

Zur statistischen Analyse von Vollerhebungen

Andreas Broscheid / Thomas Gschwend*

Inferenzstatistische Methoden müssen selbst bei fehlerfrei gemessenen Vollerhebungsdaten angewendet werden, wenn allgemeine Erklärungen damit empirisch überprüft werden sollen. Präzision und Qualität von Erklärungen, egal ob sie einer quantitativen oder einer qualitativen Forschungslogik entspringen, sind abhängig von der theoretischen Vorstellung, die wir von den Einflüssen stochastischer Faktoren auf den zu erklärenden sozialen Prozess haben. Die getroffenen Annahmen über die Form der Stochastizität müssen adressiert und rechtfertigt werden – sei es qualitativ oder quantitativ in Form von diagnostischen Tests. Aus forschungspraktischer Sicht ist die Abhängigkeit der inhaltlichen Schlussfolgerungen von den Annahmen, die über die Form der Stochastizität gemacht werden, besonders relevant und wird im vorliegenden Aufsatz anhand zahlreicher Beispiele verdeutlicht.

1. Vollerhebungen und Stochastizität

Unter welchen Umständen müssen wir stochastische Elemente in empirische Analysen einbeziehen? Wir begrüßen, dass Joachim Behnke (2005) sich des Themas der statistischen Analyse von Vollerhebungen in der PVS angenommen hat, und wir die Möglichkeit haben, einige Anmerkungen zu seinem Beitrag zu machen, der offenbar als Reaktion auf unser MPIfG-Working Paper (Broscheid/Gschwend 2003) entstanden ist. Wir stimmen dabei mit Behnke (2005) grundsätzlich überein, dass es bezüglich der statistischen Analyse von Vollerhebungen offensichtlich Klärungsbedarf gibt. Zudem möchten wir in Übereinstimmung mit Behnke (2005) den Primat der Theorie betonen – Theorie kommt immer vor der Empirie. Methodologische Diskussionen sind deshalb kein Selbstzweck, sondern sollten als Teil der Theoriebildung betrachtet werden.

Behnkes Argument lässt sich wie folgt charakterisieren: Grundsätzlich lassen sich Signifikanztests nicht auf Vollerhebungen anwenden; es sei denn, der stochastische Charakter der Daten ist nicht nur auf die Auswahlproblematik beschränkt, sondern umfasst auch den eigentlichen Messfehler im engeren Sinn, der bei der Aufzeichnung von Daten entsteht, oder die Problematik der Operationalisierung eines theoretischen Konzeptes, was Behnke (2005: O-12) den „*ontologischen stochastischen Aspekt*“ des Inferenzproblems nennt.

* Wir danken Axel Becker, Joachim Behnke, Steffen Ganghof, Martin Heipertz, Bernhard Kittel, Steffen Kühnel, Lothar Krempel sowie den anonymen Gutachtern für zahlreiche Kommentare und Anregungen zu diesem Papier. Broscheid dankt zusätzlich für die Unterstützung des Max-Planck-Instituts für Gesellschaftsforschung, Köln, das diese Arbeit ermöglichte.

Wir stimmen Behnke zu, dass Messfehler und Operationalisierungsprobleme (z.B. Proxyvariablen) zum stochastischen Charakter von Daten beitragen. Allerdings glauben wir, dass Behnke (2005) die Bedingungen unterschätzt, unter denen der stochastische Charakter von Daten relevant wird. Unser Beitrag hat das Ziel, Bedingungen dafür anzugeben, unter denen Stochastizität in irgendeiner Form in die Analyse miteinbezogen werden sollte, ob das nun durch Signifikanztests, Konfidenzintervalle, Bayesianische Höchstdichteintervalle oder durch eine sorgfältige qualitative Beschreibung der (Un-)Sicherheit der durch die Schätzungen ermittelten Vorhersagen geschieht.

Im Unterschied zu Behnke argumentieren wir, dass Stochastizität immer dann eine Rolle spielt, wenn wir *allgemeine Erklärungen* sozialer Wirklichkeit empirisch überprüfen. Hierfür werden stochastische Modelle¹ benutzt, die quantitativer oder auch qualitativer Natur sein können. Stochastizität ist notwendig, da allgemeine Theorien und Modelle indeterministisch sind, weil sie geringere Komplexität aufweisen müssen als die zu erklärende Realität; zudem muss häufig angenommen werden, dass die sozialen Prozesse, die Ausschnitte der Wirklichkeit, die es zu erklären gilt, selbst stochastisch sind. Dabei mag es aus philosophischer Sicht notwendig sein, die genaue Natur dieser Stochastizität zu ermitteln (von Mises 1951; de Finetti 1981; Hennig 2001). Ganz gleich, ob wir nun manche Aspekte der Wirklichkeit wie Behnke (2005: O-13) als deterministisch erachten, diese aber nicht perfekt erklären können, oder wie wir (Brotschoid/Gschwend 2003: 11–14) als stochastisch ansehen, die Implikationen für die Forschungspraxis sind die gleichen. Die politikwissenschaftliche Forschung sollte stochastische Komponenten noch öfter explizit in Theoriebildung und Analyse miteinbeziehen, als das die Argumentation von Behnke (2005) im Falle von Vollerhebungen bereits impliziert.

