

Statistische Modellierung auf der Grundlage von Modellergebnissen

Inhalt

- 1. Einleitung**
- 2. Modelltypen der Datengenerierung**
- 3. Konsequenzen**
- 4. Schlussbemerkungen**

Einleitung

Bei der Anwendung von Methoden der Statistik wird davon ausgegangen, dass die vorliegenden Daten ihrem Wesen nach **Beobachtungen** sind, die noch dazu eine Reihe von erwünschten Eigenschaften aufweisen.

Diese Sicht ist stark von der Situation in den Naturwissenschaften geprägt.

In den Sozial- und insbesondere in den Wirtschaftswissenschaften sind die Datengrundlagen hingegen oft selbst schon das Ergebnis von Modellberechnungen unterschiedlichster Art.

In die Generierung der „Daten“ sind bereits zahlreiche Hypothesen eingeflossen.

Die Datengrundlage ist keineswegs homogen.

Es besteht die Gefahr des „modeling on the basis of models“ (HOLUB, TAPPEINER 1997).

Modelltypen der Datengenerierung

Typ 1 - Konsolidierungsmodelle

Typ 2 - Substitutionsmodelle

- Standardfall – Imputation
- Ausgleichsverfahren
- Prognosemodelle

Typ 3 - Redefinitionsmodelle

- Überschreitung der Beschränkungen des Geltungsbereichs von Prämissen - Einsatz von Stellvertretervariablen
- Datengenerierung aus dem Systemzusammenhang

Modelltypen der Datengenerierung

Typ 1 – Konsolidierungsmodelle

Das Ziel ist immer Verdichtung von Information mit Hilfe genau definierter Bildungsgesetze.

Die Ergebnisse dieser Modellklasse können bei Kenntnis der Regeln, etwa jener der Aggregation, gut nachvollzogen werden.

Die Wahl der Bildungsgesetze und ihrer Reihenfolge enthält stets eine a priori Wertung, ist theoriegeladen.

Zwischen alternativen Modellen des Typs 1 kann nicht auf der Grundlage empirischer Tests diskriminiert werden.

Für Konsolidierungsmodelle gilt: "Concepts matter".

Modelltypen der Datengenerierung

Typ 2 - Substitutionsmodelle

Information über der Beobachtung zugängliche Einzelelemente wird durch Information über andere Elemente ersetzt.

Das grundsätzlich beobachtbare Einzelelement B wird - durch die Unterstellung funktionaler Zusammenhänge zwischen beobachtbaren Variablen (A und B) - aus dem oder den tatsächlich beobachteten Elementen A1, A2, A3, usw. generiert.

Da sowohl A als auch B beobachtbar sind, ist eine empirische Überprüfung der unterstellten funktionalen Zusammenhänge, der verwendeten Parameter, grundsätzlich möglich.

Es werden Beziehungen zwischen Größen unterstellt, die Ziel der Erkenntnis wirtschaftswissenschaftlichen Forschens sein können und sind.

Modelltypen der Datengenerierung

Typ 3 - Redefinitionsmodelle

Modelle des Typs 3 erzeugen durch die Unterstellung funktionaler Zusammenhänge das grundsätzlich nicht beobachtbare Einzelelement C aus dem oder den tatsächlich beobachteten Elementen A1, A2, A3, B1, B2 usw.

Die unterstellten funktionalen Beziehungen können empirisch nicht überprüft werden.

Auf der Grundlage einer Konvention wird der Informationsgehalt der erklärenden Variable(n) zur abhängigen Variablen umdefiniert.

Modelle des Typs 3 versuchen, Barrieren, die aus der Natur der zu beobachtenden Realität resultieren, zu überwinden.

Exkurs: Einige Charakteristika der Wirtschaftsstatistik

Schiefe der Verteilungen

Inhomogenität vieler der elementaren Beobachtungseinheiten

Arbeitsteiliger Prozess der Datengenerierung:

- 1) Elementarbeobachtungen
- 2) Von den Elementarbeobachtungen zum einzelwirtschaftlichen Informationsvorrat
- 3) Vom Informationsvorrat zu statistischen Einzelbelegen für die Einzelunits
- 4) Von Mikrodatensätzen zu editierten Mikrodatensätzen
- 5) Von editierten Mikrodatensätzen zu statistischen Ergebnissen

