

**ZUR VERSUCHSPLANUNG IN
ÖKONOMIE UND ÖKONOMETRIE**

WERNER G. MÜLLER

Forschungsbericht/
Research Memorandum No. 270

September 1990

Die in diesem Forschungsbericht getroffenen Aussagen liegen im Verantwortungsbereich des Autors/der Autorin (der Autoren/Autorinnen) und sollen daher nicht als Aussagen des Instituts für Höhere Studien wiedergegeben werden. Nachdruck nur auszugsweise und mit genauer Quellenangabe gestattet.

All contributions are to be regarded as preliminary and should not be quoted without consent of the respective author(s). All contributions are personal and any opinions expressed should never be regarded as opinion of the Institute for Advanced Studies.

This series contains investigations by the members of the Institute's staff, visiting professors, and others working in collaboration with our departments.

Abstract

Diese Arbeit soll einerseits einen Eindruck von der Verwendbarkeit experimenteller Instrumentarien im Wirtschaftssektor vermitteln und andererseits die verfügbaren statistischen Methoden umreißen. Außerdem soll aus Gründen der Praxisnähe eine bereits erfolgte Anwendung nämlich das 'Graduated Work Incentive Experiment' von Conlisk & Watts [1969] kurz beschrieben werden.

This paper shall on the one hand give an impression of the applicability of experimental tools in the economic sector, on the other describe the available statistical methods. Additionally, for being more practically, an application from the U.S., the 'Graduated Work Incentive Experiment' of Conlisk & Watts [1969], will be discussed briefly.

1 Einleitung

Trotz der engen Beziehungen zwischen klassischen statistischen Techniken (z.b. der Regressionsanalyse) und ökonomischen sowie insbesondere ökonomischen Verfahren kam es kaum zu einer gemeinsamen Entwicklung dieser Bereiche mit der von Kiefer [1959] mitbegründeten Theorie der Versuchsplanung.

Dies resultierte aus der intuitiv ansprechenden Auffassung der Wirtschaftsdisziplinen als 'nicht-experimentelle' Wissenschaftszweige im Gegensatz etwa zur Physik, Medizin, Chemie oder in der Industrie zur Qualitätskontrolle.

In den letzten zwanzig Jahren gab es aber (zwar nur vereinzelte) Versuche den großen Vorteil der experimentellen Kontrolle von Variablen auch dem Ökonomen zugänglich zu machen. Es wurden sowohl wichtige theoretische Vorschläge in dieser Richtung gemacht (z.b. Castro & Weingarten [1970] und Naylor [1972]), als auch, vor allem in den Vereinigten Staaten, konkrete Experimente unternommen (Conlisk & Watts [1969]). Dabei zeigte sich, daß gerade aufgrund der enorm hohen Kosten von Experimenten in der Wirtschaft, der Versuchsplanung hier ein eindeutig zu geringer Stellenwert zukommt.

2 Anwendungsgebiete

Aus der zerstreuten Literatur über ökonomische Anwendungen der Versuchsplanung lassen sich vier Hauptgebiete klassifizieren: (a) Real-world Experimente, (b) Inframenschliche Experimente, (c) Computersimulationsexperimente, (d) Spielexperimente.

2.1 Real-world (oder sozioökonomische) Experimente

bilden das Analogon zu den klassischen physikalischen Versuchen, bei denen Objekte verschiedenen kontrollierbaren Bedingungen unterworfen werden. In sozioökonomischen Experimenten wird das Verhalten von Wirtschaftssubjekten, also Individuen, Familien, Betrieben oder von Industriezweigen, bezie-

ungsweise der gesamten Volkswirtschaft, untersucht.

Eine umfassendere Definition dieser Art von Versuchen geben Ferber & Hirsch [1979]:

[...] the type of social experiment being carried out in economics may be defined as a rigorously designed study involving expenditures of public monies and incorporating experimental aspects applied over a period of time to one or more segments of a human population, with the aim of evaluating the economic and social effects of the experimental treatments.

Derartige Versuche sollen jedoch möglichst nicht nur direkte Auswirkungen anwendbarer 'treatments' messen, sondern auch Nebeneffekte von eventuell größerer Wichtigkeit enthüllen.

Es ist klar, daß aufgrund der besonderen Versuchsbedingungen sozioökonomische Experimente einigen spezifischen Verzerrungseffekten unterliegen:

Der bekannteste davon ist der in psychologischen Versuchen entdeckte *Hawthorne-effekt*, bei dem das Wissen der Subjekte an einem Experiment teilzunehmen von den Kontrollparametern unabhängige Auswirkungen zeigt.

Ein sehr ernstzunehmendes Problem sind auch die *Zeithorizont-effekte*. Da solche Versuche natürlich eine begrenzte Dauer haben, können veränderte Bedingungen von den Versuchsteilnehmern nicht als permanent angenommen werden, und sie werden daher ihr Verhalten nicht in dem Maß anpassen, als handle es sich um einen dauerhaften Wandel. Eine mögliche, wenn auch nicht gänzlich befriedigende Vorgangsweise bestünde in einer Aufnahme des Zeithorizonts als Kontrollvariable im Versuch.

Die in der Regel gegebene Möglichkeit der freien Entscheidung bei sozioökonomischen Experimenten teilzunehmen, führt zu einer weiteren wichtigen Art von Effekten, den *Ausfallsverzerrungen* (wenn es also zu einer Wechselwirkung zwischen Ausfallraten und Art der Behandlung kommt).

Nicht zuletzt sollten auch die eventuell auftretenden Einflüsse *exogener Variablen* nicht vernachlässigt werden. Anders als etwa bei physikalischen Versuchen ist es nicht möglich die Umwelt total zu kontrollieren, deshalb könnte selbst bei einem guten Versuchsplan durch unvorhersehbare Umweltveränderungen das gesamte Experiment ad absurdum geführt werden.

Zusätzlich zu all den angeführten Problemen erheben sich natürlich beim Umgang mit Menschen auch *ethische Probleme*, die aus der durch den Versuchsplan bedingten unterschiedlichen Behandlung der Testsubjekte resultieren.

Eine detailliertere Charakterisierung von sozioökonomischen Versuchen mit Beispielen geben Ferber & Hirsch [1979].

