

## Statistische Methoden in der Politikwissenschaft

### Wahlforschung:

- **Umfragen**
  - Umfrageumfang: je nach Kostenrahmen 800 bis 2000 repräsentativ ausgewählte Befragte
  - Auswahlkriterien: Zufalls- oder Quotenauswahl
  - meist telefonische Befragung
  - Fehlertoleranz beträgt je nach Umfang und Methode zwischen zwei und vier Prozent
  
- **Wahlprognosen**
  - beruhen auf „exit polls“
  - Wähler werden gebeten ihre Stimme noch einmal abzugeben
  - sozialstatistische Merkmale werden erfasst
    - Vorteile:
      - nur tatsächliche Wähler befragt, Zeitspanne zwischen Wahl und Befragung gering
      - größerer Umfang: bei Landtagswahlen ca. 6.000; bei Bundestagswahlen ca. 25.000
  
- **Hochrechnungen**
  - beruhen auf tatsächliches Wählerverhalten
  - Ergebnisse repräsentativ ausgewählter Stimmbezirke werden hochgerechnet
  - Stimmbezirke sind so ausgewählt, dass sie in ihrer Gesamtheit das vorangehende Ergebnis möglichst exakt abbilden
  - „Schätzung von Wahlergebnissen aus Teilergebnissen“

Übergangsverhalten der Wähler im gesamten Wahlgebiet ist homogen

Modellannahmen:

$x_{i_n}$  : Stimmen der Partei  $i$  in Gemeinde  $n$  bei der Vorwahl

$y_{j_n}$  : Stimmen der Partei  $j$  in Gemeinde  $n$  bei der neuen Wahl

Die Annahme der „homogenen Wählerströme im Wahlgebiet“ lässt sich so ausdrücken:

$$Y_{j_n} = \sum_{i=1}^J X_{i_n} p_{ij}, j=1 \dots J, n=1 \dots N$$

Stimmergebnis der neuen Wahl kommt zustande, indem in allen Gemeinden derselbe Prozentsatz jener Wähler, die bei der Vorwahl Partei  $i$  gewählt haben, bei der neuen Wahl Partei  $j$  wählen und dass diese Tatsache für alle  $i=1\dots I$  und  $j=1\dots J$  gilt.

Grundlage dieses Modells ist, dass bei beiden Wahlgängen dieselben Wähler wahlberechtigt waren, was natürlich eine vereinfachende Annahme ist.

Gesamtheit der Wahlberechtigten in allen Gemeinden:

$$S_n = \sum_{i=1}^I X_{i_n} = \sum_{j=1}^J X_{j_n}$$

Anteil der einzelnen Parteien:

$$x_{i_n} = \frac{X_{i_n}}{S_n}$$

$$y_{j_n} = \frac{Y_{j_n}}{S_n}$$

Ebenso gilt:

$$y_{j_n} = \sum_{i=1}^I x_{i_n} p_{ij}, \quad j=1\dots J, \quad n=1\dots N$$

Die Hochrechnung hat folgende Aufgabe:

Anhand der Gemeinden, wo bereits beide Wahlergebnisse (Vorwahl und aktuell Wahl) vorhanden sind, schätzt man die Übergangskoeffizienten (Wählerströmekoeffizienten) =  $\hat{p}_{ij}$ .

Mittels der Gleichung

$$\hat{y}_{j_n} = \sum_{i=1}^I x_{i_n} \hat{p}_{ij}$$

kann man nun das Ergebnis der aktuellen Wahl in einer Gemeinde  $n$  schätzen.

Durch Addition der bereits ausgezählten und der geschätzten Gemeinden, erhält man eine aktuelle Wahlhochrechnung.

Die  $\hat{p}_{ij}$  liefern gleichzeitig eine „Wählerstromanalyse“.

## Time series cross section (TSCS):

Anwendung in der vergleichenden Policyanalyse:

Sozialpolitik, Arbeitsmarktpolitik, Wirtschaftspolitik, auch in der Wirtschaftswissenschaft benutzte Methode, Umweltpolitik

- Länderanalysen (cross section analysis)
- Zeitreihenanalysen (time series analysis)
- Gepoolte Zeitreihenanalyse (time-series-cross-section analysis, TCSC)

Gepoolte Zeitreihenanalyse modellieren Zeit- und Ländereffekte

→ gegenwärtiger Standard der quantitativ vergleichenden Politikwissenschaft

mit der Verbindung von Zeitreihen- und Ländervergleichen steigt allerdings die Komplexität der Analyse und die Anzahl der zu berücksichtigenden Fehler- und Verzerrungsquellen.

### Vorteile der TSCS:

- Erhöhung der Beobachtungen (Fallzahl)  
→ Erhöhung des Freiheitsgrad einer Regressionsanalyse
- sowohl Veränderungen in der Zeitdimension als auch zwischen den Ländern lassen sich erfassen
- Besondere Stärke: Erfassung von Zusammenhängen die über Zeitperioden Gültigkeit besitzen

Mit TSCS Analysen lassen sich folglich vor allem nomothetische Zusammenhänge analysieren.

Zunächst wird Fall A für jedes Jahr in die Spalte aufgenommen, dann Fall B bis Fall N. Um allerdings eine TSCS durchführen zu können, kommt es auf die konkrete Modellspezifikation an. In der vergleichenden Politikwissenschaft hat sich ein Standardverfahren durchgesetzt. Dieses verfahren berücksichtigt die vier potentiellen Problem einer TSCS.