Konsequenzen

Beispiel 1: Berechnung der Arbeitsproduktivität

Grundansatz:

$$\frac{pr_t}{pr_o} = \frac{\frac{\sum_i \frac{q_{ti}}{q_{oi}} \cdot w_{oi}}{\sum_i w_{oi}}}{b_t}}{b_o} \quad (1)$$

pr_t	Index der Beschäftigtenproduktivität Periode t
pr_o	Index der Beschäftigtenproduktivität Periode o
q_{ti}	Produzierte Menge des Gutes i in der Periode t
q_{oi}	Produzierte Menge des Gutes i in der Basisperiode o
w_{oi}	Indexgewicht des Gutes i in der Basisperiode o
b_t	Zahl der unselbständig Beschäftigten in der Periode t
b_o	Zahl der unselbständig Beschäftigten in der Basisperiode o

Konsequenzen

Beispiel 1: Berechnung der Arbeitsproduktivität

$$\frac{pr_t}{pr_o} = \frac{\sum_i^* \frac{q_{ti}^*}{q_{oi}^*} \cdot w_{oi}^* + \sum_i^{**} \frac{q_{ti}^{**}}{q_{oi}^{**}} \cdot w_{oi}^{**} + \sum_i^{***} \frac{q_{ti}^{***}}{q_{oi}^{***}} \cdot w_{oi}^{***}}{\sum_i^* w_{oi}^* + \sum_i^{**} w_{oi}^{**} + \sum_i^{***} w_{oi}^{***}} \quad (2)$$

$$\frac{q_{ti}^{**}}{q_{oi}^{**}} = \frac{\frac{o_{ti}^{**}}{o_{oi}^{**}}}{\frac{p_{ti}^{**}}{p_{oi}^{**}}} = \frac{o_{ti}^{**} \cdot p_{oi}^{**}}{o_{oi}^{**} \cdot p_{ti}^{**}} \quad (3)$$

q_{ti}^* Produzierte Menge des Gutes i in der Periode t

q_{oi}^* Produzierte Menge des Gutes i in der Basisperiode o

o_{ti}^{**} Wert des produzierte Gutes i in der Periode t

o_{oi}^{**} Wert des produzierte Gutes i in der Basisperiode o

p_{ti}^{**} Preis des Gutes i in der Periode t

p_{oi}^{**} Preis des Gutes i in der Basisperiode o

Konsequenzen

Beispiel 1: Berechnung der Arbeitsproduktivität

$$\frac{q_{ti}^{***}}{q_{oi}^{***}} = \frac{b_{ti}^{***}}{b_{oi}^{***}} \quad (4)$$

“Of substantial importance when using the labour input as the basic information for compiling the IPI is the change in productivity, in other words the ongoing changes in the volume of the labour factor input needed to achieve a fixed amount of output, caused for example by changes in technology and organisation.

Failure to take changes in productivity into account would lead to a misinterpretation of the IPI as regards production output, growth and income generation. **When using labour input as the basic information it is necessary to include this productivity development via appropriate factors.**”

(Methodology of short-term business statistics, Interpretation and guidelines EUROSTAT 2006)

Konsequenzen

Beispiel 1: Berechnung der Arbeitsproduktivität

$$\frac{b_{ti}^{***}}{b_{oi}^{***}} = \frac{\frac{O_{ti}^{***}}{O_{oi}^{***}}}{\frac{lp_{ti}^{***}}{lp_{oi}^{***}}} = \frac{O_{ti}^{***} \cdot lp_{oi}^{***}}{O_{oi}^{***} \cdot lp_{ti}^{***}} \quad (5)$$

$$\frac{q_{ti}^{***}}{q_{oi}^{***}} = \frac{\frac{O_{ti}^{***}}{O_{oi}^{***}}}{\frac{lp_{ti}^{***}}{lp_{oi}^{***}}} \quad (6) \qquad \frac{q_{ti}^{***}}{q_{oi}^{***}} = \frac{\frac{O_{ti}^{***}}{O_{oi}^{***}} \cdot \dot{p}r_i^{***}}{\frac{lp_{ti}^{***}}{lp_{oi}^{***}}} \quad (7)$$

lp_{ti}^{***} Preis des Faktors Arbeit für die Herstellung des Gutes i in der Periode t
 lp_{oi}^{***} Preis des Faktors Arbeit für die Herstellung des Gutes i in der
 Basisperiode o

Konsequenzen

Beispiel 1: Berechnung der Arbeitsproduktivität

Ein Spezialproblem stellt der **Nichtmarktbereich** dar, in dem viele Beschäftigte tätig sind, für den aber kein auf Transaktionen beruhender Output beobachtbar ist.

