

Felix Schaefer:

## **Erfahrungen mit Prognosetests in der marktforscherischen Praxis**

---

Es gab in der jüngeren Vergangenheit eine ganze Reihe von Dokumentationen über die bemerkenswerte Genauigkeit von Prognosetests und auch meine Vorredner haben hierzu einmal mehr etliche Beispiele angefügt<sup>1</sup>

Wenn ich also heute über unsere Erfahrungen mit solchen Prognose-Tests sprechen möchte, dann geht es mir weniger darum in diesen Chor einzustimmen und weitere Erfolgs-Stories anzufügen, sondern genau das Gegenteil zu tun: ich möchte über ein paar sog. "böse" Erfahrungen berichten. Dies soll nicht geschehen, um die oft beschworene "Validität" solcher Tests zu erschüttern, sondern um zu zeigen, wie "natürlich" und folgerichtig die in diesen Tests verwendeten Modelle auf Störungen aller Art reagieren - und damit Erkenntnisse freisetzen.

Wenn alles immer so schön funktioniert, wie es ein Modell vorsieht, dann kann man sich zwar befriedigt zurücklehnen; Erfahrungen sammeln heißt aber oft "böse Erfahrungen" sammeln. Und in diesem Sinne sollen die folgenden Beispiele dienen.

### **1. Beispiel: Unterschiedliche Prognose-Ergebnisse**

Unsere Erfahrungen beruhen auf dem Prognose-Modell des Typs "Assessor", wie es in der Mai-Ausgabe des Journal of Marketing Research des Jahres 1978 von Alwin J. Silk und Glen L. Urban beschrieben worden war. Diese beiden Autoren fügten eine ganze Reihe bekannter Modell-Elemente auf eine äußerst originäre und intelligente Weise zusammen. Dies waren

- o ein Kauf-/Wiederkauf-Modell eher "klassischen" Zuschnitts, beruhend auf Überlegungen von Parfitt & Collins (1968)
- o sowie ein Präferenz-Modell, beruhend auf den Überlegungen von Axelrod (1968), Luce (1968), Allaire (1973) und Pessemier (1975).

Das Wesentliche dieser "Assessor"-Idee war, daß das daraus erwachsene Test-Modell zwei Prognosen liefert, die auf unterschiedlichem Wege zustande kommen. Liefern beide Modelle dasselbe, hat man das sichere Gefühl, daß man sich auf das Ergebnis verlassen kann. Stimmen die Ergebnisse beider Modelle nicht überein, ist man gezwungen, darüber nachzudenken, was denn nun, um Himmelswillen, passiert sein könnte. Schon Silk und Urban haben dies in ihrem 78er-Aufsatz deutlich unterstrichen:

*"Concergent results should strengthen confidence in the prediction whereas divergent outcomes signal the need for further analyses to identify sources of discrepancies and to provide bases for reconciliation ..."*

1983 haben Urban und Katz über die Spannweite solcher Unterschiede berichtet, dies aber in erster Linie benutzt, um zu zeigen, wie 'valide' die von einem solchen Modell zu erwartenden Ergebnisse sind.

---

<sup>1</sup>Vortrag, gehalten auf dem 40. Kongress der Deutschen Marktforschung, Berlin 1995

Ich möchte hier ein solches Beispiel mit unterschiedlichem Prognose-Ergebnis anfügen, um zu zeigen, daß man die "sources of discrepancies" aufdecken kann, weil das Modell einem "inneren System" folgt mit generellen Strukturen oder, wie es Prof. Ehrenberg in seinem vorangegangenen Vortrag formulierte, mit "lawlike relationships", die äußerst empfindlich auf Störungen aller Art reagieren.

Das Kauf-/Wiederkauf-Modell lieferte folgendes Ergebnis:

Erstkaufrate im Studio	25 %
x "Markenbekanntheit"	77 %
x Distribution	70 %
-----	
= Penetration	13 %
x Wiederkaufrate	36 %
-----	
= Prognose	<b>4.9 %</b>

Das Präferenzmodell hingegen lieferte dagegen einen 1.5 % höheren zu erwartenden Käufer-Anteil:

"Markenbekanntheit"	77 %
"Aktualität"	27 %
-----	
=> Wahrscheinlichkeit auf Aufnahme in den "evoked set"	27 %
x Kauf-Wahrscheinlichkeit aus Präferenz-Noten	24 %
-----	
= Prognose	<b>6.5 %</b>

Was war geschehen?

