
Gerald Schneider / Gabriele Ruoff¹

Quantitative Methoden

1 Grundlagen der quantitativen Analyse

Eines der Grundmerkmale wissenschaftlichen Arbeitens besteht darin, Informationen so zu verdichten, dass eine Struktur erkennbar ist. Ohne Theorien über den Prozess, in dem die Daten entstanden sind, bleibt jeder Forscher blind. Doch mit Theorien allein ist es in den angewandten Wissenschaften nicht getan. Ob die Erklärung zu den Daten passt, lässt sich nur mit Hilfe von Methoden prüfen, die zugleich transparent und unabhängig von der Theorie sind, die es zu überprüfen gilt. Wenn eine theoriegeleitete Wissenschaftlerin ungeeignete Methoden verwendet, um die Datenstruktur zu beschreiben, kann sie zwei Probleme erfahren: das Fata Morgana- und das Vernebelungsproblem. Die erste Schwierigkeit besteht darin, dass die Methode den Effekt einer Ursache so überschätzt, dass der Forscher sich in falscher Sicherheit wiegt. Die zweite Folge ungeeigneter Methoden ist umgekehrt, dass die Wirkungen systematisch unterschätzt werden. Die Methodenwahl ist deshalb mitentscheidend, um schlechte und gute Forschung, wahre und falsche Erkenntnisse unterscheiden zu können. Generell gibt es zwei Kriterien, um die Angemessenheit einer jeden Methode zu beurteilen, unabhängig davon, ob das Verfahren „quantitativer“ oder „qualitativer“ Natur ist. So sollte ein geschätzter Zusammenhang („der Schätzer“) zum einen unverzerrt sein, was sich auch als Erwartungstreue oder Absenz eines Bias bezeichnen lässt. Zum anderen sollten die Resultate, die ein Verfahren bei Replikationen erbringt, eine geringe Varianz aufweisen und somit „effizient“ sein (King/Keohane/Verba 1994). Da das Effizienzkriterium oft nur bei großen Samples zu erreichen ist, gilt als Ersatzmaßstab die Konsistenz eines Schätzers. Bei einem konsistenten Schätzer nähert sich bei steigender Stichprobenfallzahl der Parameter, der geschätzt wird, dem wahren Wert an.

Die meisten methodologischen Diskussionen in der Politikwissenschaft seit den 1990er Jahren drehen sich um diese Kriterien. Die Kompetenz der einschlägigen Methodenlehre ist dabei so weit vorangeschritten, dass Politologen mittlerweile selbständig effiziente, erwartungstreue Schätzverfahren zu entwickeln versuchen. Obwohl die fachliche Diskussion anspruchsvoll ist, erreicht sie einen immer größer werdenden Kollegenkreis. Im Zuge dieser Entwicklung hat sich auch die Diskussion in der Internationalen Politik intensiviert. Dies ist anhand der einschlägigen Veröffentlichungen in der bislang einzigen Methodenzeitschrift, dem Quartalsheft *Political Analysis*, und einigen Diskussionen in Zeitschriften wie *International Organization* (z.B. die Kontroverse um sog. Fixed Effect-Modelle) oder *American Political*

¹ Wir danken Glenn Palmer (University Park, PA) und J. David Singer (Ann Arbor, MI) für seine Hilfe bei der Bereitstellung der in diesem Beitrag verwendeten Daten sowie Aurelio Tobias (Madrid) für seine Bereitstellung eines STATA-Programms zur Schätzung eines Zeitreihenpoissonmodells.

Science Review (z.B. die Auseinandersetzung um Neurale Netzwerke versus Logitregressionen) ersichtlich. In diesem Übersichtsartikel wollen wir zunächst ganz praktisch die beiden Kriterien der Effizienz und der Erwartungstreue verdeutlichen. Anschließend gehen wir auf einige der zentralen Herausforderungen und neueren Entwicklungen ein, die zur Zeit diskutiert werden.