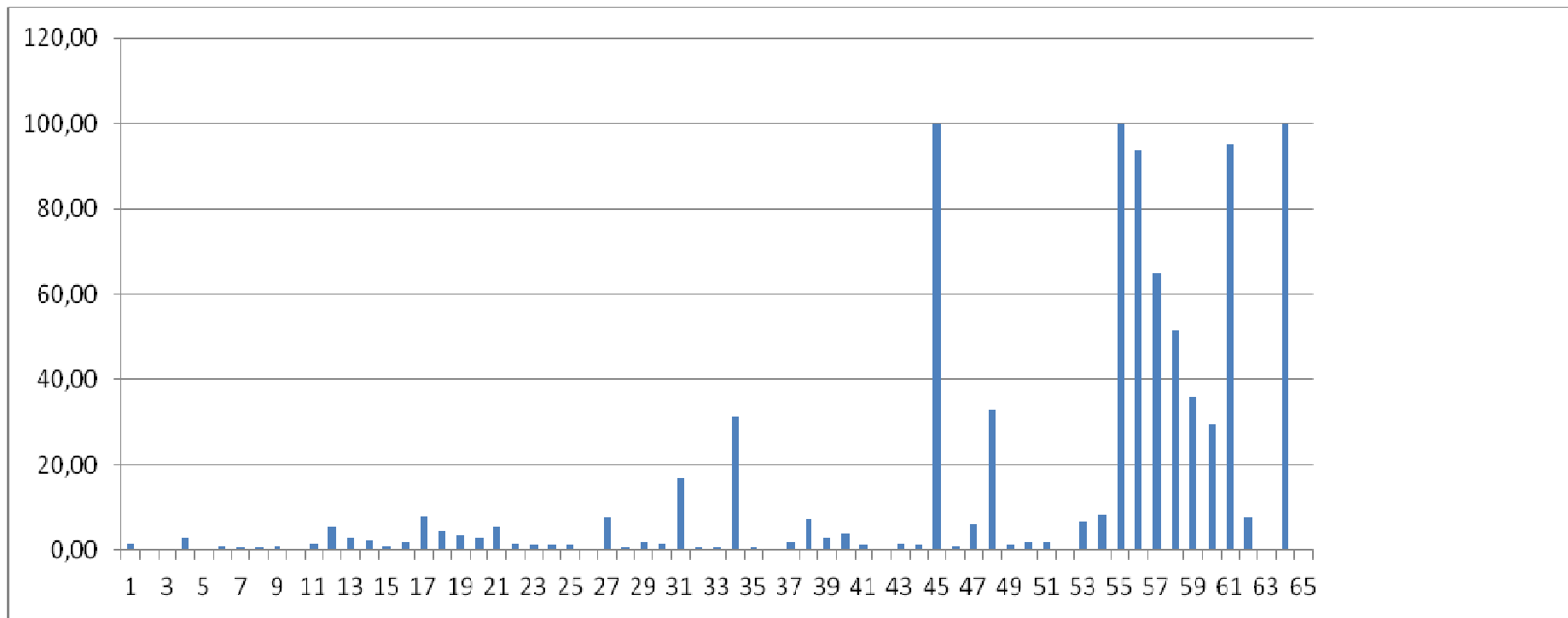
Kostenkonvention der VGR: $\text{Output} = \text{Summe der Kosten}$

Da Preise fehlen, ist die Rechnung zu konstanten Preisen in der üblichen Form nicht möglich.

Das ESVG 2010 regelt im Detail, wie die Berechnungen erfolgen sollen. Ein Beispiel: Im Bereich der nichtmarktbestimmten Gesundheits- und Bildungsleistungen etwa müssen die geschätzten Produktions- und Verbrauchsvolumen anhand direkter Outputmessungen — ohne Qualitätsanpassung — berechnet werden (ESVG 2010,10.30).

Homogenität des Datenkörpers?

Anteile der Nichtmarktproduktion an der Gesamtproduktion nach Wirtschaftsbereichen in % – Österreich 2011



Quelle: Statistik Austria; Aufkommens- und Verwendungstabelle 2011

Homogenität des Datenkörpers?

Anteile der Nichtmarktproduktion an der Gesamtproduktion für ausgewählte Wirtschaftsbereiche in % – Österreich 2011

Crop and animal production, hunting and related service activities	1,42
Forestry and logging	0,05
Fishing and aquaculture	0,00
Land transport and transport via pipelines	17,04
Travel agency, tour operator reservation service and related activities	6,73
Security and investigation activities; services to buildings and landscape activities;	8,27
Public administration and defence; compulsory social security	100,00
Education	93,84
Human health activities	65,08
Social work activities	51,39
Creative, arts and entertainment activities; libraries, archives, museums and other	35,79
Sports activities and amusement and recreation activities	29,37
Activities of membership organisations	95,19
Repair of computers and personal and household goods	7,72
Other personal service activities	0,22
Total	16,84

Quelle: Statistik Austria; Aufkommens- und Verwendungstabelle 2011

Konsequenzen

Beispiel 1: Berechnung der Arbeitsproduktivität

Je nach Aggregationsniveau ist der Anteil von Modellresultaten des Typs 3 unterschiedlich hoch.