2.2 Inframenschliche Experimente

sind eine vor allem von Castro & Weingarten [1970] forcierte Abart sozioökonomischer Experimente. Hierbei sind die Versuchsobjekte Tiere (manchmal auch Kleinkinder) aus deren Reaktionen man auf menschliches (bzw. erwachsenes) Verhalten zu schließen sucht, ohne die im obigen Abschnitt angeführten Probleme in Kauf nehmen zu müssen. Mögliche Ziele sind dabei z.B.: das Aufstellen experimenteller Nachfragefunktionen, Arbeitsangebotsfunktionen, Sparverhalten bei Preisänderungen usw..

2.3 Computersimulationsexperimente (kurz CSE)

bilden ein eher neues Anwendungsgebiet für Versuchplaner. Obwohl bereits Naylor in den späten sechziger bzw. frühen siebziger Jahren (Naylor et al. [1967]) erste Schritte in diese Richtung tat, beginnt erst jetzt das Anwenderinteresse zu erwachen.

Die neue Aufgabenstellung erwuchs aus der steigenden Komplexität von (vorwiegend ökonomischen) Computermodellen und ihrer daraus resultierenden Unüberschaubarkeit. Ein gegenüber dem Computercode (Originalmodell) stark vereinfachtes Metamodell soll es dem Anwender ermöglichen, ursächliche Zusammenhänge besser zu durchschauen und die intensive Computerbenützung (aus Kosten- oder Zeitgründen) zu verringern.

Die Konstruktion solcher Metamodelle läuft im allgemeinen in zwei Phasen ab (was konsequenterweise zwei Arten von Versuchsplänen erfordert):

- **Screening** Eine große Anzahl von im Modell implementierten Faktoren sind für eine spezielle Fragestellung unbedeutend und können daher

aussortiert werden. Im allgemeinen werden zu diesem Zweck Zufallsdesigns für stufenweise Regressionsanalyse verwendet.

- **Untersuchung der Zusammenhänge** Jetzt wird das eigentliche Metamodelle aus den substantiellen Faktoren gebildet. Es bieten sich klassische Verfahren der Versuchsplanung für Regressionsexperimente oder Interpolationsdesigns an.

Einige Ansätze dafür finden sich in Naylor et al. [1967], Kleijnen [1982] und Fedorov [1983].

2.4 Spieleexperimente

bilden ein Zwischending von (2.1) und (2.3). Es werden Individuen oder Gruppen mit einem Modell konfrontiert, das die Umwelt repräsentiert.

Zusammenfassende Betrachtungen über die Anwendungen der Versuchsplanung in der Ökonometrie geben Castro & Weingarten [1970], Naylor [1972] und Papakyriazis [1978].

3 Das statistische Instrumentarium

Das hinter einem guten Versuchsplan stehende mathematische Hauptprinzip ist absolut einleuchtend:

Gibt man sich entweder Hypothesen-testen oder Parameter-schätzen als Ziel des Experiments vor, so bleibt die Aufgabe eine gewisse (womöglich gegebene) Anzahl von Messungen so durchzuführen, daß entweder die Güte des Tests oder die Präzision der Schätzung maximiert wird.

Weiters soll auch auf die Unverzerrtheit der Methode sowie auf die Begrenzung der experimentellen Bedingungen durch eine Versuchsregion geachtet werden.

Im weiteren soll nun vor allem auf den Kern der Versuchsplanungstheorie, die von Kiefer [1959] initiierten Methoden zum 'response surface' Design und die daraus resultierenden Entwicklungen, also im Bereich der Versuchsplanung von Regressionsexperimenten, hingewiesen werden.

Grundlegendes über den zweiten großen Zweig der Versuchsplanung, dem kategorialen Design findet sich in der bekannten Monographie von Fisher [1935].

3.1 Definitionen

Man nennt den Variablensatz:

$$\begin{pmatrix} y_{11}, y_{12}, \dots, y_{1N_1}; & y_{21}, y_{22}, \dots, y_{2N_2}; & \dots; & y_{n1}, y_{n2}, \dots, y_{nN_n}; \\ \sigma_1^2; & \sigma_2^2; & \dots; & \sigma_n^2; \\ x_1; & x_2; & \dots; & x_n; \end{pmatrix}$$

ein Experiment $\mathcal{E}(n, N)$, wobei y_{ij} für die beobachteten Werte, σ_i^2 für ihre Varianzen und x_i für die Designpunkte (das sogenannte Spektrum) steht ($N = \sum_{i=1}^n N_i$).

Den Satz:

$$\begin{pmatrix} p_1, & p_2, & \dots, & p_n \\ x_1, & x_2, & \dots, & x_n \end{pmatrix}$$

wobei $p_i = N_i/N$ und $\sum_{i=1}^n p_i = 1$ nennt man den normalisierten Versuchsplan $\xi(n)$. Die p_i können allerdings auch als Präzision oder Dauer der Messung aufgefaßt werden.

Ist die Zahl der Beobachtungen groß genug, kann man ein stetiges Design ξ anstelle des diskreten $\xi(n)$ verwenden und somit zu den leichter handhabbaren Techniken der stetigen Mathematik übergehen. Dieses stetige (normalisierte) Design ξ ist dann charakterisiert durch ein Wahrscheinlichkeitsmaß $\xi(x)$:

$$\int_{\mathcal{X}} \xi(x) dx = 1, \quad \xi(x) > 0, \quad \forall x \in \mathcal{X} \quad (1)$$

wobei \mathcal{X} für die Versuchsregion steht.

Man nennt solche Designs im Gegensatz zu den in der Praxis exakt anwendbaren diskreten Designs auch approximative Designs, da man erst durch Runden der Gewichte einen geeigneten Versuchsplan erhält.

3.2 Optimalitätskriterien

Die Durchführung genauer Parameterschätzung in der Ökonometrie kann verschiedene Gründe haben: entweder handelt es sich um die Analyse ökonomischer Theorien, die exakte Vorhersage wirtschaftlicher Ereignisse oder die Bestimmung der optimalen Politik.

Mathematisch kann man die Güte einer Parameterschätzung $\hat{\theta}$ des Parameters θ am besten mit Hilfe des Bias und der Varianzkovarianz Matrix eines Schätzers ausdrücken.

