhand aggregierter Parameter (vgl. z. B. Müller et al. 2010).

Das in diesem Beitrag vorgestellte Verfahren hingegen greift ausschließlich auf aggregierte Parameter zurück. Damit könnte es die Erzeugung von Wirtschaftsstrukturen nachrangiger Bedeutung erleichtern oder die Aufwandsschwelle zur Anwendung mikroskopischer Nachfragemodelle senken. Ziel dieses Versuchs war, die praktische Eignung eines solchen Ansatzes zu untersuchen und ggf. zu verbessern, sowie die Frage zu beantworten, ob dieses im Programm r88-Structurama implementierte Syntheseverfahren grundsätzlich für die Erzeugung von Wirtschaftsstrukturen geeignet und was bei der praktischen Anwendung zu beachten ist.

Im Folgenden wird die exemplarische Anwendung des Verfahrens vorgestellt. Dabei werden zunächst die grobe Struktur der Bearbeitung, die Datenbasis sowie das im Programm r88-Structurama implementierte Syntheseverfahren vorgestellt. Im Anschluss daran wird auf die Ergebnisse der einzelnen Bearbeitungsschritte und die daraus gewonnenen Erkenntnisse eingegangen.

## **2.2 Methode**

Bei der exemplarischen Anwendung des Verfahrens wurde versucht, Teile einer am Institut für Verkehrsplanung und Logistik der TUHH auf Basis zahlreicher Realdaten erzeugten synthetischen Wirtschaftsstruktur mit Hilfe daraus abgeleiteter Parameter zu rekonstruieren. Der Bearbeitungsablauf gliederte sich wie folgt:

- Ableitung statistischer Parameter und aggregierter Strukturdaten aus der synthetischen Wirtschaftsstruktur der TUHH, die in Form von Punktdaten für einzelne Betriebe auf Straßenabschnittsebene vorliegt
- Vollsynthetische Reproduktion der Wirtschaftsstruktur mit r88-Structurama auf Basis gebietsbezogener Daten (Anzahl Arbeitsplätze, Verteilungen)
- Vergleich der erzeugten mit der ursprünglichen Wirtschaftsstruktur

Dieser Bearbeitungsablauf wurde in 12 Varianten exerziert, die sich hinsichtlich der verwendeten Basisdaten, der Anzahl der Betriebsgrößenklassen und der angestrebten räumlichen Auflösung der Wirtschaftsstruktur unterscheiden. Zwei Varianten wurden doppelt durchgerechnet, um die vom Programm erzeugten Abweichungen bei ansonsten gleicher Datenbasis zu kontrollieren.

Tabelle 2.1 zeigt die untersuchten Synthesevarianten und ihre Eigenschaften:

**Tabelle 2.1** Synthesevarianten

Variante	Räumliche Auflösung	Synthesebasis	Betriebs- größen- klassen
1	Gemeinde	Arbeitsplätze	3
2 (Kontrolle)	Gemeinde	Arbeitsplätze	3
3	Gemeinde	Arbeitsplätze nach Branche	3
4	Aggregate mit >500 Arbeitsplätzen	Arbeitsplätze	3
5	Aggregate mit >500 Arbeitsplätzen	Arbeitsplätze nach Branche	3
6	Aggregate mit >1.000 Arbeitsplätzen	Arbeitsplätze	3
7	Aggregate mit >1.000 Arbeitsplätzen	Arbeitsplätze nach Branche	3
8	Gemeinde	Arbeitsplätze	6
9 (Kontrolle)	Gemeinde	Arbeitsplätze	6
10	Gemeinde	Arbeitsplätze nach Branche	6
11	Aggregate mit >500 Arbeitsplätzen	Arbeitsplätze	6
12	Aggregate mit >500 Arbeitsplätzen	Arbeitsplätze nach Branche	6
13	Aggregate mit >1.000 Arbeitsplätzen	Arbeitsplätze	6
14	Aggregate mit >1.000 Arbeitsplätzen	Arbeitsplätze nach Branche	6

