

Zur Regionalisierung von Stichprobenergebnissen

PETER M. SCHULZE

Zusammenfassung

Es gibt eine steigende Nachfrage nach vielfältigen und schneller verfügbaren statistischen Informationen sowohl für administrative als auch für nichtadministrative Raumeinheiten. Kleinräumige Angaben sind meist nur aus Totalerhebungen erhältlich, die unregelmäßig stattfinden und nur wenige Variablen betreffen. Erforderlich sind deshalb aktuelle Schätzungen für verschiedene Charakteristika in kleinen Raumeinheiten. Um diese Anforderungen zu erfüllen, wurden verschiedene Verfahren entwickelt, die zusätzliche Informationen, z. B. aus übergeordneten Raumeinheiten oder einer Totalerhebung bei der Schätzung berücksichtigen. Zwei in diesem Zusammenhang — insbesondere in den USA — benutzte Verfahren, die Synthetische und die Regressions-Schätzung, werden in diesem Beitrag diskutiert und auf ihre Anwendbarkeit hin untersucht.

Summary

(Regionalization of sample survey results). There is growing demand for more timely and varied statistical data for administrative units and for small areas. Data for small domains are available from censuses, but not frequently and only for a few variables. Therefore it is often asked to produce intercensal estimates of various characteristics for small areas. To fulfil such requests, several methods which make use of supplementary information have been proposed. Two of the most frequently used small area estimation techniques — synthetic and regression estimates — are presented in this article together with their data requirements and limitations.

1. Einführung

Ohne Zweifel wächst die Nachfrage nach vielfältigen und schneller verfügbaren statistischen Daten sowohl für administrative als auch für nichtadministrative Raumeinheiten. Im Zusammenhang mit der Grundsatzdiskussion um Großzählungen zielen die Überlegungen immer mehr auf Stichprobenstatistiken. „Die allgemeinen Vorstellungen über die Möglichkeiten einer solchen Hilfestellung reichen von strikter Ablehnung der Einsatzfähigkeit von Stich-

proben bis hin zu einem geradezu leichtfertigen Optimismus, der in Stichproben eine Art Zaubermittel zur Beseitigung sämtlicher finanzieller — und methodischer (Hinzufügung vom Verfasser) — Hindernisse sieht“ (Nourney, 1978, S. 53). Zwar kann sich das regionalstatistische Datenangebot in der Bundesrepublik Deutschland sehen lassen, jedoch gibt es für spezielle quantitative Analysen nach wie vor Beschränkungen. Für bestimmte Fragestellungen liegen nämlich die vorhandenen Regionalstatistiken oft nicht sachlich genügend differenziert vor, oder sie sind nicht flächendeckend vorhanden bzw. nicht vergleichbar. Hinzu kommt, daß durch die gegenwärtige Datenschutzdiskussion in Zukunft insbesondere kleinräumige Analysen, die auf entsprechende statistische Daten angewiesen sind, eher behindert werden. Auch die Erhebungsprogramme der in immer größeren Zeitabständen durchgeführten Großzählungen werden zunehmend enger gefaßt, so daß der für regionalwirtschaftliche Analysen bestehende Datenbedarf auf anderem Wege befriedigt werden muß.

Man wird deshalb für kleinräumige Analysen sowohl auf vorhandenes Datenmaterial aus übergeordneten Raumeinheiten als auch auf (Teil-)Erhebungen in der konkreten Raumeinheit bzw. auf die Kombination beider zurückgreifen müssen (Hartenstein, 1976; Monheim, 1976; Böltken, Strubelt, 1985; Bretschneider, 1985).

Die Größe der betrachteten Region und die Art der Analyse beeinflussen dabei die Wahl und die Anwendbarkeit der zugrundeliegenden Datengewinnungsmethode.