Beschränkt man sich auf unverzerrte Schätzer, so ist es klar, daß ein Experiment \mathcal{E}_1 mit der Varianzkovarianz Matrix $D_1[\hat{\theta}]$ einem Experiment \mathcal{E}_2 mit $D_2[\hat{\theta}]$ vorzuziehen ist, wenn:

$$(D_2[\hat{\theta}] - D_1[\hat{\theta}]) \quad (2)$$

eine positiv-definite Matrix ist.

Im allgemeinen ist es aber nicht möglich einen eindeutigen, optimalen Versuchsplan für solch ein Experiment zu finden, weil es sich aufgrund der Dimension von $D[\hat{\theta}]$ ($k \times k$) um ein multiobjektives Optimierungsproblem handelt. Man muß also faktisch einen Kompromiß zwischen Varianzen und Kovarianzen der einzelnen Parameterschätzer schließen und zu einer skalaren Funktion von $D[\hat{\theta}]$ übergehen.

Für Kleinstquadrateschätzer (LSE) mit normalverteilten Fehlern kann man die Konfidenzregionen als Ellipsoide im \mathcal{R}^k betrachten. Für den Fall $k = 2$ ergibt sich Abbildung 1.

Es zeigt sich, daß einige Funktionen von $D_1[\hat{\theta}]$ unter Zuhilfenahme von Abb.1 vernünftig geometrisch interpretiert werden können. Sie werden deshalb in der Literatur am häufigsten als Optimalitätskriterien verwendet:

[1] 'max $|D[\hat{\theta}]^{-1}|$ ' — D-Optimalität ist aufgrund etlicher günstiger mathematischer Eigenschaften (z.b.) Skalierungsinvarianz das gebräuchlichste Kriterium. Es entspricht geometrisch der Minimierung des Volumens (der Fläche) des Konzentrationsellipsoids. (Equivalent dazu ist das Kriterium 'min $\ln|D[\hat{\theta}]|$ ').

[2] 'min $tr D[\hat{\theta}]$ ' — A-Optimalität entspricht der Minimierung der Varianzen der einzelnen Parameterschätzungen. Geometrisch ist es das Quadrat

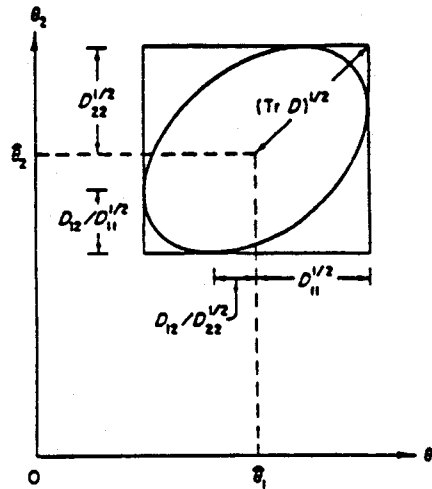


Abbildung 1: Konzentrationsellipsoid (aus Fedorov [1972])

des Abstandes vom Mittelpunkt des Ellipsoids zu den Eckpunkten des umschließenden Quaders (Rechtecks).

Ein wichtiger Zusammenhang zwischen diesen beiden Kriterien enthüllt sich beim Betrachten der Eigenwerte $\lambda_1, \dots, \lambda_k$ von $D[\hat{\theta}]$. Es gilt nämlich $|D[\hat{\theta}]| = \lambda_1 \cdot \dots \cdot \lambda_k$ und $\text{tr} D[\hat{\theta}] = \lambda_1 + \dots + \lambda_k$.

- [3] 'min max_i λ_i ' — E-Optimalität Hier wird der maximale Eigenwert der Matrix $D[\hat{\theta}]$ minimiert (geometrisch gesehen die längste Halbachse des Konzentrationsellipsoids). Man kann dieses Kriterium auch als 'min max_c $\text{Var}[c'\hat{\theta}]$ ' gegeben $c'c = 1$ also als die Minimierung der Varianz des am schlechtesten geschätzten Kontrastes $c'\hat{\theta}$ auffassen.
- [4] 'min max_i $D_{ii}[\hat{\theta}]$ ' — Maximale Varianz Optimalität ist ein einleuchtendes, wenn auch eher schwaches Kriterium. Auch hier wird, wie in Abbildung 1 ersichtlich, in gewisser Weise der Umfang des Ellipsoids minimiert ($D_{ii}[\hat{\theta}]$ bezeichnet das i-te Diagonalelement von $D[\hat{\theta}]$).

Alle diese Kriterien lassen sich relativ einfach zu einer für den Anwender attraktiveren Form verallgemeinern. Nimmt man an, daß a priori nicht alle Parameter von gleichem Interesse sind, so läßt sich mittels einer Gewichtung durch eine sogenannte Interessensmatrix P folgende Transformation durchführen:

$$f(D[\hat{\theta}]) = PD[\hat{\theta}]P' \quad (3)$$

um sodann die Kriterien (1) - (4) statt auf $D[\hat{\theta}]$ auf $f(D[\hat{\theta}])$ anwenden zu können.

Optimalität von Versuchsplänen läßt sich selbstverständlich nicht nur über optimale Parameterschätzung sondern auch durch die Präzision der Schätzungen \hat{y} , der 'response'-Werte definieren. Dies führt zu zwei weiteren natürlichen Kriterien. Die Varianz von \hat{y} sei definiert durch $d(x, \xi)$, dann nennt man:

[5] 'min $\max_{x \in \mathcal{X}_0} d(x, \xi)$ ' — G-Optimalität (für den Fall $\mathcal{X}_0 = \mathcal{X}$) Ist $\mathcal{X}_0 \in \mathcal{X}$ so spricht man vom Problem der Extrapolation.

[6] 'min $\int_{\mathcal{X}_0} w(x) d(x, \xi) dx$ ' — α -Optimalität

Gute und kurze Überblicke über Optimalitätskriterien geben zum Beispiel Fedorov [1972] oder Papakyriazis [1978].

3.3 Vorbedingungen

Will jemand Methoden der optimalen Versuchsplanung anwenden, so hat er sich ganz zu Beginn die Frage zu stellen:

Welches mathematische Modell bietet eine wenigstens annähernd gute Beschreibung des zu beobachtenden Prozesses ?