Die in diesen Varianten erzeugten, vollsynthetischen Wirtschaftsstrukturen wurden unter folgenden Gesichtspunkten mit der TUHH-Datenbank verglichen:

- Die Zuordnung der Arbeitsplätze zu einzelnen Branchen
- Die Zuordnung der Betriebe zu Branchen und Größenklassen
- Die absolute Anzahl der erzeugten Betriebe

Bei den Verteilungen (Branchen und Größenklassen) wurden die Schnittmengen (Koinzidenzverhältnisse, siehe Formel 2.1) der statistischen Vorgaben mit den in der einzelnen Gemeinde bzw. im einzelnen Aggregat vollsynthetisch erzeugten Verteilungen geprüft.

$$KV = \sum_k RH_{min,k} [-] \quad (2.1)$$

mit: KV: Koinzidenzverhältnis

$RH_{min,k}$ : kleinerer Wert der relativen Häufigkeit beider Klassen k

Die Schnittmengen wurden für jede Gemeinde bzw. jedes Aggregat einzeln berechnet. Die Auswertung weist die Minimal- und Maximalwerte aus. Da die erreichten Werte auch von der Anzahl der Klassen abhängen, wurde ergänzend ein so genannter „Verteilungsfehler“ definiert und berechnet, der die mittlere Abweichung, in Prozentpunkten, vom vorgegebenen Anteil einer Klasse ausweist:

$$V = (1 - KV)/k [-] \quad (2.2)$$

mit: V: Verteilungsfehler

Ein Verteilungsfehler von 0,01 besagt, dass der Anteil der einzelnen Klassen einer untersuchten Verteilung im Mittel um 1 Prozentpunkt von der Vorgabe abweicht. Für die Beurteilung der Ergebnisse wurde ein Verteilungsfehler von 0,01 als Akzeptanzschwelle gewählt.

In einer vertiefenden Einzelfallbetrachtung wurde außerdem überprüft, inwieweit die statistischen Vorgaben für einzelne Gemeinden bzw. Aggregate repräsentativ sind.

## 2.3 Datenbasis

### 2.3.1 Vorliegende Wirtschaftsstrukturdaten

Bei der vorliegenden, mit r88-Structurama zu rekonstruierenden Wirtschaftsstruktur handelt es sich um einen Auszug aus einer am Institut für Verkehrssysteme und Logistik der TUHH generierten Datenbank, in der knapp 300.000 einzelne Betriebe der Metropolregion Hamburg als Punktdaten mit verschiedenen Eigenschaften hinterlegt sind.