Auch im Marktbereich sind viele Teilaggregate das Ergebnis von Modellberechnungen des Typs 2.

Je nach Aggregationsniveau spiegeln die Ergebnisse der Berechnung der Arbeitsproduktivität jene Annahmen wider, welche in die Generierung der Datengrundlage eingeflossen sind.

Konsequenzen

Beispiel 2: Untersuchung des Einflusses der Unternehmensgröße auf zentrale Unternehmensmerkmale

Wesentliche Datengrundlage ist die Leistungs- und Strukturstatistik (LSE).

Die LSE ist eine „Konzentrationsstichprobe“.

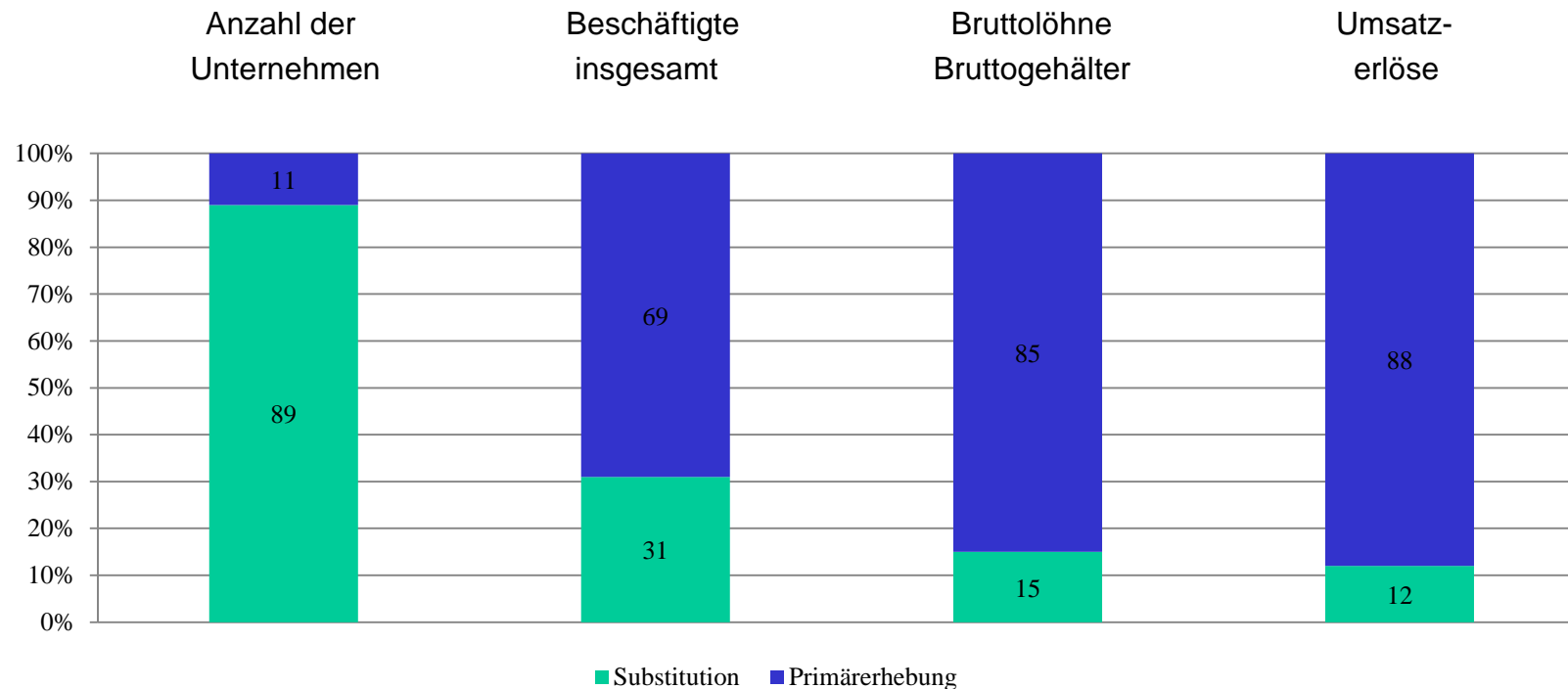
Für die Einheiten unter der Meldeschwelle stehen in den meisten Fällen Verwaltungsdaten für Beschäftigte, Umsatz und Lohn- und Gehaltssumme zur Verfügung.