Solch ein (stochastisches) Modell kann formal wie folgt definiert sein:

$$y_i = \eta(x_i, \theta) + \epsilon_i \quad \text{mit} \quad E[\epsilon_i] = 0 \quad (4)$$

oder

$$E[y/x] = \eta(x, \theta) \quad (5)$$

wobei $y = (y_1, \dots, y_i, \dots, y_n)$ ein Vektor von Beobachtungen an den n Meßpunkten $x_1, \dots, x_i, \dots, x_n$ und $\eta(x, \theta)$ ist die a-priori gegebene Struktur genannt 'response - function' (Wirkungsfunktion).

Die Beobachtungen sollten unkorreliert sein:

$$Cov[y/x] = \sigma^2 \lambda^{-1}(x) I_n \quad (6)$$

wobei $\lambda(x)$ die sogenannte Effektivitätsfunktion, die mögliche Heteroskedastizität der experimentellen Situation widerspiegelt.

Für eine genauere analytische Behandlung muß meist durch einschränkende Annahmen über die Struktur von η das Modell vereinfacht werden.

3.3.1 der klassische Regressionsansatz

In den letzten Jahren ist vor allem dem klassischen linearen polynomialen Regressionsproblem viel Zeit und Platz gewidmet worden. Oft ist die Versuchsregion ein Würfel oder eine Kugel und das Kriterium ist D-Optimalität [1].

Man betrachte die Parametrisierung:

$$E[y/x] = \eta(x, \theta) = F'\theta \quad (7)$$

wobei F eine $k \times n$ Regressorvariablenmatrix von der Form

$$\begin{pmatrix} f_1(x_1), & \dots, & f_1(x_n) \\ \vdots & \dots & \vdots \\ f_k(x_1), & \dots, & f_k(x_n) \end{pmatrix}$$

ist. Die Wahl der Meßpunkte x_1, \dots, x_n ist durch die Versuchsregion \mathcal{X} beschränkt.

Der beste lineare unverzerrte Schätzer (BLUE) [zugleich der Kleinstquadrateschätzer (LSE)] von θ ist bekanntlich:

$$\hat{\theta} = n^{-1}M(\xi)^{-1}Fy \quad (8)$$

im stetigen (approximativen) Fall mit $f(x)' = (f_1(x), \dots, f_k(x))$ und der normalisierten Fisher'schen Informationsmatrix $M(\xi) = \int_{\mathcal{X}} f(x)f(x)'\xi(x)dx$ oder $M(\xi(n)) = \sum_{i=1}^n p_i f(x_i)f(x_i)'$ im diskreten (exakten) Fall.

Für die Kovarianzmatrix dieses Schätzers ergibt sich (im Fall $\lambda^{-1}(x) = 1$):

$$D[\hat{\theta}] = \sigma^2 n^{-1}M(\xi)^{-1} \quad (9)$$

Beispiel: Unterschiede in Optimalitätskriterien bei linearen polynomialen Regressionsexperimenten

Man kann jetzt anhand eines von Fedorov [1972] entwickelten Beispiels zeigen, daß ein Experiment \mathcal{E}_1 einem anderen \mathcal{E}_2 bezüglich eines Optimalitätskriteriums vorgezogen wird, bezüglich eines anderen jedoch zurück bleibt.

Es handle sich um eine eindimensionale lineare polynomiale Regression zweiter Ordnung, also:

$$E[y_i] = \theta_1 + \theta_2 x_i + \theta_3 x_i^2 \quad (10)$$

und es gebe zwei Experimente mit der gleichen Anzahl von Beobachtungen $N = 12$ mit den Versuchsplänen:

$$\xi_1(3) = \left\{ \begin{array}{ccc} x_1 = -1, & x_2 = 0, & x_3 = 1 \\ p_1 = 4/12, & p_2 = 4/12, & p_3 = 4/12 \end{array} \right\},$$
$$\xi_2(3) = \left\{ \begin{array}{ccc} x_1 = -1, & x_2 = 0, & x_3 = 1 \\ p_1 = 3/12, & p_2 = 6/12, & p_3 = 3/12 \end{array} \right\}.$$

Mit der Definition (9) und unter der Annahme $\sigma = 1$ erhält man:

$$D_1[\hat{\theta}] = 1/12 \begin{pmatrix} 3 & 0 & -3 \\ 0 & 3/2 & 0 \\ -3 & 0 & 9/2 \end{pmatrix},$$
$$D_2[\hat{\theta}] = 1/12 \begin{pmatrix} 2 & 0 & -2 \\ 0 & 2 & 0 \\ -2 & 0 & 4 \end{pmatrix}.$$

Daraus errechnet sich:

$$\{|D_1[\hat{\theta}|\} = 6.75 \cdot (12^3)\} < \{|D_2[\hat{\theta}|\} = 8 \cdot (12^{-3})\},$$

somit hat man gemäß [1] Versuchsplan ξ_1 zu favorisieren, andererseits gilt aber auch:

$$\{tr D_1[\hat{\theta}] = 3/4\} > \{tr D_2[\hat{\theta}] = 2/3\},$$

was nach Kriterium [2] den Vorzug für ξ_2 bedeutet.

3.3.2 Bayesianisches Design für Bayesianische Regression

Mit der wachsenden Akzeptanz bayesianischer Lösungen in der statistischen Literatur ging die Entwicklung einer bayesianisch orientierten Theorie der Versuchsplanung einher. Hauptzielpunkt war natürlich die Implementation von a priori Information in den Planungsprozeß des Experiments. Vor allem die Idee einer stufenweise Anwendung von Versuchsplänen mit einhergehender sequentieller Verbesserung der Vorinformation schien besonders attraktiv.

Anstelle des Kleinstquadrateschätzers $\hat{\theta}$ verwendet man den linearen Bayesschätzer:

$$\hat{\theta}_B = (nM(\xi) + \sigma_0^2 D_0^{-1})^{-1} (F'y + \sigma_0^2 D_0^{-1} \theta_0) \quad (11)$$

mit den gleichen Annahmen wie bei (8) allerdings mit dem a priori Wissen:

$$E[\theta/\sigma] = E[\theta] = \theta_0, \quad Cov[\theta] = D_0, \quad E[\sigma^2] = \sigma_0^2 \quad (12)$$

Es werden also keine Verteilungsannahmen getroffen. Die Erfordernisse sind nur approximative Kenntnis der ersten und zweiten Momente des Parameters θ und der erwarteten Beobachtungsvarianz.