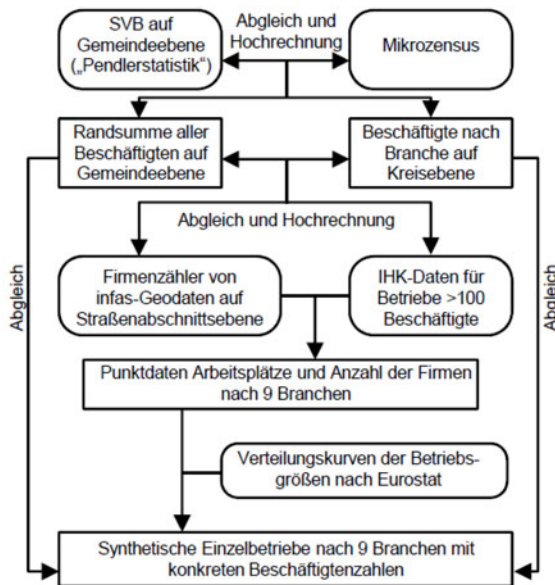


Abb. 2.1 Erstellung TUHH-Datenbank

Diese Datenbank der TUHH wurde eher „klassisch“ unter kombinierter Verwendung amtlicher Statistiken (Bundesagentur für Arbeit, Mikrozensus, Betriebshistorien-Panel, Eurostat), ergänzender Unternehmensdaten (IHK, Hoppstedt-Datenbank) und kommerzieller Geodaten (infas Geodaten) erstellt und erfordert ein vergleichsweise hohes Maß an finanziellen Mitteln, Arbeitseinsatz und methodischem Knowhow. Im Wesentlichen besteht die Erzeugung der Datenbasis aus den in Abbildung 2.1 dargestellten Schritten. Eine umfangreiche Schilderung der einzelnen Schritte findet sich im Arbeitspapier zum BMBF-geförderten Projekt „ELAN – Energiepreisentwicklung und Landnutzung“, das aktuell am Institut für Verkehrsplanung und Logistik der TUHH erstellt wird (Gertz et al, Veröffentlichung in Vorbereitung).

Aus der Wirtschaftsstrukturdatenbank der TUHH wurden Auszüge für die Landkreise Uelzen und Lüchow-Dannenberg verwendet, bei denen aufgrund der geringen Siedlungsdichte die größten Probleme mit der Rekonstruktion der synthetischen Wirtschaftsstruktur erwartet wurden. Beide Kreise sind dem BBR-Kreistyp „ländliche Räume geringer Dichte“ zugeordnet. Für die exemplarische Vollsynthese mit r88-Structurama wurden die Branchenzugehörigkeit und die Betriebsgröße als zu reproduzierende Attribute ausgewählt.

### 2.3.2 Sonstige Daten

Unerlässliche Voraussetzung für die Anwendung des Programms ist eine DBF-Datei, die Angaben zur Menge der in den gegebenen Raumeinheiten zu erzeugenden Arbeitsplätze enthält. Hierfür wurde die Datei „gemeinden-umringe“ aus dem WebFeatureService des Landes Niedersachsen (LGLN 2012) mit den Betriebsdaten der TUHH verschnitten. Die Ergebnisdatei enthält die Arbeitsplatzzahlen nach Branche und in Summe. Daraus wurden manuell Aggregate mit mindestens 500 bzw. 1.000 Arbeitsplätzen abgeleitet. Dabei wurde von der Gemeinde mit der niedrigsten Arbeitsplatzzahl ausgegangen und diese mit derjenigen benachbarten Gemeinde vereint, die die wenigsten Arbeitsplätze hat. Je nach Fragestellung können andere Methoden sinnvoller sein (Vereinigung von Gemeinden ähnlicher Arbeitsplatzdichte usw.).

## 2.4 Das Programm r88-Structurama

r88-Structurama wurde ursprünglich zur Erzeugung synthetischer Bevölkerungen auf Grundlage aggregierter, georeferenzierter Bevölkerungsdaten und allgemeiner, statistischer Angaben zur Verteilung von Attributen dieser Bevölkerung entwickelt. Die Bedienung erfolgt mit einem Graphical User Interface und erfordert keine besonderen Programmierkenntnisse. Die Programmiersprache ist Python 2.5.4, und das Programm steht im GitHub-Account von Timotheus Klein<sup>7</sup> mit GPL 3.0-Lizenz zur freien Verwendung zur Verfügung.

---

<sup>7</sup> <https://github.com/TimotheusKlein>

Das Programm erzeugt Tabellen in einer SQLite-Datenbank, deren Zeilen einzelne Personen, Arbeitsplätze oder anderes repräsentieren, und in deren Spalten beliebige, zum Teil voneinander abhängende Attribute belegt werden können. SQLite-Datenbanken wiederum können über den in Firefox eingebauten SQLite-Manager begutachtet und anderen Anwendungen zugänglich gemacht werden.