2. Zur stochastischen Natur von Erklärungen

Wir wollen Joachim Behnkes Frage nach den Bedingungen, unter denen Stochastizität in Analysen einbezogen werden muss, hier fortführen. Behnke nennt die Stichprobenatur vieler Daten und die mögliche Existenz (und Messbarkeit!) von Messfehlern als Bedingungen, unter denen Inferenzstatistik angewandt werden muss. Im Folgenden fügen wir noch eine weitere Bedingung hinzu. Um diese zusätzliche Bedingung zu verstehen, hilft es, ein Beispiel zu betrachten, in dem wir es weder mit einer Stichprobe noch mit Messfehlern zu tun haben. Ein für die vergleichende Politikwissenschaft typisches Beispiel wären Policyentscheidungen einer abgrenzbaren Gruppe von Regierungen; wir verwenden Fortbildungsanforderungen für Mediziner in den 50 US-Staaten. Diese Anforderungen sind gesetzlich – per Gesetz oder Verordnung – festgelegt und

¹ Behnke (2005: O-13) bezweifelt die Notwendigkeit stochastischer Modelle insbesondere für „alle Gegebenheiten, die wir als unmittelbare Folgen bewussten Handelns auffassen“. Folgen bewussten Handelns sind jedoch nicht deterministisch. Wenn das so wäre, müssten wir menschliches Verhalten vollständig erklären können. Dem ist nicht so. Zudem müssten wir die Folgen bewussten Handelns immer perfekt erklären können. Tatsächlich wird es aber immer eine Vielzahl von Faktoren geben, die eine solche komplette Erklärung unmöglich macht.

gut dokumentiert, sie können fehlerfrei und vollständig erfasst werden. Nach Behnkes Argumentation wäre also Inferenzstatistik nicht unbedingt notwendig.

Wir stimmen Behnke in diesem Fall auch zu, dass Inferenzstatistik in unserem Beispiel unter bestimmten Umständen nicht notwendig ist. Wenn wir z.B. die rechtlichen Regelungen zur Ärztefortbildung in den Vereinigten Staaten *beschreiben* wollen, reicht es völlig aus, unsere Vollerhebungsdaten darzulegen; Signifikanztests o.Ä. sind dann nicht notwendig. Ähnlich sieht die Situation aus, wenn wir die Ärztefortbildungsregeln verwenden, um andere Prozesse zu erklären. Wenn wir etwa wissen wollen, ob die rechtlichen Regeln zur Ärztefortbildung die Einführung neuer Behandlungsmethoden erklären, dann können wir unsere Daten über die Ärztefortbildung als „fix“ oder nicht-stochastisch akzeptieren.

Die Situation ist aber völlig anders, wenn wir die rechtlichen Regelungen zur Ärztefortbildung *erklären* wollen. In diesem Fall sind wir gezwungen, die rechtlichen Regelungen als Ausdruck eines Prozesses zu betrachten, der mit einer systematischen und einer nicht-systematischen, d.h. „stochastischen“ Komponente (King 1989: 8–11) modelliert wird.² Nehmen wir als Beispiel an, wir wollten die Fortbildungsregeln als linearen additiven Effekt verschiedener Faktoren erklären:

$$\text{Fortbildungsregeln}_i \text{ (in Stunden/Jahr)} = a + b_1 \times \text{Unabhängigkeit der Zulassungsbehörde}_i + b_2 \times \text{Proportionaler Anteil an Nicht-Medizinern in der Zulassungsbehörde}_i + b_3 \times \text{Bevölkerungsanteil über 65 Jahren}_i + b_4 \times \text{Stärke ärztlicher Interessenverbände}_i + b_5 \times \text{Mittleres Bevölkerungseinkommen}_i + b_6 \times \text{Region}_i.^3$$

Dieses Modell stellt ein System von 50 Gleichungen mit sieben Unbekannten dar – die Daten der abhängigen und unabhängigen Variablen sind gegeben, und die „a“- und „b“-Parameter müssen ermittelt werden. Solange die Datenprofile nicht Vielfache von lediglich sieben Profilen sind, wird sich diese Gleichung nicht lösen lassen. In anderen Worten: Die Erklärung, die hier vorgeschlagen wird, ist niemals perfekt (oder deterministisch), sie kann nicht alle Fälle genau vorhersagen. Zur Modellierung eines solchen Prozesses reicht die systematische Komponente nicht aus. Dies hat aber nichts mit dem gewählten Beispiel zu tun. Vielmehr muss jede *allgemeine* Erklärung eines Sachverhalts die Komplexität dieses Sachverhalts reduzieren. Die Abweichung der tatsächlichen Beobachtung von der (hier linearen) Vorhersage durch die systematische Komponente des Modells stellt aber ein Element der Erklärung dar, das üblicherweise mit ϵ bezeichnet wird und als stochastisch definiert werden kann. Sie wird als die stochastische Komponente des Modells bezeichnet. Die (lineare) Erklärung hat dann also folgende Form:

$$\text{Fortbildungsregeln}_i \text{ (in Stunden/Jahr)}_i = a + b_1 \times \text{Unabhängigkeit der Zulassungsbehörde}_i + b_2 \times \text{Proportionaler Anteil an Nicht-Medizinern in der Zulassungsbehörde}_i + b_3 \times \text{Bevölkerungsanteil über 65 Jahren}_i + b_4 \times \text{Stärke ärztlicher Interessenverbände}_i + b_5 \times \text{Mittleres Bevölkerungseinkommen}_i + b_6 \times \text{Region}_i + \epsilon_i.$$

² Wir verwenden Anführungszeichen, da Stochastizität in diesem Zusammenhang verschiedene Interpretationen erlaubt, von denen sich einige nicht auf Stochastizität im eigentlichen Sinne beziehen. Mehr davon unten.

³ Der Subskript „i“ ist ein Index für die einzelnen Staaten/Beobachtungen. Zur substantziellen Begründung dieser Variablen und empirischen Ergebnissen siehe Broscheid und Teske (2003).

Für die substanzielle Interpretation von ϵ , der stochastischen Komponente des Modells, bieten sich zwei Argumente an. Erstens kann ϵ als Zufallsfaktor aufgefasst werden. In dieser Interpretation erklärt die lineare Kombination unabhängiger Variablen und Parameter alles, was systematisch erklärbar ist. ϵ bezieht sich dann auf reine Zufallsschwankungen, die nicht weiter erklärt werden können. Diese Interpretation von ϵ mag gerade in den Sozialwissenschaften überzeugen, da menschliches Verhalten als nicht vollständig erklärbar aufgefasst werden kann. Beispielsweise kann der Zufall im politisch-administrativen Entscheidungsprozess, der die ärztlichen Fortbildungsregeln bestimmt, in dem Sinne eine Rolle spielen, dass unterschiedliche Entscheidungsträger unklare Vorstellungen über die Konsequenzen verschiedener Regelungen haben. Diese unterschiedlichen Vorstellungen können als zufällig bestimmt betrachtet werden.

Das soeben genannte Beispiel kann aber auch die zweite Interpretation von ϵ motivieren: Der „Zufallsfaktor“ mag eben nicht rein zufällig bestimmt sein, sondern durch die Einflüsse vieler, nur schwer (oder nicht) messbarer Faktoren, zum Beispiel durch die individuellen Vorstellungen der Entscheidungsträger über die Angemessenheit verschiedener Fortbildungsregeln. Prinzipiell ist es vorstellbar, dass irgendwann einmal Wissenschaftler diese Faktoren messen können, aber faktisch ist dies nicht der Fall. Aus epistemologischer Sicht werden wir zudem nicht alle möglichen Einflussfaktoren in unsere Analyse einbeziehen wollen, selbst wenn wir dies könnten, da ansonsten der Komplexitätsgrad unserer Erklärungen zu hoch wäre, und diese dadurch ihren allgemeinen Charakter einbüßten.

Unabhängig davon, ob wir ϵ als „reinen“ Zufallsfaktor interpretieren oder als Sammelmaß für alle relevanten aber ignorierten Einflussfaktoren (oder als Kombination beider Interpretationen) – um das Erklärungsmodell der Fortbildungsregelung zu schätzen, müssen wir eine Theorie über ϵ aufstellen bzw. ϵ in unsere bestehende Theorie integrieren, das heißt: Annahmen machen. Diese Annahmen können vielfältig sein: Um eine lineare Regression nach dem OLS-Verfahren zu schätzen,⁴ müssen wir etwa annehmen, dass die einzelnen Fehlerwerte ϵ_i gleiche Varianzen haben, unabhängig voneinander und im Durchschnitt gleich Null sind, nicht mit den unabhängigen Variablen korrelieren, und dass sie beliebige Werte annehmen können (was impliziert, dass die abhängige Variable stetig und unbegrenzt ist). Wenn wir diese Annahmen nicht rechtfertigen können oder diagnostische Tests diese Annahmen nicht stützen – wenn etwa die einzelnen ϵ_i nicht unabhängig voneinander sind – dann müssen wir dies explizit ins Modell miteinbeziehen und schätzen.⁵ Dabei gibt es theoretisch zwei Wege. Erstens

4 Wir benutzen die lineare Regression nur, weil sie vielen vertraut ist. Einfache Mittelwertvergleiche aber auch ANOVA können genauso wie viel kompliziertere Modelle (Faktoranalysen, Logit-, Probit- und andere nichtlineare Modelle) als Regressionsmodelle dargestellt und mit derselben Logik behandelt werden. Das Schätzmodell, das benutzt werden muss, richtet sich natürlich nach den Annahmen, die über ϵ gemacht werden.