Somit ergibt sich die Bayes'sche Varianzkovarianz Matrix:

$$D_B[\hat{\theta}] = n^{-1} \sigma^2 (M(\xi) + n^{-1} \sigma_0^2 D_0^{-1})^{-1} \quad (13)$$

Im Prinzip kann nun wieder jedes der in [1] - [6] angeführten Optimalitätskriterien zur Bestimmung des Versuchsplans verwendet werden, Chaloner [1984] behandelt jedoch ein Kriterium, das der bayesianischen Philosophie (Inkorporation von Vorinformation wann immer möglich) besser entgegenkommt.

Ziel des Experiments sei die Schätzung einer Klasse von Funktionen $c'\theta$, mit einem auf c definiertem Wahrscheinlichkeitsmaß μ . Nun soll das erwartete präposteriore Risiko minimiert werden, also:

$$[7] \text{ 'min } tr \Psi \cdot (nM(\xi) + \sigma_0^2 D_0^{-1})^{-1} \text{ ' — Bayes-A-Optimalität mit } \Psi = E_\mu[cc'].$$

Einen zusammenfassenden Überblick über bayesianische Methoden der Versuchsplanung geben Bandemer et al. [1987].

3.3.3 das Nichtlinearitätsproblem

Trifft man keine beschränkenden Annahmen über die Struktur von $\eta(\theta)$, kann der Fall auftreten, daß der optimale Versuchsplan von den 'wahren' Werten der zu schätzenden Parameter abhängt (konkret im Fall nichtlinearer Terme in θ).

Ein Ausweg aus diesem Paradoxon besteht darin, diese Terme durch eine Taylorreihe um einen a-priori Schätzwert θ_0 zu approximieren.

Ein weit attraktiverer und vielversprechenderer Ansatz entstammt aber auch hier der Bayes-schule. Chaloner [1987] schlägt eine stufenweise sequentielle bayesianische Designmethode vor. Diese vereinigt sowohl theoretische (Verwendung der bestmöglichen Information über θ) als auch praktische Vorteile (Informationsgewinn durch neue Daten).

3.3.4 Zeitreihenmodelle

Der für die Ökonometrie wohl bedeutsamsten Modellsparte wurde wohl aus Komplexitätsgründen bisher die geringste Aufmerksamkeit geschenkt. Papakyriazis [1978] betrachtet in einem Grundlagenartikel einige Spezialfälle eines 'distributed lag' Modells:

$$y_i = \sum_{i=1}^n \theta_i(L)x_{it} + \epsilon_{it} \quad (14)$$

wobei $x_{it} = (\omega_i(L)/\delta(L))b_{it}$ und $\epsilon = (\beta(L)/\psi(L))a_t$ entsprechen. $\delta(L)$, $\beta(L)$, $\psi(L)$, $\omega(L)$ und $\theta_i(L)$ bezeichnen Polynome in L (dem Lagoperator definiert als $L^j x_{it} = x_{i(t-j)}$). a_t und b_{it} stehen für unabhängiges weißes Rauschen.

Papakyriazis [1978] weist auf die große Generalität dieses Modells hin, da es praktisch alle relevanten Zeitreihenmodelle beinhaltet.

Eine weitere, sehr ausführliche Behandlung des Themas findet sich in Cambanis [1985]. Er betrachtet drei Designprobleme im Rahmen von Zeitreihenproblemen: nämlich für die Schätzung von Integralen von Zufallsgrößen, für die Schätzung von Regressionskoeffizienten von Zeittrends und das Aufspüren von 'signals in noise'. Zur Behandlung der Problematik schlägt er sowohl Zufallsdesigns als auch deterministische Versuchspläne vor.

$\Phi(M)$	$\varphi(x, \xi) = \text{tr}\{\partial\Phi/\partial M^{-1}\}M^{-1} - \lambda(x)f(x)'M^{-1}\{\partial\Phi/\partial M^{-1}\}M^{-1}f(x)$
$\ln M^{-1} $	$k - \text{tr}M^{-1}f(x)$
$\ln P'M^{-1}P $	$\text{rg}P - \text{tr}M^{-1}P(P'M^{-1}P)^{-1}P'M^{-1}f(x)$
$\text{tr}P'M^{-1}P$	$\text{tr}P'M^{-1}P - \text{tr}M^{-1}PP'M^{-1}f(x)$

Tabelle 1: Optimalitätskriterien mit dualen Funktionen.

3.4 Der Kern der Versuchplanungstheorie

Eine der Hauptkonsequenzen der approximativen Theorie ist die Möglichkeit Zusammenhänge zwischen verschiedenen Optimalitätskriterien herzustellen. Man spricht auch davon das sogenannte 'duale' Problem zu einem bestimmten Kriterium zu finden. Federführend in diesem Bereich waren Kiefer & Wolfowitz [1960] mit der Veröffentlichung ihres gefeierten generellen Äquivalenztheorems, das für ein optimales Design ξ^* die Gleichartigkeit der folgenden drei Eigenschaften feststellte:

- (i) ξ^* ist D-optimal
- (ii) ξ^* ist G-optimal
- (iii) $\max_{x \in \mathcal{X}} d(x, \xi^*) = k$: Anzahl der linear unabhängigen Parameter

Ein allgemeines Resultat für alle konvexen Optimalitätskriterien gibt Fedorov [1980]. Daraus kann Tabelle (1) mit den gebräuchlichsten Kriterien Φ und ihrem jeweiligen dualen Kriterium φ bestimmt werden.

Angesichts der Attraktivität einer solchen Betrachtungsweise sollte aber nicht vergessen werden, daß ihre Anwendung auf approximative Designs (natürlich auch asymptotisch exakte Versuchspläne) beschränkt bleibt. Exakte Designs müssen kombinatorisch gelöst werden.

Übersichten zu dieser Problemstellung finden sich in Fedorov [1972], Silvey [1980] sowie mit Blickrichtung Ökonomie in Aigner [1979].

Aufgrund der Äquivalenzen in der approximativen Theorie können hier Optimierungsalgorithmen eingesetzt werden. Sie basieren auf dem Prinzip, daß man einen suboptimalen Versuchsplan durch Hinzufügen oder Entfernen

von Gewichten zu bestimmten Punkten x_s der Versuchsregion \mathcal{X} verbessern kann.