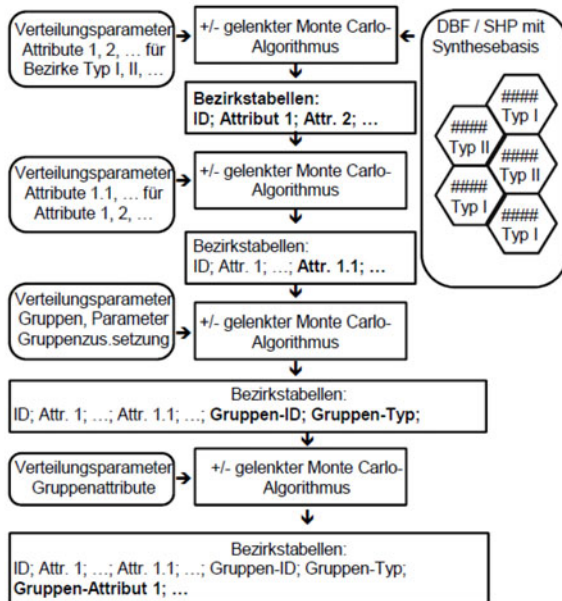
In erster Linie hängen die Attribute bzw. die Anzahl der in der Tabelle angelegten Zeilen von aggregierten Strukturdaten ab, die aus einer \*.dbf-Datei eingelesen werden. Damit stehen alle Daten zur Verfügung, die durch Shapefiles hinterlegt sind. Zu jeder dort verzeichneten Mengenangabe können stetig oder diskret verteilte Attribute und ihre Ausprägungen definiert werden. Das Programm erzeugt eine entsprechende Anzahl Einträge, deren Attribute unter Berücksichtigung der vorgegebenen Verteilung wahlweise mit einem Monte Carlo-Algorithmus oder, bei stetig verteilten Attributen, durch Interpolation belegt sind. Im Beispiel dieser Untersuchung handelt es sich bei der Mengenangabe um Arbeitsplätze und beim Attribut um die Branchenzugehörigkeit. Die in diesem ersten Schritt belegten Attribute werden im Programm als „Attribute (mit Abhängigkeit) 1. Ordnung“ bzw. „Dependency 1st Order“ bezeichnet.

Attribute 2. Ordnung sind dieser Terminologie entsprechend Attribute, deren Wert von den Attributen 1. Ordnung abhängt. Im Fall der branchenspezifischen Arbeitsplätze (1. Ordnung) bestünde die Möglichkeit, als Attribute 2. Ordnung Bildungsgrade, Einkommen usw. zuzuordnen.

Als Attribut 3. Ordnung kann die Zugehörigkeit der einzelnen Individuen zu unterschiedlich definierten Gruppen festgelegt werden. Dabei wird neben der Verteilung von Gruppentypen der Aufbau der Gruppen vorgegeben, so können z.B. verschiedene Typen von Arbeitsplätzen in einem Betrieb kombiniert werden.

Bei der Zuordnung der Individuen zu den Gruppen erfolgt vor jeder neu zu bildenden Gruppe eine Gegenüberstellung der Typenverteilung bereits gebildeter Gruppen mit der vorgegebenen Typenverteilung. Daraus ergibt sich eine Priorisierung der Gruppentypen für die nächste zu bildende Gruppe. Kann der Typ mit der höchsten Priorität mangels passender Individuen bzw. Arbeitsplätze nicht mehr gebildet werden, wird versucht, den Typ mit der nächstniedrigeren Priorität zu bilden. Kann überhaupt keine Gruppe mehr gebildet werden, werden die übrig gebliebenen Individuen aus der Datenbank entfernt und der Gesamtumfang der Bevölkerung bzw. der Arbeitsplätze oder sonstiger Randsummenbedingungen durch Kopieren entsprechender vorhandener Gruppen wieder hergestellt. Dabei kann es zu einer Verzerrung der zuvor reproduzierten Verteilungen kommen.

Abschließend besteht mit den Attributen 4. Ordnung die Möglichkeit, den zuvor generierten Gruppen ihrerseits Attribute zuzuordnen. Denkbar wären u.a. Attribute zum Fahrzeugbestand, jedoch wurde im Rahmen dieser Untersuchung darauf verzichtet. Die Abbildung 2.2 zeigt den konkreten Ablauf der Synthese mit r88-Structurama.