Die anderen Ausprägungen für die Einheiten unter der Meldeschwelle werden durch komplexe Modellrechnungen geschätzt (s. DINGES, HAITZMANN, ZACH 2010).

Konsequenzen

Beispiel 2: Untersuchung des Einflusses der Unternehmensgröße auf zentrale Unternehmensmerkmale

Merkmalskonzentration – Primärabdeckung in der LSE 2008



Quelle: Statistik Austria, Leistungs- und Strukturstatistik 2008

Konsequenzen

Beispiel 2: Untersuchung des Einflusses der Unternehmensgröße auf zentrale Unternehmensmerkmale

Modellbasis: Auswahl der Donoren nach Art der Wirtschaftstätigkeit (differenziert nach 540 Wirtschaftszweigen) und nach Größenklassen des Umsatzes.

Zur Schätzung der Hauptmerkmale wird branchenspezifisch die Least Trimmed Squares Regression angewandt.

Zentrale Merkmale wie etwa die Wertschöpfung werden in Abhängigkeit von verfügbaren Merkmalen wie Umsatz oder Zahl der Beschäftigten geschätzt.

Sowohl Umsatz wie Zahl der Beschäftigten stehen für die Unternehmensgröße.

Konsequenzen

Beispiel 2: Untersuchung des Einflusses der Unternehmensgröße auf zentrale Unternehmensmerkmale

Die modellbasierte Datenergänzung für die Unternehmen unter der Meldeschwelle entspricht dem Modelltyp 2.

Für viele Untersuchungen ist die Relation „Bruttowertschöpfung pro Beschäftigten“ ein zentraler Indikator.

Wird die Untersuchung dieser Maßzahl für die kleinen Unternehmen auf dem gleichen Aggregationsniveau durchgeführt, das für die modellbasierte Datenergänzung (MDE) gewählt wurde, wird sie genau die Schätzung reproduzieren.

In allen anderen Fällen wird das Ergebnis sowohl die Reproduktion der Modellansätze als auch von Aggregationseffekten widerspiegeln.

Konsequenzen

Beispiel 3: Untersuchung der Wirtschaftsleistung nach Regionen

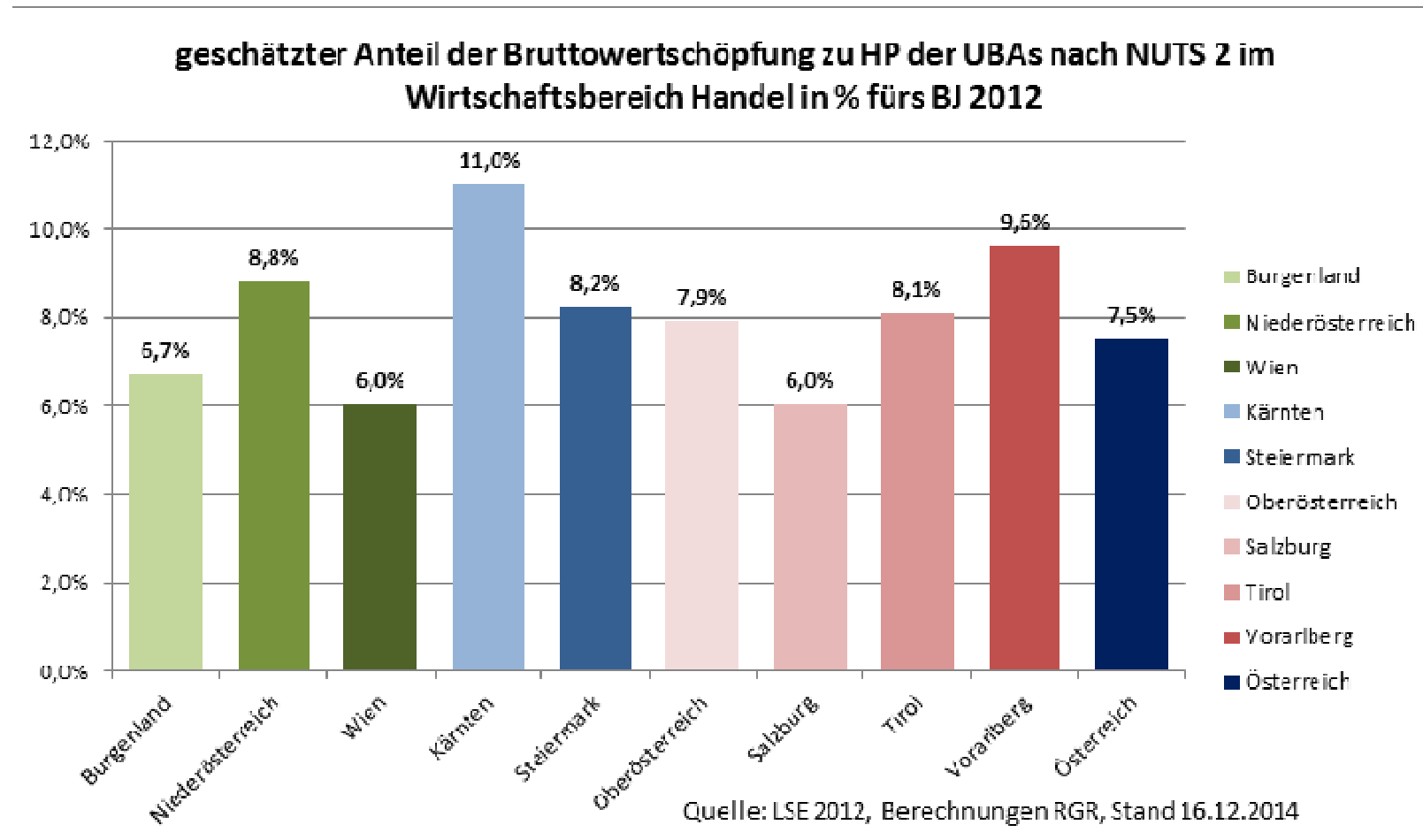
Unternehmen, die nur in einer Region tätig sind, können ohne Probleme mit allen ihren Merkmalen dieser Region zugeordnet werden.