Auf jeder Stufe eines Algorithmus wird also solch ein Schritt durchgeführt. Die Lage des Punktes x_s (der Meßstelle) wird durch das Maximum oder Minimum der dualen Funktion $\varphi(x, \xi)$ (z.B. aus Tabelle 1) bestimmt.

Zwei typische Algorithmen sollen hier kurz beschrieben werden:

- eine iterative Prozedur erster Ordnung
- ein 'exchange - type' Algorithmus

— Ersterer wird in der folgenden Art ausgeführt:

$$\xi_{s+1} = (1 - \alpha_s)\xi_s + \alpha_s\xi(x_{s+1}) \quad (15)$$

wobei $\xi(x_{s+1})$ ein Design bezeichnet, daß sein ganzes Gewicht an der Stelle:

$$x_{s+1} = x_{s+1}^+ = \underset{x}{\operatorname{argmin}} \varphi(x, \xi_s) \quad \forall x \in \mathcal{X} \quad (16)$$

konzentriert (im Falle der D-optimalität also am Punkt mit der größten Varianz).

Die gleiche iterative Vorgangsweise und die gleiche Idee liegt der Reduktion von Gewichten an 'schlechten' Stellen des Versuchsplanes zugrunde (dies nennt man 'backward' (also rückwärts) Prozedur). Hier ist:

$$x_{s+1} = x_{s+1}^- = \underset{x}{\operatorname{argmax}} \varphi(x, \xi_s) \quad \forall x \in \mathcal{X}_s \quad (17)$$

und

$$\alpha_s = \begin{cases} -\gamma_s & p(x_s^-) \geq \gamma_s \\ -p(x_{s+1}^-)/(1 - p(x_{s+1}^-)) & p(x_s^-) < \gamma_s \end{cases} \quad (18)$$

\mathcal{X}_s steht für die Menge der Punkte, die zum Design ξ_s gehören, $p(x_s)$ ist das Gewicht von Punkt x_s im Versuchsplan ξ_s .

Es gibt mehrere Möglichkeiten für die Wahl von γ_s . Zwei davon werden hier verwendet:

$$\gamma_s = 1/(n_0 + s) \quad (n_0 \text{ ist die Zahl der Meßstellen eines Initialversuchsplans } \xi_0)$$

$\gamma_s = c_0$ (c_0 ist eine kleine zu wählende Konstante)

Der Nachteil der letzteren Definition liegt in der Nichterfüllung der üblichen Konvergenzbedingungen:

$$\lim_{s \rightarrow \infty} \gamma_s = 0, \quad \sum_s \gamma_s = \infty, \quad \sum_s \gamma_s^2 < \infty \quad (19)$$

Dies kann aber durch die Bedingung: ist $\Phi_s > \Phi_{s-1}$ dann sei $c_0 = c_0/2$ erreicht werden.

Im Fall der Vorwärtsprozedur (16) wird $\alpha_s = \gamma_s$ gesetzt.

— Beim Auswechsel('exchange')-algorithmus wurden einige Berechnungsschritte vereinfacht. Zum Beispiel ist es jetzt nicht mehr notwendig, den Versuchsplan auf jeder Stufe neu zu berechnen sondern nur, wenn Meßpunkte hinzu- oder wegkommen. Dieser Algorithmus hat die folgende Struktur:

Kann man die Versuchsregion \mathcal{X} durch ein Gitter mit den Elementen Δ_s auf Stufe s repräsentieren, dann gehe man vor:

- [1] Es gibt einen Versuchsplan $\xi_{ns} = (x_{1s}, \dots, x_{ns})$, wobei die x_i 's irgendwelche Punkte (womöglich die Eckpunkte von \mathcal{X}) von Δ_s sind. Dann muß:

$$x_s^+ = \arg \max_{x \in \tilde{\Delta}_s} \varphi(x, \xi_{ns}), \quad \tilde{\Delta}_s = \Delta_s \text{supp} \xi_s \quad (20)$$

gefunden werden und das neue Design:

$$\xi_{(n+1)s} = (x_{1s}, \dots, x_{ns}, x_s^+)$$

wird konstruiert.

- [2] Punkt:

$$x_s^- = \arg \min_{x \in \text{supp} \xi_{(n+1)s}} \varphi(x, \xi_{(n+1)s}) \quad (21)$$

muß vom Versuchsplan $\xi_{(n+1)s}$ entfernt werden.

Man kann diese Prozedur entweder mit Vorwärtsschritten [1] oder Rückwärtsschritten [2] starten. Im zweiten Fall muß die Länge des Schrittes kleiner sein als $s - k$, wobei k die Anzahl der zu schätzenden Parameter bezeichnet.

Es ist notwendig anzuführen, daß es für die Konvergenz dieser Methode wichtig ist, daß $d(\Delta_s) \rightarrow 0$, aber für praktische Zwecke reicht es aus ein genügend kleines $d(\Delta_s) = d$ mit einem sehr dichten Gitter zu wählen (d steht für die Distanzen zwischen benachbarten Punkten aus Δ).

Für ökonomische Problemstellung bietet sich besonders der 'exchange-type' Algorithmus an, da man hier leicht die Beschränkung auf nur eine mögliche Beobachtung pro Zeitpunkt, wie bei ökonomischen Zeitreihen häufig der Fall, verwirklichen kann (siehe z.B. Fedorov [1989]).

4 Ein Anwendungsbeispiel: Income Maintenance Experiments (Unterhaltsversuche)

Diese Ende der sechziger, Anfang der siebziger Jahre in den Vereinigten Staaten gestartet und mit einer Laufzeit von bis zu zwanzig Jahren geplanten, also teilweise noch andauernden, Versuche waren die ersten ihrer Art. Niemals zuvor war in solch großem Maßstab sozioökonomisch experimentiert worden, was klarerweise zu völlig neuartigen und ungewohnten Situationen führte.

Das Ziel dieser Versuche war die empirisch Erforschung der Effekte von staatlichen Unterhaltszahlungen auf die individuelle Arbeitsbereitschaft, beziehungsweise auf das gesamte Arbeitsangebot am Markt.