2 Typische Probleme der Datenanalyse anhand von Beispielen

In der Politikwissenschaft hat sich seit der Publikation von King (1989) die Auffassung durchgesetzt, dass eine Methode zum Messniveau der Daten passen muss, die es zu untersuchen gilt. So passt ein OLS-Regressionsmodell (d.h. ein Regressionsmodell, in dem von linearen Beziehungen ausgegangen wird und in dem nach der Methode der kleinsten Quadrate geschätzt wird) eigentlich nur auf intervallskalierte Daten, während sich für die Analyse von Ordinalskalen Logit- und Probitmodelle eignen. Für die Analyse von Nominalskalen mit mehr als zwei Kategorien empfiehlt sich die multinominale logistische Regression. Für Häufigkeiten wie die Zahl von Streiks oder Konflikten werden am besten Poisson-, Negativ Binomial- oder verwandte Regressionstechniken eingesetzt, und Verläufe (d.h. die Zeit, bis ein bestimmtes Ereignis wie Demokratisierung einsetzt) lassen sich mit Verlaufsmodellen wie einem Weibull-Survival-Modell schätzen. Ein weiterer zentraler Gesichtspunkt, der die Verfahrenswahl beeinflusst, besteht darin, ob die Daten längsschnitt- oder querschnittsorientiert sind. Eine Kombination dieser beiden Möglichkeiten besteht in Paneldatensätzen, die in der Politikwissenschaft nach der Definition von Beck (2001) zumeist sog. TSCS-Datensätze (*Times series cross section*) sind, da die Zahl der Zeitpunkte häufig die Zahl der Einheiten (Länder, Staaten, etc.) übersteigt. Besonders bei der Analyse von Zeitreihen besteht ein Problem darin, dass die Fälle meist nicht unabhängig voneinander sind und dass somit eine zentrale Annahme der linearen Regression verletzt ist. Dies kann sich in Autoregression äußern (die Werte aufeinanderfolgender Schätzungen korrelieren untereinander; d.h. das BIP in diesem Jahr ist beeinflusst vom BIP des vergangenen Jahres) oder Autokorrelation (die Fehler aufeinanderfolgender Schätzungen korrelieren untereinander). In Querschnitten ist die Annahme der Unabhängigkeit dadurch gefährdet, dass die Wirtschaft oder Politik von geographisch benachbarten Regionen oder Staaten miteinander korrelieren. In diesem Zusammenhang ist dann von räumlicher Autokorrelation die Rede, für deren Analyse in den letzten Jahren vor allem Michael Ward und seine Ko-Autoren (z.B. Hoff/Ward 2004) zentrale Ergebnisse geliefert haben. Ein ähnliches Problem wie die Autokorrelation ist die Heteroskedastizität. Ergebnisse, die unter dieser Schwierigkeit leiden, verletzen die Annahme, dass die Varianz der Fehler für alle Werte von X gleich sein soll. Dies führt zu ineffizienten Schätzungen und verunmöglicht aussagekräftige Hypothesentests.

Die pragmatische Haltung gegenüber solchen Problemen besteht darin, die Residuen nach einer ersten provisorischen Analyse graphisch und mit Hilfe von Testverfahren (White, Beusch-Pragan etc.) auf Heteroskedastizität hin zu überprüfen. Liegt das Problem vor, besteht der übliche Ausweg darin, auf geeignetere Verfahren auszuweichen. In der Zeitreihenökonomie bestehen überdies Techniken, die bei Hochfrequenzdaten wie täglichen Börsenkursen eine Schätzung von Heteroskedastizität erlaubt. Die Clusterung der Fehler ist