Viele Unternehmen sind aber nicht „regionalhomogen“. Alle Merkmale, wie vor allem die Wertschöpfung, sind nur für das Unternehmen insgesamt beobachtbar, nicht aber für die in den Regionen aktiven Arbeitsstätten (genauer LKAUs).

Auf dieser Ebene sind nur wenige Merkmale (wie die Beschäftigung) deskriptiv erfassbar.

Für die Berechnung der für die Ableitung der Wertschöpfung erforderlichen Angaben (Erlös- und Vorleistungskomponenten) für die LKAU wird üblicherweise ein als „Pseudo-Bottom-up“ bezeichneter Ansatz gewählt.

Homogenität des Datenkörpers?



Konsequenzen

Beispiel 3: Untersuchung der Wirtschaftsleistung nach Regionen

Für jedes Unternehmen werden die Unternehmensmerkmale proportional zu den Beschäftigtendaten den örtlichen Einheiten zugerechnet (Modellansatz des Typs 3).

Die so ermittelten Schätzwerte für die einzelnen LKAU werden nach Wirtschaftsbereichen und Regionen aggregiert.

In diesen Aggregationsschritten werden Elemente unterschiedlichen kognitiven Charakters zusammengefasst.

Der Modellgehalt differiert nach Wirtschaftsbereichen und Regionen wesentlich und ist oft beträchtlich.