5 Intuitiv stehen weniger Informationen zur Verfügung, wenn die einzelnen Beobachtungen nicht unabhängig voneinander sind. Wird trotzdem angenommen, dass die Beobachtungen unabhängig sind, dann werden die geschätzten kausalen Effekte sicherer erscheinen als sie es wirklich sind. Das hat wichtige inhaltliche Konsequenzen, die betont werden sollten. Im Übrigen werden, abgesehen von experimentellen Daten, in den Sozialwissenschaften niemals völlig unabhängige Beobachtungen erhoben. Deshalb wird trotzdem, und vernünftigerweise, oft die Unabhängigkeit angenommen, wenn die Natur der Daten nicht weit von diesem theoretischen

können die Verletzungen der Annahmen als *Petitessen* betrachtet und statistisch korrigiert werden, um zu effizienten und unverzerrten Schätzungen zu kommen. Beispielsweise stehen im Rahmen des GLS-Ansatzes (*generalized least squares*) viele Tests und Methoden bereit, um räumliche bzw. zeitliche Abhängigkeiten sowie ungleiche Varianzen der Residuen zu diagnostizieren und gegebenenfalls zu korrigieren (z.B. robuste Standardfehler; vgl. Gschwend/Leuffen 2005: Tabelle 1). Zweitens können diagnostizierte Verletzungen der Modellannahmen über die ϵ_i als substantiell interessant betrachtet werden. Das kann zu präziseren Theorien über ϵ , die Stochastizität des sozialen Prozesses, führen, die explizit modelliert werden können und so letztlich tiefere Erklärungen ermöglichen (z.B. Gschwend 2004: 64–65).

Unabhängig davon, welcher Weg beschritten wird, können prinzipiell vielfältige Annahmen über die stochastische Komponente ϵ , die sich in Annahmen über die Verteilung der abhängigen Variablen übersetzen lassen, in empirische Modelle integriert werden. Die Einbeziehung von ϵ in die Analyse lässt sich mit verschiedenen Methoden und Ansätzen verfolgen. Das genaue methodische Vorgehen hängt dabei von epistemologischen (z.B. klassische oder bayesianische Statistik), methodologischen (z.B.: Sind Signifikanztests eine adäquate Methode des Hypothesentests?) und theoretischen Entscheidungen (Verteilung von ϵ , exakte Form des Erklärungsmodells) ab. Egal, wie unsere Entscheidungen im Einzelnen ausfallen mögen, es ist sicher, dass wir solche Entscheidungen nicht vermeiden können.

Wenn unsere Daten auf einer Stichprobe beruhen, spielen die gerade beschriebenen Zufallsfaktoren übrigens auch eine Rolle. Zusätzlich zu diesen kommt dann aber auch noch die Stochastizität der einfachen Zufallsauswahl hinzu. Der große Vorteil der einfachen Zufallsauswahl ist die Tatsache, dass die Unabhängigkeit der Beobachtungen, und dadurch auch der Fehlerwerte, theoretisch garantiert werden kann. Wenn die Daten nicht auf einer einfachen Zufallsauswahl beruhen oder diagnostische Tests zu einem anderen Schluss kommen, müssen eventuelle Abhängigkeiten zwischen Beobachtungen oder Fehlerwerten explizit modelliert werden.

3. Konkretisierung an Beispielen

Fassen wir das bisherige Argument zusammen: Uns ging es um die Frage, unter welchen Umständen wir stochastische Elemente in statistischen Analysen einbeziehen müssen. Mit Behnke (2005) stimmen wir darin überein, dass bei Stichprobenanalysen und – bedingt – bei Messfehlern stochastische Elemente Teil der Schätzung sein müssen. Im Unterschied zu Behnke (2005) fügen wir aber hinzu, dass selbst bei fehlerfrei gemessenen Vollerhebungsdaten stochastische Elemente eine Rolle spielen, wenn die Daten erklärt, also als abhängige Variable modelliert werden.⁶

Maximalstandard abweicht. Allerdings gibt es quantitative Methoden, um diese Abhängigkeiten zu korrigieren oder explizit zu modellieren.

⁶ Präziser als der Begriff der abhängigen Variablen wäre der der endogenen Variablen. Unabhängige Variablen können nämlich auch abhängige Variablen sein, etwa in einem hierarchischen oder nicht-rekursiven Gleichungssystem. Solche Variablen werden als exogene Variablen bezeichnet und müssen auch als stochastisch analysiert werden. Wir verzichten auf eine detaillierte Diskussion dieses Falles, da dies für unser Hauptargument nicht notwendig ist.