### Potentielle Probleme einer TSCS:

- a) Autokorrelation
- b) Heteroskedastizität
- c) gleichzeitige Fehlerkorrelation durch gemeinsame externe Faktoren
- d) das gleichzeitige Auftreten von a und b

Behandlung, indem eine zeitverzögerte abhängige Variable als unabhängige Variable sowie Länderdummies und Jahresdummies in das Modell aufgenommen werden

Grundmodell:

$$Y = a + bY_{it-1} + \sum bX_{it} + \sum b * \text{Länderdummies} + \sum b * \text{Jahresdummies} + e_{it}$$

Das wichtigste Problem besteht darin, dass die verzögerte abhängige Variable  $Y_{t-1}$ , plus Länder- und Zeitdummies kaum noch Varianz für substantielle Variablen übriglassen.

Des Weiteren ist es problematisch, wenn eine zeitverzögerte abhängige Variable mit Länderdummies in Untersuchungen mit geringer Periodizität (kleines  $t$ ) kombiniert wird.

Technisch betrachtet, kann die Korrelation zwischen Länderdummies und der zeitverzögerten abhängigen Variable zu Verzerrungen der Schätzer führen.

- Zeitdummies dienen dazu, unkontrollierte Zeiteffekte wie Schocks und andere Ereignisse aufzufangen.
- Länderdummies eliminieren den Effekt von relativ zeitinvarianten Variablen wie Korporatismus, Föderalismus etc. weil sie mit diesen stark korrelieren.

Die meisten TSCS Analysen gehen von gleichen Zeitverzögerungseffekten für alle Fälle und Variablen aus. Meistens werden alle Variablen mit einer Zeitverzögerung von einem Jahr modelliert.

### **Lineare gemischte Modelle/ linear mixed (effect) models:**

Zwei grundsätzlich verschiedene Typen:

- feste Effekte (fixed effects): beeinflussen nur den Mittelwert von  $Y$
- zufällige Effekte (random effects): beeinflussen nur die Varianz von  $Y$

→ Mixed effect model berücksichtigt beide Arten von Effekten

Daten:

$$\{(y_i, z_i) = (y_{i1}, \dots, y_{in}, z'_{i1}, \dots, z'_{in}) : i = 1, \dots, m\}$$

bestehen aus wiederholten Messungen für metrische Zielvariable  $y$  und metrischen oder binär kodierten Kovariablen  $z = (x', u')$

→ das lineare gemischte Modell ist gegeben durch

$$y_{ij} = x'_{ij}\beta + u_{ij}b_i + \epsilon_{ij}, \quad j = 1, \dots, n; i = 1, \dots, m$$

bzw.

$$y_i = X_i\beta + U_i b_i + \epsilon_i, \quad i = 1, \dots, m$$

Voraussetzung:

zufällige Effekte:  $b_i, i = 1, \dots, m$  ist unabhängig und identisch normalverteilt  $b_i \stackrel{iid}{\sim} N(0, D)$

Fehler:  $\epsilon_i, i = 1, \dots, m$  sind unabhängig und normalverteilt  $\epsilon_i \sim N(0, \Sigma_i)$

$b_1, \dots, b_m, \epsilon_1, \dots, \epsilon_m$  sind unabhängig,  $D$  ist positiv semi-definit und  $\Sigma_1, \dots, \Sigma_m$  ist positiv definit

### Vorteile der Analyse mit gemischten Modellen:

Die Berücksichtigung individuenspezifischer Information kann im Vergleich zur Schätzung eines einfachen linearen Modells zu einer verbesserten Schätzgenauigkeit führen.

Die Schätzer für die zufälligen Effekte ermöglichen individuelle Prognosen, die in einem herkömmlichen Regressionsmodell nicht möglich sind.

### Quellenverzeichnis:

Nathaniel L. Beck; Political Methodology: A Welcoming Discipline; Journal of the American Statistical Association (651-654)

Dieter Roth; Empirische Wahlforschung: Ursprung, Theorien, Instrumente und Methoden; 2.Auflage; VS Verlag für Sozialwissenschaften

Karl-Rudolf Korte; Wahlen in der Bundesrepublik Deutschland; Bundeszentrale für politische Bildung

Detlef Jahn, Einführung in die vergleichende Politikwissenschaft; VS Verlag für Sozialwissenschaften

Hans-Joachim Lauth, Gert Pickel, Susanne Pickel; Methoden der vergleichenden Politikwissenschaft; VS Verlag

Dielman, Terry E. ; Pooled Cross-Sectional and Time Series Data Analysis; CRC Press

S.Albers, D.Klapper, U.Konradt, A.Walter, J.Wolf; Methodik der empirischen Forschung; Gabler

N. Beck, J.N. Katz; What to do With Times Series Cross Section Data; American Political Science Review,634-647

[http://www.statistik.lmu.de/institut/ag/fda/mixedmodels\\_ose11/downloads/kapitel2.pdf](http://www.statistik.lmu.de/institut/ag/fda/mixedmodels_ose11/downloads/kapitel2.pdf)

[http://brandenburg.geoecology.uni-potsdam.de/users/schroeder/teaching/ss2006/statistics/material/vl\\_stat\\_2006\\_06\\_27.pdf](http://brandenburg.geoecology.uni-potsdam.de/users/schroeder/teaching/ss2006/statistics/material/vl_stat_2006_06_27.pdf)

<http://www.kai-arzheimer.com/Lehre-Regression/stunde4b-print.pdf>