Einer der in Frage kommenden Ansätze ist die Regionalisierung von Stichproben. In diesem Zusammenhang lassen sich zwei verschiedene Fragestellungen verfolgen. Zum einen besteht das Problem der Gewinnung von Regionalergebnissen aus Stichproben darin, „für die angestrebte Regionalisierungsebene ... solche Auswertungstabellen aufzustellen, die sich mit dem erhobenen Datenmaterial auffüllen lassen und deren Ergebnisse eine bestimmte, vorgegebene Genauigkeitsgrenze nicht unterschreiten“ (Gatzweiler, Philippczyk, 1979, S. 3). Hierbei geht es also darum, regionale Stichproben, die bereits erhoben vorliegen, für bestimmte Analysen weiter auszuwerten, zu regionalisieren, und zwar unter Einhaltung der Nebenbedingung vorgegebener Genauigkeitsanforderungen. Diese Fragestellung („Methode der Regionalaufbereitung“) ist insbesondere in der deutschsprachigen Literatur erörtert worden (Händler, 1970; Gatzweiler, Philippczyk, 1979; Derenbach, 1979).

Zum anderen kann daran gelegen sein, für eine bestimmte Region je nach Art der zugrundeliegenden Raumeinheit und der verfügbaren Daten (z. B. von Stichproben aus übergeordneten Raumeinheiten oder aus früheren Totalerhebungen) Schätzungen interessierender Parameter zu erhalten. Zu diesem letztgenannten Problem sollen im folgenden zwei oft benutzte Ansätze — nämlich die sogenannte Synthetische Schätzung und die Regressions-Schätzung — in ihren Grundzügen dargestellt werden. Diese Schätzungen lassen sich auch als eine besondere Art „gebundener Hochrechnungen“ interpretieren.

2. Schätzmethoden

2.1 Synthetische Schätzungen (synthetic estimators)

Die Synthetische Schätzung benutzt Stichprobenwerte, die auf höherer räumlicher Aggregationsstufe für die interessierende Variable in verschiedenen Gruppen (Klassen, Charakteristika) der Gesamtheit vorliegen, um eine Schätzung für eine Teilregion vorzunehmen. Sie gewichtet die Schätzungen in der übergeordneten Region im Verhältnis zum Vorkommen der Gruppe in der interessierenden Region (Gonzales, 1973, S. 33; Laake, 1978, S. 57). Der Begriff „synthetisch“ wird deshalb benutzt, weil die Schätzungen nicht direkt aus Erhebungsergebnissen gewonnen werden.

Es gebe in einer Gesamtregion (z. B. Bundesland) r ($r = 1, \dots, R$) Teilregionen (z. B. Landkreise) mit jeweils g ($g = 1, \dots, G$) Gruppen. Außerdem seien W_{rg} die (nichtstochastischen) Gewichtungsfaktoren der sich ausschließenden G Gruppen in der Teilregion r , so daß

$$(1) \quad \sum_r W_{rg} = 1.$$

Dann ist die Synthetische Schätzung x_r^* für den Totalwert in der Teilregion r

$$(2) \quad x_r^* = \sum_g W_{rg} x_{g,r}$$

wobei $x_{g,r}$ den Totalwert in der Gesamtregion für die Gruppe g darstellt. Dies zeigt auch das folgende Schaubild:

	1	2	...	g	...	G	
1							$x_{1,1}$
2							$x_{2,1}$
\vdots							\vdots
r				$x_{r,g}$ (nicht vorhanden)			x_r (unbekannt)
\vdots							\vdots
R							$x_{R,1}$
	$x_{1,1}$	$x_{2,1}$...			$x_{G,1}$	
				$x_{g,1}$ (gegeben)			

Die Gewichtungsfaktoren können beispielsweise Bevölkerungszahlen der Teilregion r in der Gruppe g (Z_{rg}) im Verhältnis zu Bevölkerungszahlen in der

Gruppe g aller Teilregionen ($Z_{.g}$) aus einer letzten Totalerhebung (Volkszählung) sein. Die Datenmatrix läßt sich allgemein hierzu wie folgt angeben:

	1	2	...	G	
1	Z_{11}	Z_{12}	...	Z_{1G}	$Z_{.1}$
2	Z_{21}	Z_{22}	...	Z_{2G}	$Z_{.2}$
\vdots	\vdots	\vdots		\vdots	\vdots
R	Z_{R1}	Z_{R2}	...	Z_{RG}	$Z_{.R}$
	$Z_{.1}$	$Z_{.2}$...	$Z_{.G}$	