Wir können dieses Argument nun auf die Beispiele anwenden, die Behnke (2005) anführt: die Geschlechterverteilung in seinen Kursen und die unterschiedlichen Werbe-strategien von CDU und SPD. Die Geschlechterverteilung kann als nicht-stochastische Variable angesehen werden, solange wir die Verteilung nur beschreiben oder aber andere Variablen mit ihr erklären wollen (etwa den Studienerfolg weiblicher und männlicher Studenten). Sobald wir aber die Verteilung selbst erklären wollen (denken wir etwa an die Frage, ob männliche und weibliche Studenten unterschiedliche Kursthemen bevorzugen), müssen wir die Variable bzw. ihre Erklärung als eine stochastische ansehen und stochastische Elemente in unsere Schätzung einbeziehen. Das statistische Modell, welches den Prozess repräsentiert, der die Geschlechterverteilung hervorbringt, muss also neben einer systematischen auch eine stochastische Komponente beinhalten. Aufgabe sozialwissenschaftlicher Erklärungen ist es, Aussagen über die Ursache von sozialen Prozessen zu machen und die Unsicherheiten der resultierenden Schlussfolgerungen abzuschätzen (King et al. 1994: 7–9; Schnell et al. 1999: 6).

In Bezug auf die möglichen Unterschiede in der Verteilung der Wahlkampfplakate stimmen wir Behnke wieder zu, dass ein Signifikanztest nicht sinnvoll ist, solange es wirklich nur um den Anteil bestimmter Wahlkampfplakate eines Jahres geht. Wir können Behnkes Modell als lineares Modell umformen:

$$\text{Proportion sozialer Themen}_i = a + b \times \text{SPD}_i + \epsilon_i.^7$$

„SPD“ ist dabei eine Dummy-Variable, die den Wert „1“ annimmt für die zur SPD gehörende Proportion sozialer Themen, und Subskript „i“ kann die Werte „SPD“ und „CDU“ annehmen. Da wir zwei Datenpunkte und zwei Unbekannte haben (die Koeffizienten „a“ und „b“), ist $\epsilon = 0$. Eine stochastische Komponente muss nicht angenommen werden. Aber das Modell ist eben kein allgemeines Erklärungsmodell, da die Zahl der Beobachtungen der Zahl der Erklärungsfaktoren gleicht.⁸

Anders sähe es aus, wenn wir den Inhalt jedes einzelnen Plakats erklären wollten. Dann hätten wir eine große Anzahl von Fällen mit zwei Erklärungsfaktoren, da nämlich jedes Plakat eine Beobachtung wäre, die durch die Konstante und die „SPD“-Variable „erklärt“ würde. Da es vier verschiedene Datenprofile gibt (CDU-Sozial, SPD-Sozial, CDU-Wirtschaft, SPD-Wirtschaft), werden die zwei Erklärungsfaktoren der systematischen Komponente des Modells die Inhalte der einzelnen Plakate nicht deterministisch erklären. Der Erklärungsfehler kann als stochastische Fluktuation verstanden werden. Die stochastische Komponente spielt also selbst bei Vollerhebungsdaten eine wichtige Rolle, wenn die Daten als abhängige Variable erklärt werden sollen. Sie muss Bestandteil des Modells sein. Es kommt also auf die genaue theoretische Frage an, die geklärt werden soll, will heißen: die abhängige Variable, die erklärt wird.⁹

7 Natürlich ist die von Behnke (2005: O-8ff.) postulierte Null-Hypothese, dass 50 Prozent der Plakate beider Parteien sich mit sozialen Themen befassen, hier irrelevant. Behnkes Frage ist, ob die Themenverteilung bei beiden Parteien gleich ist; diese Gleichheit kann es in Bezug auf alle Proportionen geben. Dieser Fehler ändert aber nichts an der grundsätzlichen theoretischen Fragestellung.

8 Alternativ könnte man auch die Differenz zwischen beiden Proportionswerten als abhängige Variable nehmen. Dann hätten wir nur eine Beobachtung mit einer unabhängigen Variablen – wieder kein allgemeines Modell.

9 In diesem Fall ist die abhängige Variable mit ihrem Fokus auf jedes einzelne Plakat substanziell

4. Was hat das mit konkreter Forschung zu tun?

Zur Erklärung sozialer Prozesse muss Stochastizität integriert werden, selbst wenn die Überprüfung der Erklärung auf Vollerhebungsdaten basiert. Wie relevant ist unser Argument für Sozialwissenschaftler, die hauptsächlich an der Untersuchung substanzieller Forschungsfragen interessiert sind? Die Forschungsrelevanz des Disputes um die Einbeziehung von Stochastizität selbst bei Vollerhebungsdaten ist inhaltlich motiviert. Nach den Naturwissenschaften sollten auch die Sozialwissenschaften anerkennen, dass ihre Schlussfolgerungen inhärent unsicher sind (z.B. Pollock 1979: 1). Die Präzision und Qualität der Erklärungen, egal ob sie einer quantitativen oder einer qualitativen Forschungslogik entspringen, sind abhängig davon, welche theoretische Vorstellung wir von den Einflüssen stochastischer Faktoren auf den zu erklärenden sozialen Prozess haben. Die getroffenen Annahmen über die Form der Stochastizität müssen spezifiziert und gerechtfertigt werden – sei es qualitativ oder quantitativ in Form von diagnostischen Tests.