Konsequenzen

Beispiel 3: Untersuchung der Wirtschaftsleistung nach Regionen

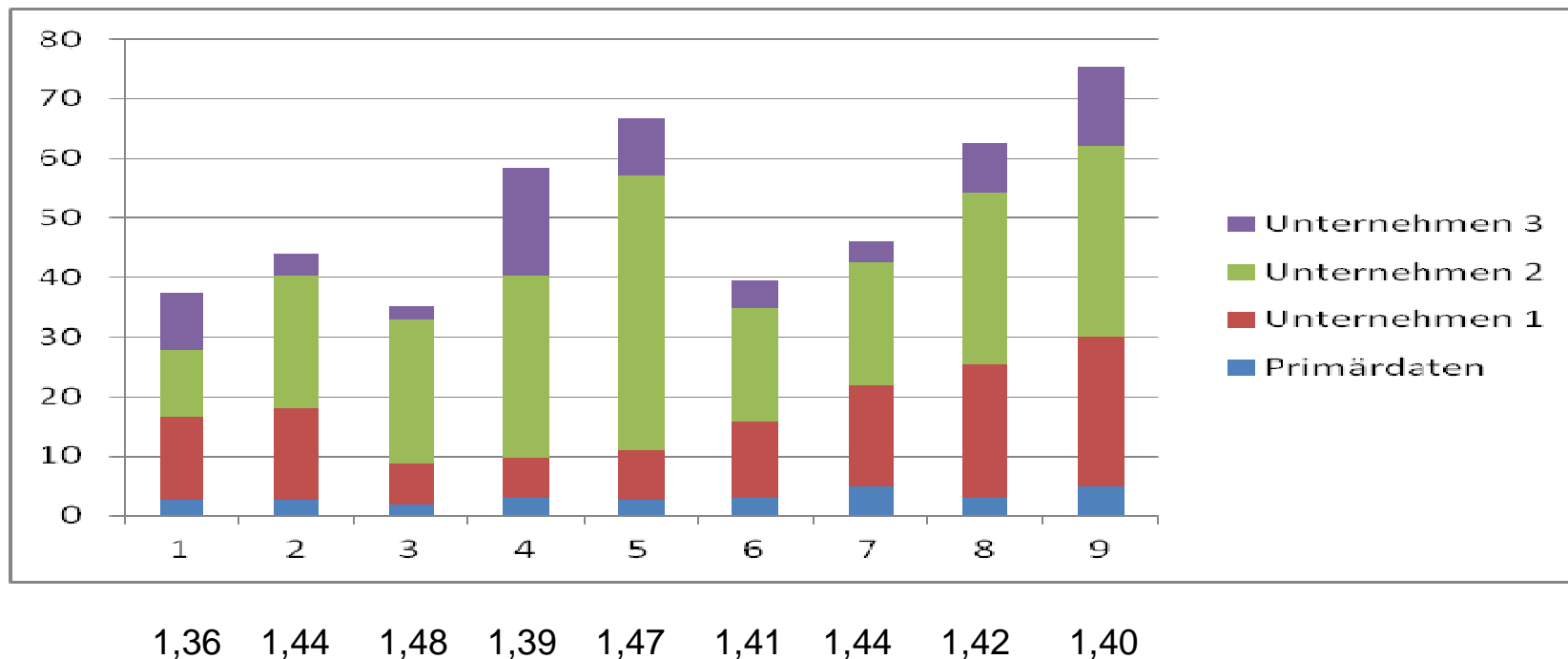
Die „ermittelte“ Wertschöpfung in einem Wirtschaftszweig und in einer Region entspricht ihrem Informationsgehalt oft in einem wesentlichen Teil nur einer „gewichteten Beschäftigtenzahl“.

Der auch für die Regionalanalyse oft als zentral angesehene Indikator „Bruttowertschöpfung pro Beschäftigten“ kann in vielen Wirtschaftsbereichen nicht den gewünschten Informationsgehalt transponieren.

Konsequenzen

Beispiel 3: Untersuchung der Wirtschaftsleistung nach Regionen

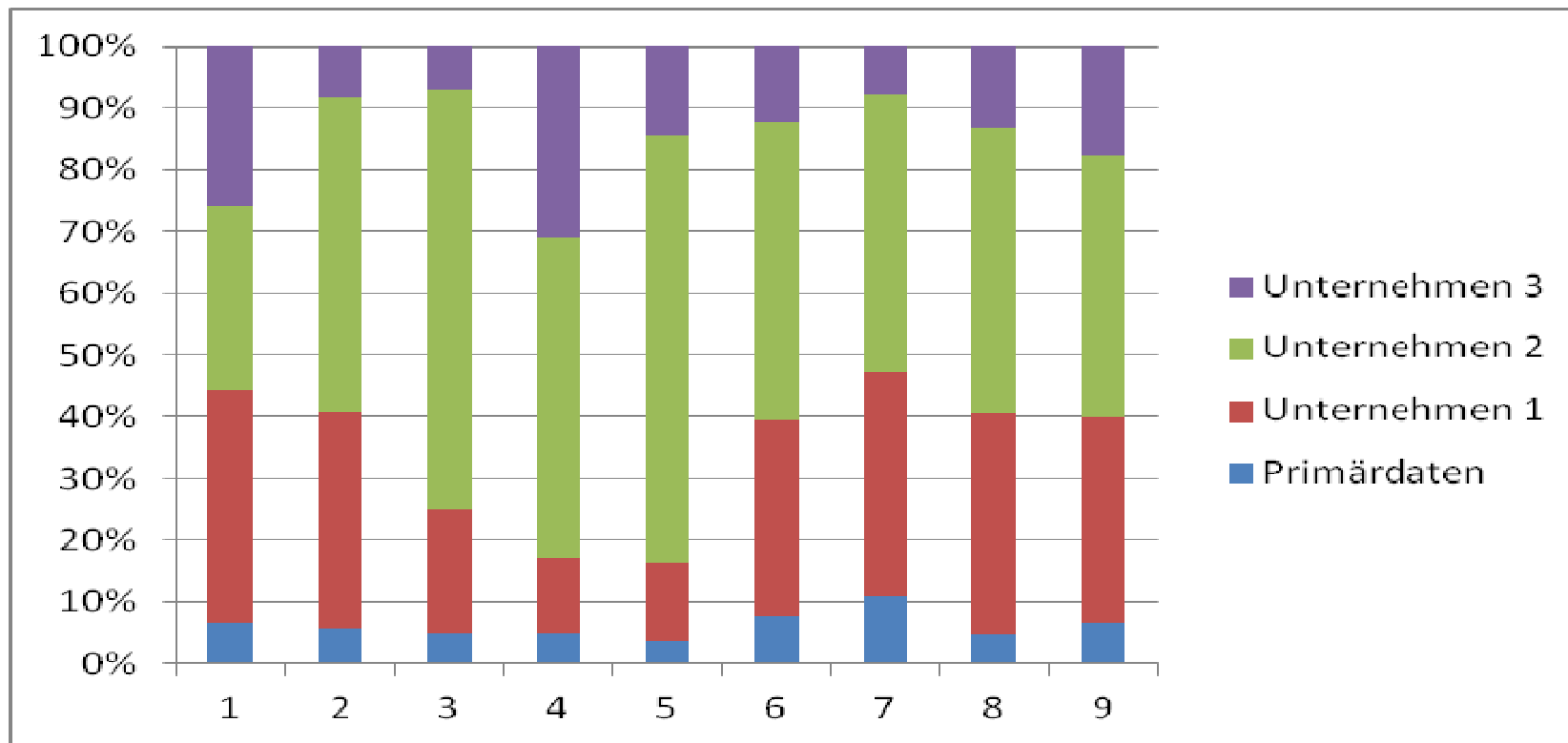
Eine Illustration für neun Regionen, BWS bzw. BWS pro Beschäftigten