Deshalb kann (2) geschrieben werden als

$$(3) \quad x_r^* = \sum_g \frac{Z_{rg}}{Z_{.g}} x_{.g}$$

Wird über alle R Raumeinheiten summiert, so erhält man eine Schätzung für die Gesamtregion

$$(4) \quad \sum_r x_r^* = \sum_r \sum_g \frac{Z_{rg}}{Z_{.g}} x_{.g} = \sum_g x_{.g} = x_{..}$$

Die vorangegangenen Überlegungen für den Totalwert lassen sich auf die Schätzung eines Mittelwertes übertragen. Angenommen, es soll beispielsweise das durchschnittliche Einkommen in einer Region geschätzt werden. Der vorhandene Beobachtungsumfang aus einer einfachen Stichprobe sei in Region r für eine verlässliche Schätzung jedoch zu klein. Es mögen aber wiederum zusätzliche Informationen — etwa aus einer früheren Volkszählung — über die Verteilung der Einkommen auf bestimmte Gruppen, z. B. nach Alter, Geschlecht oder Wirtschaftszweigen, vorliegen.

Der synthetische Durchschnittsparameter \bar{x}_r^* in Teilregion r ist dann mit

$$(5) \quad \bar{x}_r^* = \sum_g \frac{Z_{rg}}{Z_{.r}} \bar{x}_{.g}$$

gegeben.

Im Gegensatz zu (3) wird bei (5) im allgemeinen durch Mittelwertbildung für die Gesamtregion ein Fehler auftreten, d. h.

$$(6) \quad \frac{1}{R} \sum_r \bar{x}_r^* = \frac{1}{R} \sum_r \sum_g \frac{Z_{rg}}{Z_{.r}} \bar{x}_{.g} \neq \bar{x}$$

Dies rührt daher, daß beide Schätzungen zwar Z_{rg} benutzen, aber (5) verwendet es für das Verhältnis der Gruppen innerhalb der Teilregion, wohingegen (3) es als Verhältnis der Teilregionen innerhalb der g Gruppen ansetzt.

Synthetische Schätzungen sind im allgemeinen aus zwei Gründen nicht erwartungstreu (Gonzales, Hoza, 1978, S. 7). Zunächst wird in (3) unterstellt, daß sich die Merkmalssumme $x_{.g}$ gleichmäßig auf die Teilregionen verteilt. In (5) wird angenommen, daß $\bar{x}_{.g}$ regionsinvariant ist, d. h. für alle Regionen gleich ist. Diese Annahmen können u. U. wichtige örtliche Schwankungen in den Teilregionen ignorieren. Außerdem beziehen sich die Gewichtungsfaktoren W_{rg} auf — eventuell weit zurückliegende — Daten einer Totalerhebung und können sich mittlerweile verändert haben.

Die Verzerrung läßt sich formal für den Totalwert wie folgt ausdrücken (Purcell, Kish, 1979, S. 371):

$$(7) \quad E(x_r^* - X_r) = E\left(\sum_g \frac{Z_{rg}}{Z_{.g}} x_{.g} - \sum_g X_{rg}\right) = \sum_g Z_{rg} \left[\frac{X_{.g}}{Z_{.g}} - \frac{X_{rg}}{Z_{rg}}\right],$$

wobei die X die wahren Werte der Variablen in der Grundgesamtheit bezeichnen. Es ergibt sich nur dann eine erwartungstreue Schätzung, wenn

$$\frac{X_{.g}}{Z_{.g}} = \frac{X_{rg}}{Z_{rg}}$$

für alle g Gruppen gilt, was im allgemeinen nicht angenommen werden kann.

Die Anwendung dieses Ansatzes wird deshalb durch die Nichtberücksichtigung kleinräumiger Besonderheiten und die Benutzung veralteter Totalerhebungsdaten eingeschränkt.