Die Abhängigkeit der Schlussfolgerungen von den Annahmen, die über die Form der Stochastizität gemacht werden, möchten wir durch ein konkretes Beispiel aus der quantitativ-komparativen Politikforschung verdeutlichen, welches sich mit Fragen der Politikbildung unter dem Einfluss der Globalisierung beschäftigt (Kittel/Winner 2005). Kittel und Winner untersuchen den Zusammenhang zwischen Globalisierung, Regierungsbildung und Staatsausgaben. Die verwendeten Daten beziehen sich auf 17 OECD-Staaten im Zeitraum zwischen 1961 und 1993 und können als Vollerhebung der 17 Staaten über diese Periode angesehen werden. Wir möchten hier auf ein hochinteressantes Ergebnis hinweisen, über das Kittel und Winner berichten (vgl. Kittel/Winner 2005: Tabelle 4). Wenn wir zunächst nur auf die Größe der geschätzten Koeffizienten schauen, dann zeigt sich, dass die Regierungsbeteiligung christdemokratischer Parteien einen kleinen, aber substanziellen Einfluss auf Staatsausgaben zu haben scheint. Diese Modelle errechnen eine Senkung der Staatsausgaben zwischen 0,04 und 0,6 Prozentpunkte des BIP durch eine christdemokratische Regierung verglichen mit einer Regierung ohne Beteiligung christdemokratischer Parteien. Wenn wir zudem von der heroischen Annahme ausgehen wollten, dass die Staatsausgaben fehlerfrei gemessen würden, dann wäre nach der von Behnke (2005) vertretenen Meinung die Analyse beendet. Fazit: Die Höhe der Staatsausgaben in den 17 OECD-Staaten wird unter anderem durch die parteipolitische Zusammenstellung der Regierung bestimmt.

Das Problem ist aber, dass hier Staatsausgaben *erklärt* werden sollen. Obwohl die Autoren eine Reihe weiterer unabhängiger Variablen einbeziehen, ist es unmöglich, alle möglicherweise relevanten und einflussreichen Faktoren zu messen und in der Analyse zu berücksichtigen.¹⁰ Die nicht analysierten Faktoren werden von der stochastischen

wohl nicht interessant für Politologen, auch wenn man keine Frage von vornherein als irrelevant ausschließen sollte. Die Schätzung des Modells könnte in diesem Fall auch nicht mit OLS vorgenommen werden, sondern müsste die binäre Natur der abhängigen Variablen modellieren.

10 Es lohnt, sich mögliche unmessbare Faktoren vorzustellen: Erkrankungen, die höhere Belastungen der Gesundheits- und Sozialkosten hervorrufen, individuelle Entscheidungen, staatliche Förderungen in Anspruch zu nehmen, größere und kleinere Naturereignisse, die die Staatskos-

Komponente des Modells absorbiert, und die Frage stellt sich, ob der errechnete Zusammenhang zwischen der Regierungsbeteiligung christdemokratischer Parteien und Staatsausgaben nicht ein Artefakt der stochastischen Komponente sein mag. Dieser Frage gehen Signifikanztests nach. Die von Kittel und Winner durchgeführten Signifikanztests (2005: Tabelle 4) deuten dann auch darauf hin, dass der errechnete Zusammenhang rein zufällig zustande kommt und über den allgemein akzeptierten Schwellenwerten für Signifikanztests liegt. Diese Modellschätzungen begründen deshalb unserer Auffassung nach nicht die Annahme, wonach die Regierungsbeteiligung christdemokratischer Parteien die Staatsausgaben beeinflusst. Die Einbeziehung von Stochastizität zur Erklärung der unterschiedlichen Höhe der Staatsausgaben führt zu wesentlich konservativeren und – wie wir behaupten – damit zu realistischeren Schlussfolgerungen.

Aus forschungspraktischer Sicht ist die Arbeit von Kittel und Winner (2005) zusätzlich besonders erhellend, weil die Autoren vorführen, dass Fehlerkonzeption und -schätzung einerseits und substanzielle empirische Untersuchung andererseits nicht voneinander zu trennen sind. Gerade Daten, die nicht auf Zufallsstichproben beruhen, weisen komplexe stochastische Prozesse auf (Greene 2000: 592–608) – etwa räumliche und zeitliche Abhängigkeiten zwischen Beobachtungseinheiten –, die substanzielle Bedeutung haben. Nach Kittel und Winner können Annahmen über stochastische Prozesse die Schätzung der Parameterwerte und damit die darauf beruhenden Schlussfolgerungen beeinflussen und stellen gerade deshalb einen wichtigen Teil der Theoriebildung empirischer Untersuchungen dar. Ihren Analysen zufolge (Kittel/Winner 2005: Tabelle 1, erste Spalte) legt die Nichtbeachtung räumlicher und zeitlicher Abhängigkeiten, die das naive Poolen aller Daten impliziert, den Schluss nahe, dass linke Regierungen systematisch die Staatsausgaben im Schnitt um fast sechs Prozentpunkte steigern verglichen mit einer Regierung ohne Beteiligung linker Parteien. Der entsprechende Koeffizient wechselt allerdings das Vorzeichen, wenn für unbeobachtete zeit- und länderspezifische Effekte kontrolliert wird. Gemäß einer solchen Theorie über stochastische Effekte würden linke Regierungen die Staatsausgaben um fast 1,3 Prozentpunkte senken (Kittel/Winner 2005: Tabelle 1, zweite Spalte). Das sind signifikant und substanziell verschiedene Ergebnisse, die nur durch die Annahme einer anderen Theorie über den Einfluss der Stochastizität zustande kommen. Entsprechende diagnostische Tests weisen die Autoren aber darauf hin, dass die mit beiden Schätzungen verbundenen Annahmen nicht haltbar sind. Die Theorie über das Wirken stochastischer Effekte muss demnach weiter verfeinert werden, was die Autoren vorbildlich leisten. Es zeigt sich (Kittel/Winner 2005: Tabelle 4), dass Parteipolitik keinen vom Zufall verschiedenen Einfluss auf die Staatsausgaben der untersuchten 17 OECD-Staaten hat.