Aus der mikroökonomischen Theorie ist bekannt, daß der Gesamteffekt von Unterhaltszahlungen auf das Arbeitsangebot sich aus Einkommens- und Substitutionseffekt zusammensetzt. Unter den üblichen Annahmen von Freizeit als normalem Gut und von konvexen Indifferenzkurven zwischen Freizeit und Arbeit, ist dieser Gesamteffekt, wegen negativem Einkommens- und Substitutionseffektes bei Preiserhöhung von Arbeit, negativ. Eine mögliche Inferiorität des Gutes Arbeit könnte jedoch zu positiven Gesamteffekten führen (Giffens Paradoxon, siehe Abbildung 2).

Zur Erforschung der Richtung dieses aus der Theorie nicht abschätzbaren Effektes, sowie der zugrundeliegenden Parameter ist ein empirischer Nachweis wünschenswert. Die hohen Kosten eines derartigen Projektes machen eine sorgfältige Versuchsplanung unerlässlich.

The total effect of an increase in the price of X is to increase the quantity of X demanded. This happens because the negative substitution effect (the movement from X^* , Y^* to point B) is outweighed by a strong positive income effect resulting from the inferiority of good X . Not every inferior good need exhibit Giffen's paradox.

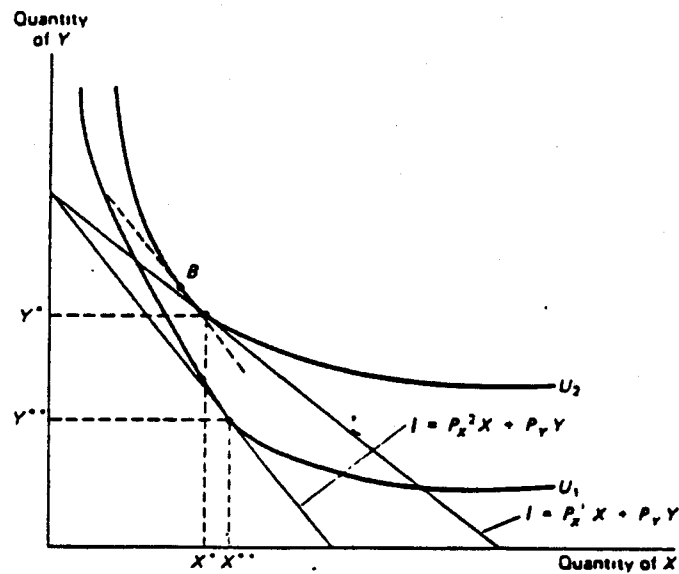


Abbildung 2: Giffens Paradoxon (aus Nicholson [1984])

Als Bewertungsvariable für die geschilderten Auswirkungen diene eine Variable R , die die Steigung respektive die Senkung des Haushaltseinkommens während des Experiments messen sollte:

$$R = \frac{\text{aktuelles Haushaltseinkommen}}{\text{Einkommen vor dem Versuchsbeginn}}$$

Den Versuchsfamilien wurden über die Dauer des Versuchs (mindestens drei Jahre) wöchentliche Zuschüsse z gewährt, die nach dem folgenden Schema errechnet wurden:

$$z = \alpha g - ty(w) \quad (22)$$

wobei α die Armutseinkommensgrenze beschreibt, g kann man den Unterstützungsgrad nennen, t den internen Steuersatz, der auf das Wocheneinkommen entoben und um den also der Maximalzuschuß verringert wurde.

Die Variable w (Einkommen vor Versuchbeginn / α) soll den Einfluß der unterschiedlichen Startbedingungen der verschiedenen Familien in die Analyse einbringen.

Somit hat die Variable R , die sich aus dem Arbeitseinkommen y und dem Zuschuß z zusammensetzt, die funktionale Form:

$$R = f(g, t, w) + \epsilon \quad (23)$$

Es ist offensichtlich, daß die Variablen g, t direkter Einflußnahme unterliegen, während hingegen w durch Schichtung von Stichproben kontrolliert werden muß.

Klarerweise wirken noch einige andere exogene Variable auf R . Sie können zum Beispiel g, t oder w beeinflussen (Abhängigkeit der Armutsgrenze von der Familiengröße, Abhängigkeit von w von einer Vielzahl von Variablen wie Alter, Rasse, Erziehung, etc.). Andere unkontrollierbare Einflußgrößen mit generell weit geringerem Gewicht wurden zwecks Simplifizierung dem Fehlerterm zugeschlagen.

Die Versuchsregion \mathcal{X} bestand sodann aus 27 zulässigen Kombinationen von g, t und w , welche bereits unter Bedachtnahme von Kostenüberlegungen aber auch der gewünschten großen Variabilität selektiert wurden (der Steuersatz wurde zum Beispiel auf die drei levels 30%, 50% und 70% gesetzt).

Wegen der Vorgabe der Lokation der Meßpunkte wurde das Problem auf die Aufgabe, die optimalen Gewichte p_1, \dots, p_{27} festzusetzen, reduziert. In der Praxis ist das Auffinden der Sample-größen N_1, \dots, N_{27} (Anzahl der Versuchsfamilien) für die jeweilige Variablenkombination dazu equivalent.

Dies ähnelt dem in 3 angeführten Designproblemen. Unter Zuhilfenahme einer konkreten Regressionsfunktion, die einer sehr sorgfältigen Auswahl unter mehreren Alternativen bedurfte, konnte nach der Wahl eines entsprechenden Optimalitätskriteriums das Optimierungsproblem formuliert werden (im konkreten Fall wurde eine erweiterte Version von A-optimalität [2] verwendet):

$$\min \operatorname{tr} \left\{ A \left(\sum_{i=1}^{27} N(i) f(x_i) f(x_i)' \right)^{-1} \right\} \quad (24)$$

Zahlreiche zusätzliche Einschränkungen mußten der Optimierung auferlegt werden. Besonders wichtig war in diesem Zusammenhang natürlich die Kostenschranke:

$$\sum_{i=1}^{27} c_i N_i \leq C \quad N_1 \geq 0, \dots, N_{27} \geq 0 \quad (25)$$

wobei C das Gesamtbudget bezeichnet und c_i die Kosten einer Beobachtung an den Regressoren x_i .

In einem ökonomischen Kontext entspricht die Fragestellung einem Nutzenmaximierungsproblem, wobei das Kriterium ein inverses Nutzenmaß darstellt, die N_i Gütermengen, c_i die Preise und C das verfügbare Einkommen beschreiben.