Um die Genauigkeit der synthetischen Schätzung zu erkennen, ist die Berechnung des mittleren quadratischen Fehlers (MSE)¹⁾ wünschenswert (Laake, 1978, S. 58)

$$(8) \quad \text{MSE}(x_r^*) = E(x_r^* - X_r)^2 = \text{Var}(x_r^*) + \text{Bias}^2(x_r^*) \\ = \sum_g W_{rg}^2 \text{Var}(x_{.g}) + [E(x_r^*) - X_r]^2$$

Hierbei ist X_r der wahre Wert in Teilregion r . Für große Stichproben kann die $\text{Var}(x_r^*)$ relativ klein sein im Verhältnis zur $\text{Var}(x_r)$, einer Schätzung, die beispielsweise aus einer reinen Zufallsauswahl in Teilregion r gewonnen wurde. Aber dieser Vorteil kann durch den Bias (x_r^*) zunichte gemacht werden. Die Verzerrung läßt sich bei praktischen Anwendungen jedoch nicht berechnen, da X_r unbekannt ist.

Empirische Untersuchungen für die USA haben gezeigt, daß Synthetische Schätzungen für einige Variablen zu brauchbaren Ergebnissen führen, für andere aber nicht (DiGaetano, Waksberg, MacKenzie, Yaffe, 1980; Schaible, Brock, Casady, Schnack, 1979).

¹⁾ Die mögliche Schätzung eines durchschnittlichen MSE zeigt Gonzales, 1973, S. 34.

Die Verzerrung hängt mit der Wahl der Gruppierungsvariablen zusammen. Wenn diese nicht mit der interessierenden Variablen hoch korrelieren, werden die Synthetischen Schätzungen dazu tendieren, nahe am Mittel für die Gesamtregion zu liegen, und damit die lokalen Faktoren kaum widerspiegeln. Wenn die Gruppierungsvariablen jedoch sorgfältig ausgewählt werden, können die Schätzungen durchaus in manchen Fällen zu brauchbaren Ergebnissen führen (Purcell, Kish, 1979, S. 371). Wird bereits bei der Erhebungsplanung die Möglichkeit späterer Synthetischer Schätzungen ins Auge gefaßt, so lassen sich Verbesserungen der Schätzungen erreichen (Tam, 1982).

Ohne Zweifel liegt der Hauptvorteil Synthetischer Schätzungen in ihrer Plausibilität und der einfachen Berechnungsmöglichkeit.

2.2 Regressionschätzungen

Dieser Ansatz benutzt die multiple Regressionsanalyse zur Bestimmung interessierender Größen in der Raumeinheit r (vgl. zu diesem Ansatz auch Purcell, Kish, 1979; Purcell, Kish, 1980).

Die Funktionalbeziehung wird geschätzt zwischen der interessierenden abhängigen Variablen (criterion variable) und den charakteristischen unabhängigen Variablen (symptomatic variables) in Form von Verhältnisgrößen (ratios).

Im einfachsten Fall wird unterstellt, daß die gleiche Beziehung zwischen der interessierenden und den charakteristischen Variablen besteht, und zwar sowohl zwischen zwei Totalerhebungen ($t = 1, 2$) als auch für Zeitpunkte danach (z. B. $t = 3$). Zunächst wird für jede der p charakteristischen Variablen X das Verhältnis zwischen der Größe in der Raumeinheit r und dem zugehörigen Wert im Gesamttraum für die Zeitpunkte der beiden Totalerhebungen berechnet²⁾, also

$$(9) \quad P_{rti} = \frac{X_{rti}}{\sum_r X_{rti}}$$

Hierbei ist X_{rti} der Wert für die i -te charakteristische Variable ($i = 1, \dots, p$) in der Raumeinheit r ($r = 1, \dots, R$) zum Zeitpunkt t . Diese Größen müssen für die Raumeinheiten laufend verfügbar sein, sowohl für die Zeitpunkte der Totalerhebungen als auch für spätere Zeitpunkte (z. B. Geburten-, Sterbe-, Beschäftigtenzahlen aus amtlichen Registern).