Manche Schlussfolgerungen beruhen offenbar auf ungeprüften Annahmen über das Wirken stochastischer Prozesse, die theoretisch wie empirisch nicht haltbar sind. Wichtig ist daher, die gemachten Annahmen über das Wirken stochastischer Einflüsse zu adressieren und zu rechtfertigen.

ten beeinflussen, usw. Es ist offensichtlich, dass diese Faktoren durchaus als zufällig aufgefasst werden können.

5. Fazit

Zum Schluss möchten wir noch einmal hervorheben, dass wir – bei allen Unterschieden – mit Joachim Behnke grundsätzlich übereinstimmen, dass die sozialwissenschaftliche Datenanalyse sorgfältiger theoretischer und methodischer Überlegungen bedarf. Wir stimmen Behnke auch darin zu, dass es Umstände gibt, unter denen Daten im Prinzip als deterministisch angesehen werden können – nämlich dann, wenn weder Stichproben noch Messfehler vorhanden sind, und wenn die Datenanalyse rein deskriptive Zwecke verfolgt. Sobald wir aber Wirklichkeit allgemein erklären wollen – und hier scheint unser Argument von dem Behnkes abzuweichen –, müssen wir Stochastizität in Betracht ziehen. In dieser Ansicht sehen wir uns nicht nur durch unsere theoretischen Argumente bestätigt, sondern auch durch die internationale Forschungspraxis, die diesen methodologischen Punkt bereits akzeptiert, wie eine Inhaltsanalyse der Modellbildung vergleichender Makrostudien ergeben hat (Bollen et al. 1993: 335–337), die insbesondere für die Problematik bei Vollerhebungen relevant sind.

Was sind die Implikationen unseres Arguments für die Forschungspraxis? Erstens bedeutet es, dass wir unsere empirischen Ergebnisse einer methodisch fundierten Kritik unterwerfen müssen, indem wir Stochastizität nicht ignorieren, sondern, soweit es geht, in unsere Analysen einbeziehen. Anstatt lediglich auf beeindruckende Beziehungen in unseren Daten hinzuweisen, sollten wir auch die Unsicherheit bestimmen, mit der diese Beziehungen tatsächliche soziale Wirklichkeiten widerspiegeln. Zweitens müssen wir möglichst präzise und schätzbare Theorien der Stochastizität unserer Daten entwickeln und testen. In einigen Fällen ist dies relativ einfach möglich (Kittel/Winner 2005), in anderen Fällen müssen wir so weit wie möglich gehen und neue Ansätze und Methoden erkunden, die es uns ermöglichen, die Unsicherheit unserer empirischen Ergebnisse zu bestimmen. Es ginge hier zu weit, bayesianische Analyseansätze zu diskutieren; in diesem Bereich ließen sich aber Möglichkeiten finden, wie auch bei geringer Fallzahl Stochastizität analysiert werden kann (z.B. Buckley 2004) – wenn auch unter grundsätzlich anderen Annahmen als in der klassischen Statistik. Drittens bedeutet es, dass Signifikanztests auch bei Vollerhebungen anwendbar sind, aber nicht das einzige Mittel der Stochastizitätsanalyse darstellen müssen. Viertens bedeutet die Hinwendung zu Unsicherheitsabschätzungen der getroffenen Schlussfolgerungen nicht, dass wir nun alle wie am Weihnachtsbaum nach den Koeffizienten schauen, die mit besonders vielen Sternen versehen sind. Statistische Signifikanz ist bekanntlich ja nur eine notwendige, beileibe aber keine hinreichende Bedingung für die substantielle Relevanz gefundener Beziehungen.