Unser mobiles Teststudio stand auf dem Parkplatz eines großen Verbrauchermarktes, in dem zufälligerweise bei Testbeginn eine Sonderaktion für genau die Marke lief, an deren Marktbedeutung sich unser Kunde bezüglich der zu erzielenden Markenbekanntheit (s.o.: 77 %) orientieren wollte.

Es zeigte sich nun, daß die Werbung für diese Referenz-Marke wirkte und zwar - bei uns im Test.

Schon die Interviewer hatten berichtet, daß diese Marke in unserem Teststudio in besonderem Maße "gekauft" wurde. Bei der Analyse der Daten zeigten sich dann weitere Wirkungen:

- Die regressionsanalytische Betrachtung der Beziehung zwischen Bekanntheit der Marken (gestützt) und der Wahrscheinlichkeit, daß diese Eingang in den "relevant set" der Verbraucher finden, zeigte,

- daß die Referenz-Marke **über** der Kurve lag, die den erwarteten Verlauf charakterisierte - mit einem Plus von, relativ gesehen 25 %;
- die Beziehung zwischen Bekanntheit und Aktualität (= ungestützte Bekanntheit) lieferte statt der erwarteten 27 % gar 43 %, - ein Plus von 60 %
- und bezogen auf die Werbe-Erinnerung ergab sich gar ein Plus von rd. 100 %!

Die 'Kaufrate' für diese Referenz-Marke war davon unmittelbar betroffen. Bekannt war, daß sie einen Käufer-Anteil von ca. 10 % haben mußte. Aber 15 % der Testteilnehmerinnen gaben an, sie "zuletzt" gekauft zu haben und die Kaufrate im Studio lag dann gar bei 36 %.

Dies mußte sich klar zu Ungunsten der neuen Marke und damit in einem zu niedrigen Wert für diese (wie auch für alle anderen angebotenen Marken) niederschlagen.

Auf der anderen Seite hatte sich dieser 'run' auf die Referenz-Marke noch nicht in der grundsätzlichen Einstellung gegenüber diese Marke niedergeschlagen. Die Wahrscheinlichkeit der Aufnahme in den "relevant set" war zwar gestiegen, allerdings noch nicht so, wie man dies aufgrund der erhöhten Aktualisierung erwarten durfte.

Und auch der Präferenzbildungsprozeß hatte noch nicht zu grundsätzlichen Verschiebungen geführt, weil ja zunächst einmal nur verstärkt Probierkäufe stattgefunden hatten. Die Attraktivität der Marke mußte sich erst noch in der folgenden Verwendungs-Situation beweisen.

Die Folge war: von dieser Sonderaktion war zwar die Erstkauftrate für die neue Marke betroffen - und damit das Ergebnis des Kauf-/Wiederkauf-Modells - nicht aber das Präferenz-Modell.

In Absprache mit dem Auftraggeber haben wir daraufhin die Erstkaufraten korrigiert: Der ermittelte Anteil für die Referenz-Marke wurde auf das Normalmaß zurückgeführt und die Erstkaufraten aller anderen angebotenen Marken wurden proportional zu diesem Anteil entsprechend angehoben. Und siehe da, nun stimmen beide Prognosen fast peinlich genau überein:

<b>a) Kauf-Wiederkauf-Modell</b>	
<i>"justiert":</i>	
Erstkauftrate im Studio	33 % (statt 25 %)
x Markenbekanntheit	77 %
x Distribution	70 %
= Penetration	18 %
x Wiederkauftrate"	36 %
= Prognose	<b>6.5 %</b>
-----	
<b>b) Präferenz-Modell</b>	<b>6.5 %</b>

## 2. Beispiel: "USP" falsch betont

In einem anderen europäischen Land sollte ein Vollwaschmittel eingeführt werden, das als USP eine "eingebaute" Weichspüler-Eigenschaft auswies.

Produkttests waren durchgeführt worden, die zeigten, daß die Produktleistung stimmte und ebenfalls durchgeführte Packungstests und Werbemitteltests ließen das ganze Paket als 'rund' erscheinen und auf einen Erfolg hoffen.

Der durchgeführte Prognose-Test zeigte (einmal mehr), daß die Summe aller Einzelteile nicht das Ganze zeigen muß:

- Schon die im Test erzielte Erstkauftrate lag mit 20 % so weit unter den Erwartungen, daß der Auftraggeber an einen Abbruch des Tests dachte. Man hatte diesen Wert mit der zu erwartenden Markenbekanntheit und Distribution und einer durchschnittlichen Wiederkauftrate multipliziert und war damit zu einer als unzureichend befundenen Marktanteils-Schätzung gekommen.