Entsprechend wird für die interessierende Variable Y das Verhältnis

$$(10) \quad Q_{rt} = \frac{Y_{rt}}{\sum_r Y_{rt}}$$

²⁾ Die Verhältnisbildung wird vorgenommen, um Variation und Schiefe der zugehörigen Verteilungen zu reduzieren.

gebildet. Dann werden zwei Verhältnisgrößen V für die Variablen der Regressionsbeziehung aus beiden Zeitpunkten der Totalerhebungen (Index a) berechnet, nämlich

$$(11) \quad V_{rai} = \frac{P_{r2i}}{P_{r1i}}$$

und

$$(12) \quad V_{ra} = \frac{Q_{r2}}{Q_{r1}}$$

Eine zweite Gruppe von Verhältnisgrößen zeigt die Veränderung zwischen Zeitpunkt 2 (letzte Totalerhebung) und Zeitpunkt 3 (aktuelle Datenerhebung; Index b) bei den charakteristischen Variablen

$$(13) \quad V_{rbi} = \frac{P_{r3i}}{P_{r2i}}$$

Nun läßt sich zunächst unter Benutzung der Verhältnisgrößen (11) und (12) aus allen betrachteten Raumeinheiten die Beziehung

$$(14) \quad V_{ra} = \beta_0 + \beta_1 V_{rai} + \dots + \beta_p V_{rap} + \varepsilon_r$$

schätzen. Die geschätzten Koeffizienten $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ beruhen auf den Verhältnissen zwischen den beiden Basisperioden ($t = 1, 2$). Sie werden anschließend benutzt, um zusammen mit V_{rbi} aus (13) das Verhältnis V_{rb} schätzen zu können

$$(15) \quad \hat{V}_{rb} = \hat{\beta}_0 + \hat{\beta}_1 V_{rbi} + \dots + \hat{\beta}_p V_{rbp}$$

Ist \hat{V}_{rb} nach (15) bestimmt, so können dann die interessierenden aktuellen Variablen für die Raumeinheit r (Y_{r3}) berechnet werden, indem man \hat{V}_{rb} multipliziert mit der Verhältnisgröße von Y aus der letzten Totalerhebung ($t = 2$) und dem aktuellen Wert ($t = 3$) für alle Regionen

$$(16) \quad \hat{Y}_{r3} = \hat{V}_{rb} \left[\frac{Y_{r2}}{\sum_r Y_{r2}} \right] \sum_r Y_{r3}$$

Die Methode beruht auf der Annahme, daß die festgestellte Funktionalbeziehung zwischen den Totalerhebungen ($t = 1, 2$) auch im aktuellen Zeitpunkt ($t = 3$) gilt.

Diese Regressionschätzung läßt sich dahingehend modifizieren, daß die β_i der Regressionsbeziehung (14) nicht aus Totalerhebungsdaten, sondern aus Stichprobendaten berechnet werden (Ericksen, 1973; Ericksen, 1974). Die Stichprobeninformation kann unterschiedlicher Natur sein. In manchen Fällen liegen Stichprobenergebnisse für Gesamtregionen (übergeordnete Raumeinheiten) vor, in anderen dagegen Stichprobenschätzungen für nur einige interessierende Raumeinheiten, und in wiederum anderen Fällen können Totalerhebungsergebnisse für eine Stichprobe interessierender Raumeinheiten vorliegen.

Für die Raumeinheiten, für die Stichprobenwerte in $t = 3$ vorhanden sind, wird zunächst die interessierende (abhängige) Variable mit

$$(17) \quad \hat{v}_{rb} = \frac{\hat{q}_{r3}}{q_{r2}}$$

geschätzt. Mit den korrespondierenden Werten für die p charakteristischen Variablen läßt sich die zugehörige Regressionsgleichung mit

$$(18) \quad \hat{v}_{rb} = \hat{\beta}_0 + \hat{\beta}_1 V_{rb1} + \dots + \hat{\beta}_p V_{rbp} + e_r$$

angeben. In diese Gleichung werden dann die Werte für die charakteristischen Variablen eingesetzt, um Schätzungen von \hat{v}_{rb} für alle interessierenden Raumeinheiten zu erhalten, gleichgültig ob sie in der Stichprobe enthalten waren oder nicht. Dann ist auch analog zu (16) die Variable Y auf Stichprobenbasis zu errechnen.