Konstruiert man die Gewichte $p = N_i/N$ so kann man das Problem zu:

$$\min 1/c \cdot \left(\sum_{i=1}^{27} p_i c_i \right) \cdot \operatorname{tr} \left\{ A \left(\sum_{i=1}^{27} p_i f(x_i) f(x_i)' \right)^{-1} \right\} \quad (26)$$

verkürzen.

Problem (26) kann nun entweder mit den in 4 geschilderten approximativen Verfahren oder mit Techniken der Programmierung (integer programming) gelöst werden. Zweiteres wurde letztendlich verwendet, um eine Lösung nach 45-minütiger Rechenzeit zu erhalten.

Hier soll angemerkt werden, daß ein Computerlauf selbstverständlich keine eindeutige Lösung bringen kann, daß im Gegenteil erst mit Hilfe einer Zwischenlösung eine Erweiterung der Modelle und eine Verfeinerung der Methoden erreicht wird.

Conlisk & Watts [1969] stellen aber doch gewisse Hauptcharakteristika der möglichen optimalen Versuchspläne für das 'Graduated Work Incentive Experiment' fest:

- (a) Nichtorthogonalität des Versuchsplans;
- (b) an etlichen der 27 zulässigen Kombinationen war keine Beobachtung notwendig;
- (c) der Großteil des Budgets wurde für wenige teure Variablenkombinationen verbraucht, während hingegen die meisten Messungen an kostengünstigen Kombinationen durchgeführt wurden;
- (d) der neugeschaffene Versuchsplan war in Bezug auf das Optimalitätskriterium wesentlich effizienter als alle vor Erarbeitung des Designs 'intuitiv' vorgeschlagenen Pläne;

5 Abschließende Bemerkungen

Innerhalb der geschilderten Problembereiche erhalten nun Ökonomen Ökonometriker die Möglichkeit zur ursprünglichen Mitwirkung am Reifen und an der Durchführung eines Projektes im Gegensatz zum üblichen konsultativen Eingreifen in den Spät- (Auswertungs)phasen einer Studie.

Die Problematik der optimalen Versuchsplanung ist, wie schon von Conlisk & Watts [1969] beschrieben, von einer in der Ökonomie bekannten Art:

[...] - how to get the most of some desirable output from limited inputs of financial and other resources, while observing various additional constraints.

In der Regel ist in diesem Fall der begrenzte Input die finanzierbare Anzahl der Messungen, und die Einschränkungen werden über die Versuchregion und über die Funktion λ aufgestellt.

Die Anwendung der Methodik in dem gewiß beschränkten Aufgabenbereich Ökonomie/Ökonometrie scheint im Rahmen der aufgezeigten Möglichkeiten deshalb doch richtig und wichtig zu sein. Insbesondere aber auch, weil in der wirtschaftlichen Anwendung der Hauptkritikpunkt an der optimalen Versuchsplanung, nämlich zu stark an der jeweiligen statistischen Struktur verhaftet zu sein, nicht trifft. Wirtschaftswissenschaftler sind traditionell sehr modellorientiert, sodaß eine Beschäftigung mit der Versuchsplanungstheorie keinen, wie auch immer gearteten methodischen Bruch bedeuten würde.

Danksagung

Mein Dank geht an V. Fedorov für viele wichtige Hinweise und Tips im methodischen Teil, sowie an T.Url für ökonomische Erläuterungen und Korrekturen.

6 Literaturliste

- [1] D.J. Aigner. A brief introduction to the methodology of optimal experimental design. *Journal of Econometrics*, 11:7–26, 1979.
- [2] H. Bandemer. Optimal experimental design for regression models. *Mathematische Operationsforschung und Statistik, Series Statistics*, 18(2):171–217, 1987.
- [3] S. Cambanis. *Handbook of Statistics*, chapter Sampling Design for Time Series, pages 337–362. Volume 5, Elsevier Science Publishers, 1985.
- [4] B. Castro and K. Weingarten. Toward experimental economics. *Journal of Political Economy*, 78(2):171–217, 1970.
- [5] K. Chaloner. *Optimal Bayesian Design for Nonlinear Estimation*. Technical Report, University of Minnesota, 1987.
- [6] K. Chaloner. Optimal bayesian experimental design for linear models. *The Annals of Statistics*, 12(1):283–300, 1984.
- [7] J. Conlisk and H. Watts. A model for optimizing experimental design for estimating response surface. *Proceedings of the American Statistical Association, Social Statistics Section*:150–156, 1969.
- [8] V.V. Fedorov. *Analysis and Design of Simulation Experiments for the Approximation of Models*. Technical Report WP-83-71, International Institute for Applied System Analysis, Laxenburg, 1983.
- [9] V.V. Fedorov. Convex design theory. *Mathematische Operationsforschung und Statistik, Series Statistics*, 11(3):403–413, 1980.
- [10] V.V. Fedorov. *Optimal Design for Experiments*. Academic Press, New York, 1972.
- [11] V.V. Fedorov. Optimal design with bounded density: optimization algorithms of the exchange type. *Journal of Statistical Planning and Inference*, 22:1–13, 1989.

- [12] R. Ferber and W.Z. Hirsch. Social experiments in economics. *Journal of Econometrics*, 11:77-115, 1979.
- [13] R.A. Fisher. *The Design of Experiments*. Oliver & Boyd Ltd., Edinburgh, 1935.
- [14] J. Kiefer. Optimal experimental designs (with discussion). *Journal of the Royal Statistical Society*, B:272-319, 1959.
- [15] J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 14:363-366, 1960.
- [16] J.P.C. Kleijnen. Experimentation with models: statistical design and analysis techniques. In F.E. Cellier, editor, *Progress in Modelling and Simulation*, pages 173-185, Academic Press, 1982.
- [17] T. Naylor. Computer simulation experiments with economic systems: the problem of experimental design. *Journal of the American Statistical Association*, LXII:1315-1337, 1967.
- [18] T. Naylor. Experimental economics revisited. *Journal of Political Economy*, 80:347-352, 1972.
- [19] W. Nicholson. *Microeconomic Theory*. The Dryden Press, third edition, 1984.
- [20] P.A. Papakyriazis. Optimal experimental design in econometrics. *Journal of Econometrics*, 7:351-372, 1978.
- [21] S.D. Silvey. *Optimal Design*. Chapman and Hall, London, 1980.