Wir sollten auch darauf hinweisen, dass wir die mehr oder minder blinde Anwendung vorgefertigter Computer-Routinen, die dann zur Affirmation durch Signifikanztests legitimierter Ergebnisse führt, für falsch halten. Gerade bei Vollerhebungsdaten sind Standardmethoden oft nicht anwendbar und müssen durch komplexere, schwerer zu schätzende statistische Modelle ersetzt werden. Ja, zuzeiten mag es durchaus nicht möglich sein adäquate Modelle zu schätzen. Unseres Erachtens wäre es aber auch in solchen Fällen falsch, Daten und Erklärungen dann als deterministische zu analysieren, da so überhaupt keine Unsicherheitsabschätzungen möglich sind. Außerdem hat es sich in der Vergangenheit immer wieder gezeigt, dass Modelle, die einst als unschätzbar gal-

ten oder nur durch starke Annahmen schätzbar wurden, durch neue Verfahren ohne Rückgriff auf starke Annahmen an die Daten geschätzt werden konnten.¹¹ Methodologische Schwierigkeiten sollten uns also nicht dazu verleiten, auf die theoretisch wichtige Schätzung stochastischer Elemente völlig zu verzichten.

Die Debatten um verschiedene Analyseansätze und -methoden sind lebhaft und kontrovers. Man denke nur an die immer wieder aufflammende Diskussion über Sinn und Unsinn von Signifikanztests (Gill 1999; Morrison/Henkel 1970). Aber hinter diesen Kontroversen sehen wir die Notwendigkeit, stochastische Prozesse in unsere Erklärungen der Wirklichkeit und in die empirische Überprüfung dieser Erklärungen einzu beziehen. Die deutsche Politikwissenschaft hat in der Vergangenheit großen Wert auf eine sorgfältige und kritische Anwendung qualitativer Methoden gelegt. Wir glauben, dass zusätzlich eine stärkere Betonung der quantitativen Methodenausbildung gefordert ist, die es den Forschenden ermöglicht, die Zufallsbedingtheit politologischer Theorien auch adäquat empirisch umzusetzen.

Literatur

- Behnke, Joachim*, 2005: Lassen sich Signifikanztests auf Vollerhebungen anwenden? Einige essayistische Anmerkungen, in: Politische Vierteljahresschrift 46, O-1–O-15, in: <http://www.vs-verlag.de/pvs>.
- Bollen, Kenneth A./Entwisle, Barbara/Alderson, Arthur S.*, 1993: Macrocomparative Research Methods, in: Annual Review of Sociology 19, 321–351.
- Broscheid, Andreas/Gschwend, Thomas*, 2003: Augäpfel, Murmeltiere und Bayes: Zur Auswertung stochastischer Daten aus Vollerhebungen. MPiFG Working Paper 03/7.
- Broscheid, Andreas/Teske, Paul*, 2003: Public Members on Medical Licensing Boards and the Choice of Entry Barriers, in: Public Choice 114, 445–459.
- Buckley, Jack*, 2004: Simple Bayesian Inference – for Qualitative Political Research, in: Political Analysis 12, 386–399.
- de Finetti, Bruno*, 1981: Wahrscheinlichkeitstheorie. Wien.
- Gill, Jeff*, 1999: The Insignificance of Null Hypothesis Significance Testing, in: Political Research Quarterly 52, 647–674.
- Greene, William H.*, 2000: Econometric Analysis. Englewood Cliffs, NJ.
- Gschwend, Thomas*, 2004: Strategic Voting in Mixed-Electoral Systems. Reutlingen.
- Gschwend, Thomas/Leuffen, Dirk*, 2005 (im Erscheinen): Divided We Stand – Unified We Govern? The Issue of Cohabitation in the French Elections of 2002, in: British Journal of Political Science.
- Hennig, Christian*, 2001: Was ist Wahrscheinlichkeit? Antworten, konstruktivistisch betrachtet, in: <http://www.math.uni-hamburg.de/home/hennig/papers/ww.pdf>.
- King, Gary*, 1989: Unifying Political Methodology. The Likelihood Theory of Statistical Inference. Cambridge.
- King, Gary/Keohane, Robert O./Verba, Sidney*, 1994: Designing Social Inquiry: Scientific Inference in Qualitative Research. Princeton.
- Kittel, Bernhard/Winner, Hannes*, 2005: How Reliable is Pooled Analysis in Political Economy? The Globalization-Welfare State Nexus Revisited, in: European Journal of Political Research 44, 1–25.
- Mises, Richard von*, 1951: Wahrscheinlichkeit, Statistik und Wahrheit. 3. Auflage. Wien.

¹¹ Man denke etwa an die Schätzung des multinomialen Probit-Modells durch Monte-Carlo-Methoden (Quinn et al. 1999).

- Morrison, Denton E./Henkel, Hamon E.*, 1970: The Significance Test Controversy. A Reader. Chicago.
- Pollock, David S.*, 1979: The Algebra of Econometrics. Chichester.
- Quinn, Kevin M./Martin, Andrew D./Whitford, Andrew B.*, 1999: Voter Choice in Multi-Party Democracies: A Test of Competing Theories and Models, in: American Journal of Political Science 43, 1231–1247.
- Schnell, Rainer/Hill, Paul B./Esser, Elke*, 1999: Methoden der empirischen Sozialforschung. München/Wien.