Schätzt man die Regressionskoeffizienten aus den strukturellen Beziehungen, die zum früheren Zeitpunkt von Totalerhebungen bestanden, so bringen Änderungen im Zeitablauf aufgrund der eingehenden Verhältnisgrößen Fehler in die laufenden Schätzungen. Diese Fehler lassen sich dadurch vermeiden, daß man die Regressionskoeffizienten aus laufenden Stichproben in wenigstens einem Teil der Raumeinheiten schätzt. Das zuletzt skizzierte Vorgehen wird allerdings erkaufte durch einen Stichprobenfehler in den Regressionskoeffizienten $\hat{\beta}_j$. Dieser hat zwei Komponenten: eine „between area component“, die dann auftritt, wenn nur in einer Stichprobe von Raumeinheiten Stichproben gezogen werden, und eine „within area component“, die auf den Stichprobenfehler in den einzelnen Raumeinheiten zurückzuführen ist. Der mittlere quadratische Fehler der Stichprobenregression läßt sich wie folgt angeben (zur Ableitung vgl. Ericksen, 1974, S. 875):

$$(19) \quad \text{MSE}(\hat{v}_{rb}) = \frac{(n - p - 1)}{n} \sigma_u^2 + \frac{(p + 1)}{n} \sigma_v^2.$$

Hierbei ist σ_u^2 die durch die p unabhängigen Größen unerklärte Varianz zwischen den Stichproben-Raumeinheiten und σ_v^2 ist die Varianz innerhalb der Stichproben-Raumeinheiten.

Eine weitere Modifikation des Regressionsansatzes ist möglich, wenn man auf die unter Punkt 2.1 diskutierten synthetischen Variablen zurückgreift und diese als erklärende Größen in die Regressionsbeziehung aufnimmt, um die Information über die interessierende Raumeinheit zu erhöhen (Gonzales, Hoza, 1978, S. 10 ff.)³⁾.

³⁾ Zu weiteren, weniger benutzten Regressionsansätzen und Anwendungshinweisen vgl. Purcell, Kish, 1979, S. 327 ff.; Purcell, Kish, 1980, S. 5 f.

3. Abschließende Bemerkungen

Zunächst ist zu sagen, daß es keine Methode zur Regionalisierung von Stichproben gibt, die für alle Situationen angezeigt ist. Denn für konkrete Fragestellungen gibt es hinsichtlich der prinzipiell zur Verfügung stehenden Daten große Unterschiede in den Quellen und deren Qualität, seien es nun Werte aus Totalerhebungen (z. B. Volkszählung), Stichproben (z. B. Mikrozensus) oder amtlichen Registern (z. B. Melderegister).

Darüber hinaus hängt die Wahl des Verfahrens und der Datenquellen auch von der Raumeinheit, auf die sich die Analyse beziehen soll, ab. Außerdem ist bei Totalerhebungsdaten einerseits und Stichprobenwerten andererseits abzuwägen zwischen Überalterung bei ersteren — wenn diese zeitlich lange zurückliegen — und dem Stichprobenfehler bei letzteren (Kish, 1979).

Bei entsprechend verfügbaren Daten haben die skizzierten Regressionsansätze im Vergleich zur Benutzung von charakteristischen (symptomatischen) Variablen bessere Ergebnisse gebracht, z. B. bei der Schätzung regionalen Bevölkerungswachstums (Purcell, Kish, 1979, S. 376).

Die unter Punkt 2.1 erörterte Synthetische Schätzung erscheint aber als der einfachste, am weitesten verbreitete und anpassungsfähigste Ansatz (Purcell, Kish, 1979, S. 377). Es gibt keine ausdrücklichen Modellannahmen, deren Gültigkeit ansonsten — wie z. B. bei der Regressionsschätzung — meist stillschweigend unterstellt wird. Vielmehr wird die Stabilität verschiedener Wechselbeziehungen zwischen der interessierenden Variablen und den Variablen, die die Untergruppen beschreiben, angenommen. Sind allerdings diese impliziten Annahmen nicht erfüllt, so werden Verzerrungen auftreten.