Wie Höger (1987) in seinem bemerkenswerten Aufsatz dargestellt hat, kann man solche rechnerischen Prognosen heute dank umfangreichen Daten(-Bank)-Materials noch wesentlich verfeinern; aber vermutlich wäre damit letztendlich kein beruhigendes Ergebnis herausgekommen.

Der Test wurde aber trotz alledem zu Ende geführt, u.a. auch deshalb, um herauszufinden, ob man mit allen anderen Teil-Testergebnissen ebenso schief gelegen haben könnte.

Und in der Tat, die Bewertung des Produktes nach der Probierphase des Prognosetests bestätigte die Produkttest-Ergebnisse und zeigte, daß das Waschmittel alle wichtigen Eigenschaften in deutlich mehr als zufriedenstellender Weise besaß.

Es zeigte sich aber auch eine deutliche Diskrepanz zwischen den Erwartungen an das Waschmittel und diesen tatsächlich vorhandenen und die Wiederkauftrate deutlich positiv beeinflussenden Eigenschaften.

Was war geschehen? Die Einführungswerbung zeigte ein junges Mädchen in einer Straßenbahn, bekleidet mit einem luftigen Sommerkleid, das von einem jungen Mann heftig angeflirtet wurde.

Off-Text:

*"NEU" wäscht fabelhaft sauber und pflegt fabelhaft weich.  
Du fühlst Dich wohl, man sieht Dir's an - die Sauberkeit  
zieht Blicke an ...*

80 % erinnerten sich, in der nachfolgenden Befragung, den Werbespot gesehen zu haben. Die höchsten Bewertungen erzielte der Film bezüglich:

- ist besonders auffällig
- ganz anders als herkömmliche Werbung
- ist besonders sympathisch.

In besonderem Maße wurde vermittelt:

- die Wäsche fühlt sich angenehm an
- macht die Farben leuchtend
- macht die Wäsche besonders weich
- macht die Wäsche wieder frisch
- gibt der Wäsche einen besonderen Duft.

Die "Uniqueness" des Produktes war also vermittelt worden und zwar auf neuartige und anmutende Art und Weise. Nur, es wurde vergessen, die für 'Hausfrauen' entscheidenden Eigenschaften eines Vollwaschmittels (und ein solches sollte es ja trotz seines USP's weiterhin sein) zu vermitteln:

- wäscht das Waschmittel überhaupt sauber?
- ist es für alle Waschtemperaturen geeignet?
- löst es hartnäckigen Schmutz?
- macht es strahlend weiß?

Diese Einführungswerbung bildete damit ein typisches Beispiel für *'Misdirection'*, wie es Rosser Reves der 'Erfinder' des USP in seinem bemerkenswerten Interview in Advertising Age (April '65) als eine typische Todsünde vieler interessanter, lustiger, ansprechender - und vermutlich preisgekrönter Werbespots, Anzeigen oder Plakate genannt hat.

Das hier gefundene Ergebnis zeigt damit u.a., daß man mit dem 'halben' Prognose-Test oder auch vielen der üblichen Werbemitteltests zu kurz springen kann. Erst im Zusammenhang mit dem Produkt und dessen Produkteigenschaften und -leistungen zeigt sich, ob Werbung wirkt.

Prof. Ehrenberg, der gestern zu 'lawlike relationships' sprach, hat auch dies schon an anderer Stelle einmal deutlich gemacht.

### **Das 3. Beispiel beschäftigt sich mit "überschätzten Erwartungen"**

Unser Kunde hatte auf einem anderen europäischen Markt mit zwei Zahncreme-Marken eine marktbeherrschende Stellung mit ca. 65 % Marktanteil. Diese Position sollte durch die Einführung einer dritten Marke zwecks Besetzung eines weiteren Segments gefestigt werden.

Die Hauptfragestellung des Prognose-Tests war damit, herauszufinden, welcher Käuferanteil zu erwarten war. Genau so wichtig war aber, der Frage nachzugehen, wie stark die Kannibalisierung sein würde, d.h. zu ermitteln, was "netto" hinzu kommen würde.

Aus der Erfahrung mit hunderten von Untersuchungen, bei denen wir die Überschneidungen von Marken-Verwender und -Anhängerschaften gemessen haben, konnte man das Ergebnis in allgemeiner Weise "vorhersehen":

- ⇒ Wenn der Hersteller zwei Drittel des Marktes hat,
- ⇒ werden ca. 2/3 der Käuferschaft der neuen Marke von den eigenen Marken kommen.