**Abb. 2.2** Ablauf Synthese mit r88-Structurama

## 2.5 Ergebnisse

### 2.5.1 Statistische Parameter

Die statistischen Parameter wurden auf Grundlage der Betriebs-Punktdaten in den Landkreisen Uelzen und Lüchow-Dannenberg berechnet. Die Arbeitsplätze verteilen sich dort zu den in Tabelle 2.2 dargestellten Prozentsätzen auf die unterschiedlichen Branchen und werden in den entsprechenden Synthes Szenarien unmittelbar für die Zuordnung der „Attribute 1. Ordnung“ verwendet. Die prozentuale Belegung der Betriebsgrößenklassen in den einzelnen Branchen bei Verwendung von drei standardisierten Betriebsgrößenklassen zeigt Tabelle 2.3.

**Tabelle 2.2** Prozentuale Anteile der Branchen an Arbeitsplätzen in den Landkreisen Uelzen und Lüchow-Dannenberg

Bau	Oef	DiLei	Fnz	Gastro	Handl	Ind	MeSoz	Land
7%	15%	16%	3%	5%	16%	14%	20%	4%



**Tabelle 2.3** Prozentuale Anteile der Betriebsgrößen in unterschiedlichen Branchen in den Landkreisen Uelzen und Lüchow-Dannenberg. Bezug Branche/ Betriebe gesamt

Größe	Bau	Oef	DiLei	Fnz	Gastro	Handl	Ind	MeSoz	Land
Groß	0,3%	0,6%	0,5%	0,9%	0,2%	0,2%	6,3%	0,6%	0,0%
101-500	0,02%	0,07%	0,11%	0,04%	0,01%	0,05%	0,17%	0,04%	0,00%
Mittel	10,7%	26,9%	14,6%	18,2%	10,5%	13,2%	36,6%	7,8%	22,6%
11-100	1,03%	3,58%	3,37%	0,75%	0,54%	2,76%	1,01%	0,48%	3,38%
Klein	89,0%	72,6%	84,9%	81,0%	89,3%	86,5%	57,1%	91,5%	77,4%
1-10	8,57%	9,66%	19,57%	3,35%	4,62%	18,03%	1,58%	5,60%	11,61%

**Tabelle 2.4** Prozentuale Anteile von sechs Betriebsgrößen unterschiedlicher Branchen im BBR-Kreistyp „ländliche Kreise geringerer Dichte“, optimierte Klassengrenzen (Anzahl Arbeitsplätze, Prozent Bezug Betriebe gesamt)

Größe	Bau	Oef	Dilei	Fnz	Gastro	Handl	Ind	MeSoz	Land
Groß (1)	24-117	63-177	126-315	28-88	61-208	72-163	299-668	386-1571	10-20
„XXL“	1,649%	2,535%	8,972%	2,043%	0,640%	4,652%	0,640%	2,658%	11,200%
Groß (2)	10-23	23-62	20-125	15-27	18-60	22-71	122-298	124-385	6-9
„XL“	2,991%	2,745%	7,680%	0,591%	1,120%	4,431%	0,960%	2,818%	1,022%
Mittel (1)	4-9	10-22	9-19	7-14	6-17	9-21	50-121	65-123	4-5
„L“	1,415%	4,898%	3,422%	0,714%	2,400%	8,849%	0,615%	0,234%	0,382%
Mittel (2)	3-3	3-9	3-8	3-6	3-5	3-8	18-49	27-64	3-3
„M“	2,597%	1,711%	1,649%	0,320%	0,591%	1,723%	0,418%	0,320%	0,751%
Klein (1)	2-2	2-2	2-2	2-2	2-2	2-2	6-17	11-26	2-2
„S“	0,554%	1,354%	1,268%	0,382%	0,406%	1,145%	0,098%	0,074%	1,169%
Klein (2)	1-1	1-1	1-1	1-1	1-1	1-1	1-5	1-10	1-1
„XS“	0,418%	0,074%	0,062%	0,086%	0,012%	0,037%	0,025%	0,012%	0,468%

Für die Bestimmung von sechs Größenklassen wurde aus der synthetischen Wirtschaftsstruktur der TUHH ein neuer Datensatz erzeugt, der die einzelnen Betriebe mit ihren Beschäftigtenzahlen ausweist. Die Einteilung der Größenklassen wurde mit Hilfe eines k-means-Algorithmus durchgeführt (Neilson 2012). Tabelle 2.4 zeigt diese Größenklassen und ihren Anteil an den Betrieben in den beiden betrachteten Landkreisen. Auffällig sind die hohe Zahl sehr kleiner Betriebe und die insgesamt ausgewogeneren Klassengrößen.