Konsequenzen

Beispiel 3: Untersuchung der Wirtschaftsleistung nach Regionen

Illustration für neun Regionen - BWS - Anteile



Konsequenzen

Respezifikation des Modellansatzes

Ursprüngliche Spezifikation: $K_i^t = f(E^t, E^{t-1}, p_i^t, \dots)$

Datengenerierender Prozess: $\overline{K_i^t} = K_i^{t^o} \cdot \frac{U_j^t}{U_j^{t^o}}$

Respezifizierter Ansatz: $K_i^{t^o} \cdot \frac{U_j^t}{U_j^{t^o}} = f(E^t, E^{t-1}, p_i^t, \dots)$

Schlussbemerkungen

Funktion: Messung und ex-post Beschreibung

Modelle des Typs 1 unverzichtbar

Modellergebnisse des Typs 2 oft brauchbare Approximationen

Modellergebnisse des Typs 3 Umetikettierung von Information

Funktion: Grundlage der Wirtschaftsforschung – Modellierung

Modelle des Typs 1 unverzichtbar

Modellergebnisse des Typs 2

- Respezifikation des Modells
- Gefahr der Tautologie

Modellergebnisse des Typs 3

- Respezifikation des Modells
- Suggestion eines nicht gegebenen Wirklichkeitsbezugs

Homogenität des Datenkörpers?



Schlussbemerkungen

"In too many instances sophisticated statistical analysis is performed on a data set whose exact meaning and validity are unknown to the author" (LEONTIEF 1971).

"Somehow econometricians, theorists, and economic analysts of all stripes have lost essential communication with the compilers and synthesizers of their data. As a consequence, popular discourse, policymaking, and basic principles of economics have suffered inordinate confusion" (EISNER 1989).

"I have insisted that econometrics must have relevance to concrete realities — otherwise it degenerates into something which is not worthy of the name econometrics, but ought rather to be called playometrics" (FRISCH 1970).

Schlussbemerkungen

“Econometric theory is like an exquisitely balanced French recipe, spelling out precisely with how many turns to mix the sauce, how many carats of spice to add, and for how many milliseconds to bake the mixture at exactly 474 degrees of temperature.

But when the statistical cook turns to raw materials, he finds that hearts of cactus fruit are unavailable, so he substitutes chunks of cantaloupe; where the recipe calls for vermicelli he uses shredded wheat; and he substitutes green garment dye for curry, ping-pong balls for turtle's eggs and, for Chalifougnac vintage 1883, a can of turpentine“ (VALAVANIS 1959).

“A science that purports to deal with the real world but that ignores its empirical and observational side is likely to appear a rather empty and unproductive discipline” (EVANS, HOFFENBERG 1955).

Danke für die Aufmerksamkeit