Und hätten wir damals schon das Dirichlet-Modell von Andrew Ehrenberg näher gekannt, hätten wir dies vermutlich im Detail berechnen - und dem Kunden die Durchführung dieses aufwendigen Prognose-Tests ersparen können. Denn dieser zeigte: es würde noch "schlimmer" kommen als vorher geschätzt; nicht zwei Drittel sondern rd. 80 % des zu erwartenden Marktanteils.

teilen wurden von der Verwenderschaft der bisherigen Marken des Kunden kommen. Beide Prognosen stimmten dabei in auffälliger Weise überein:

a) Kauf-Wiederkauf-Modell:		
	<b>Verwender der zwei Marken</b>	<b>Nicht Ver- wender</b>
Penetration	27 %	14 %
x Retention	46 %	34 %
=	13 %	5 %
x Verwender-Anteil	65 %	35 %
= Prognose	<b>8.2 %</b>	<b>1.7 %</b>

b) Präferenz-Modell:		
	<b>Verwender der zwei Marken</b>	<b>Nicht Ver- wender</b>
Kaufwahrscheinlichkeit	34 %	28 %
x Wahrscheinlichkeit der Aufnahme in "evoked set"	36 %	16 %
=	12.5 %	4.5 %
x Verwender-Anteil	65 %	35 %
= Prognose	<b>8.1 %</b>	<b>1.6 %</b>

Woran lag es nun, daß die bisherigen "Nichtverwender" nicht in einem größeren Maße für dieses neue Produkt zu begeistern gewesen waren? Da waren zunächst einmal die Erwartungen an das Produkt gewesen; sie waren sehr hoch, was sich in der Erstkauftrate von ca. 35 % niederschlug.

Dann zeigte sich aber bei der Bewertung des Produktes, daß diese Erwartungen, besonders bezüglich des neuartigen, auf Parodontose-Wirkung zielenden Charakters der Zahncreme nicht in befriedigendem Umfang nachvollzogen werden konnten.

Die Wiederkauftrate fiel damit deutlich zurück und besonders bisherige 'Nichtverwender' sahen nicht ein, warum sie von "ihren" Zahncremes zu diesem neuen Angebot wechseln sollten.

## Fazit

Das den hier beschriebenen Fällen zugrunde liegende Modell vereinigt eine Reihe grundlegender Erkenntnisse:

1. die Beziehungen zwischen Erst- und Wiederkauf
2. die Vorstellung des "relevant set" von Marken
3. die Bildung von Präferenzen als Meßinstrument für Kaufwahrscheinlichkeiten.

Es berücksichtigt ferner die Beziehung einer Reihe von Meßwerten untereinander, wie

- "Marken-Bekanntheit"
- "Aktualität"
- "Attraktivität"
- Marken-Verwendung & Kauf;

ebenso wie die Kenntnis der "Überschneidungen" von Markenverwenderschaften.

Die Chance, mit einem solchen Testmodell eine verlässliche Marktprognose zu erhalten, ist hoch. Findet man Abweichungen vom Erwarteten, lassen diese wichtige Rückschlüsse auf die Marktbedingungen zu.

In diesem Sinne wollten wir mit unseren Beispielen den Kreis schließen von "Gesetzmäßigkeiten oder lawlike relationships" über "chaotische Zustände", wie sie in verschiedenen, uns hier über 2 Tage lang vorgetragenen Bereichen herrschen, zu den praktischen Auswirkungen solcher Erkenntnisse für unsere tägliche Arbeit als Marktforscher.

\* \* \*

## Literatur:

ARF Midwest Workshop (1988):	Pre Test Market Research
Allaire (1973):	The Measurement of Heterogeneous Semantic, Perceptual, and Preference Structures, MIT-Dissertation
Axelrod (1968):	Attitude Measures that Predict Purchase, JAR
Blanchard & Harding (1983):	ASSESSOR - 10 Years of Application, ESOMAR-Seminar
Ehrenberg (1969):	The Discovery and Use of Laws of Marketing, JAR
Höger (1987):	Detector - Marktanteilsprognosen mit Hilfe von Erfahrungen aus Datenbanken, pl+a-Symposium
Silk & Urban (1978):	Pre Test Market Evaluation of New Packaged Goods: A Model and Measure Methodology, JMR
Urban & Katz (1983):	Pre-Test-Market Models: Validation and Managerial Implications, JMR