## 2.5.2 Vollsynthetische Wirtschaftsstruktur

### Schnittmengen

Die erreichten Schnittmengen sind in Tabelle 2.5 zusammengestellt. Bei Verwendung der Arbeitsplatzzahlen nach Branche werden die Arbeitsplätze einer Verteilung mit nur einer Klasse zugeordnet, sodass im Einzelfall Schnittmengen von 0 (keine Arbeitsplätze in der Branche, Division durch 0) oder 1 berechnet werden.

**Tabelle 2.5** Schnittmengen (AP – Arbeitsplätze, APB – Arbeitsplätze nach Branche)

	Gemeinde, AP	Gemeinde, AP (Kontrolle)	Gemeinde, APB	Aggregat >500 AP	Aggregat >500 APB	Aggregat >1.000 AP	Aggregat >1.000 APB
Branchenzuordnung							
Minimum	0,75	0,74	0,00	0,95	1,00	0,96	1,00
Mittel	0,92	0,91	0,88	0,97	1,00	0,98	1,00
Maximum	0,99	0,99	1,00	0,99	1,00	0,99	1,00
Betriebstyp- / -größenzuordnung							
Minimum	0,34	0,34	0,35	0,67	0,68	0,75	0,71
Mittel	0,66	0,65	0,63	0,78	0,77	0,81	0,80
Maximum	0,90	0,86	0,85	0,88	0,84	0,88	0,91

**Tabelle 2.6** Verteilungsfehler. Minimum, Maximum bezogen auf die zugrunde liegenden Schnittmengen (AP – Arbeitsplätze, APB – Arbeitsplätze nach Branche)

	Gemeinde, AP	Gemeinde, AP (Kontrolle)	Gemeinde, APB	Aggregat >500 AP	Aggregat >500 APB	Aggregat >1.000 AP	Aggregat >1.000 APB
Branchenzuordnung							
Minimum	0,028	0,029	0,111	0,006	0,000	0,005	0,000
<b>Mittel</b>	<b>0,009</b>	<b>0,010</b>	<b>0,014</b>	<b>0,003</b>	<b>0,000</b>	<b>0,003</b>	<b>0,000</b>
Maximum	0,001	0,001	0,000	0,001	0,000	0,001	0,000
Betriebstyp- / -größenzuordnung							
Minimum	0,025	0,024	0,024	0,012	0,012	0,009	0,011
<b>Mittel</b>	<b>0,013</b>	<b>0,013</b>	<b>0,014</b>	<b>0,008</b>	<b>0,008</b>	<b>0,007</b>	<b>0,008</b>
Maximum	0,004	0,005	0,006	0,005	0,006	0,005	0,003
Akzeptabel:	Nein	Nein	Nein	Ja	Ja	Ja	Ja

Die Schwankungen zwischen zwei Synthesedurchläufen sind moderat, mit der Größe der Aggregate steigen die erreichten Schnittmengen. Die mit einem Verteilungsfehler von maximal 1 Prozentpunkt akzeptablen Ergebnisdatenbanken sind in

Wirtschaftsverkehr 2013

Datenerfassung und verkehrsträgerübergreifende

Modellierung des Güterverkehrs als

Entscheidungsgrundlage für die Verkehrspolitik

Clausen, U.; Thaller, C. (Hrsg.)

2013, XII, 176 S. 60 Abb., Hardcover

ISBN: 978-3-642-37600-9