Ein direkter Vergleich zwischen Synthetischer und Regressionsschätzung scheint wenig geeignet, da beide Methoden unterschiedliche Datenanforderungen stellen. Es ist aber zu überlegen, ob — wie oben erwähnt — zur Verbesserung der Schätzungen beide Verfahren kombiniert werden können. Von den gegenwärtig verfügbaren Ansätzen besitzen diese beiden augenscheinlich das größte Potential zur Regionalisierung von Stichprobenergebnissen.

Man sollte aber auch hier — wie ohnehin aus der Regressionsanalyse bekannt — beachten, daß eine Verfeinerung der Schätzmethode oder die Hinzunahme weiterer Informationen nicht zwangsläufig zu einer Verbesserung der Schätzergebnisse, d. h. zu einem Effizienzgewinn, führt.

Literatur

- Böltken, F., Strubelt, W., 1985: Möglichkeiten und Voraussetzungen der Nutzung von Daten aus der Umfrageforschung für den Ausbau der Laufenden Raumbbeobachtung. In: Informationen zur Raumentwicklung, Heft 5.
Bretschneider, M., 1985: Umfragen in der Kommunalstatistik. In: Informationen zur Raumentwicklung, Heft 5.

- Derenbach, R., 1979: Die Wohnverhältnisse im regionalen Vergleich. Bundesforschungsanstalt für Landeskunde und Raumordnung, Bonn.
- DiGaetano, R., Waksberg, J., MacKenzie, E., Yaffe, R., 1980: Synthetic Estimates for Local Areas from the Health Interview Survey. In: Proceedings of the Section on Survey Research Methods (American Statistical Association).
- Erickson, E. P., 1973: A Method for Combining Sample Survey Data and Symptomatic Indicators to Obtain Population Estimates for Local Areas. In: Demography 10.
- Erickson, E. P., 1974: A Regression Method for Estimating Population Changes of Local Areas. In: Journal of the American Statistical Association 69.
- Gatzweiler, H. P., Philippczyk, W., 1979: Gewinnung von Regionalergebnissen aus Stichproben. Bundesforschungsanstalt für Landeskunde und Raumordnung, Bonn.
- Gonzales, M. E., 1973: Use and Evaluation of Synthetic Estimates. In: Proceedings of the Social Statistical Section (American Statistical Association).
- Gonzales, M. E., Hoza, C., 1978: Small-Area Estimation with Application to Unemployment and Housing Estimates. In: Journal of the American Statistical Association 73.
- Händler, H., 1970: Untersuchung der Verwendungsmöglichkeiten allgemeiner Stichprobenerhebungen in der Regionalforschung. Marburg.
- Kish, L., 1979: Samples and Censuses. In: International Statistical Review 47.
- Hartenstein, W., 1976: Bessere Informationen zur Raumentwicklung: Sechs Thesen und zwei Strategien. In: Raumforschung und Raumordnung 34.
- Laake, P., 1978: An Evaluation of Synthetic Estimates of Employment. In: Scandinavian Journal of Statistics 5.
- Monheim, H., 1976: Räumliche Planung und amtliche Statistik. In: Raumforschung und Raumordnung 34.
- Nourney, M., 1978: Regionalisierung von Stichproben. In: Allgemeines Statistisches Archiv 62.
- Purcell, N. J., Kish, L., 1979: Estimation for Small Domains. In: Biometrics 35.
- Purcell, N. J., Kish, L., 1980: Postcensal Estimates for Local Areas (or Domains). In: International Statistical Review 48.
- Schaible, W. L., Bock, D. B., Casady, R. J., Schnack, G. A., 1979: Small Area Estimation: An Empirical Comparison of Conventional and Synthetic Estimators for States. U.S. Department of Health, Education, and Welfare Public Health Service, Hyattsville, Md.
- Tam, S. M., 1982: Postcensal Estimates for Local Areas Using Current Samples with Census as the Source of Sampling Frame. In: International Statistical Review 50.