ja nicht einfach eine unappetitliche Begleiterscheinung der Daten, sondern unter Umständen ein Phänomen, das zu schätzen sich lohnt. In der Politikwissenschaft haben solche Verfahren – es handelt sich im Wesentlichen um GARCH-Modelle und ihre Erweiterungen² – über die Analyse der politischen Determinanten von Finanzmarktdaten eine gewisse Verbreitung erfahren (Leblang/Mukerjee 2004, Schneider/Tröger 2004). Wer sich der Natur seiner Daten bewusst ist, hat bei der quantitativen Analyse schon halb gewonnen. Wohin aber die falsche Methodenwahl führen kann, wollen wir anhand eines klassischen Textes zeigen. Der Artikel von Bremer, Singer und Stuckey (1972, fortan BSS) gilt als einer der ersten Veröffentlichungen, in dem in der Internationalen Politik ein multivariates OLS-Verfahren zur Anwendung gelangte. Was retrospektiv durchaus als bahnbrechender Beitrag zur Theorie der Internationalen Politik zu gelten hat, ist vom methodischen Anspruch heute veraltet. Wie zu zeigen ist, stehen so in heutigem Licht auch die Schlussfolgerungen von BSS auf tönernen Füßen. In unserem Forschungsdesign halten wir uns, so gut es geht, an die Untersuchungsanlage, wie sie im Originaltext beschrieben ist. Leider lässt sich die Studie aber nicht vollständig replizieren, weil die Daten trotz unserer Rückfragen nicht archiviert und nicht alle Operationalisierungsschritte ausreichend dokumentiert sind. Daher weichen unsere Daten minimal von den Originaldaten ab, vor allem weil wir anstelle einer Fünfjahresperiode jährliche Daten verwenden.³ Kernaussage von BSS ist in Anlehnung an die Diskussionen zwischen Waltz und Deutsch/Singer der 1960er Jahre, dass die Konzentration der Machtressourcen im internationalen System einen Einfluss auf die Kriegshäufigkeit hat. Es werden zwei konkurrierende Modelle getestet: zum einen das Vorherrschafts- bzw. Stabilitätsmodell, welches besagt, dass die Kriegshäufigkeit ansteigt, sobald das Staatensystem sich von einer hohen Konzentration von Machtressourcen wegbewegt. Im Gegensatz dazu bedeutet das Paritäts- bzw. Fluiditätsmodell, dass die Kriegswahrscheinlichkeit abnimmt, wenn sich das Staatensystem von hoher Konzentration zu einem ausgeglicheneren Zustand der Verteilung von Machtressourcen bewegt. BSS verwenden drei unabhängige Variablen, um ihre abhängige Variable, Anzahl der Kriegsmonate pro Jahr, zu schätzen. Dies sind (1) die Konzentration von Machtressourcen (Konzentration), wobei Machtressourcen durch einen additiven Index gemessen wird, der aus einer demographischen, einer industriellen und einer militärischen Dimension besteht; (2) die Veränderung der Konzentration und (3) die Verschiebung der Konzentration. Während die Veränderung der Konzentration unverändert bleiben kann, da ein Machtverlust eines Staates sich in einem identischen Machtgewinn eines anderen Staates äußern kann, misst die Verschiebungsvariable die Summe aller Veränderungen.

Wir schätzen fünf Modelle. Während sich das erste Modell, eine OLS-Schätzung, auf die Vorgehensweise von BSS stützt, korrigieren die zwei nachfolgenden Schätzmethode – Prais-Winsten und Cochrane Orcutt – die Zeitreihe um das Problem der Autokorrelation. Das vierte Modell schließlich berücksichtigt, dass die abhängige Variable nicht auf einer Inter-

² Die Abkürzung steht für Generalized Autoregressive Conditional Heteroskedasticity. Bei der Entwicklung dieser Modelle hat R. Engle, der 2003 zusammen mit C. Granger den Nobelpreis für Wirtschaftswissenschaft erhielt, pionierhafte Vorarbeiten geleistet.

³ Eine solche Nachlässigkeit würde heute bei den besten Journalen nicht mehr durchgehen und die Publikation eines Artikels verhindern. Die meisten führenden Zeitschriften haben mittlerweile eine strenge Replikationspolitik eingeführt und verlangen die Publikation der Daten, die für eine Untersuchung verwendet wurden. Siehe Bueno de Mesquita et al. 2003.

vallskala gemessen wird, sondern eine Häufigkeit auszählt. Wir verwenden als Häufigkeitsmodell ein Poisson-Modell. Zusätzlich rechnen wir als letztes Modell eine Poisson-Regression, die für Autokorrelation korrigiert.

Tabelle 1: Modelle zur Erklärung der Anzahl der Kriegsmonate in einem Untersuchungsjahr von 1816 bis 2001.

	(1) OLS	(2) Prais- Winsten	(3) Cochrane- Orcutt	(4) Poisson Regression	(5) Poisson- Zeitreihen- modell
Konzentration von Machtressourcen	6.694 (9.030)	15.904 (17.628)	18.918 (18.042)	2.079 (1.016)**	3.490 (2.207)
Veränderung der Konzentration	56.698 (23.322)**	23.217 (15.931)	21.529 (16.091)	5.880 (1.935)***	0.857 (4.203)
Verschiebung der Konzentration	117.152 (27.989)***	-9.632 (23.987)	-10.686 (24.058)	22.093 (2.512)***	9.878 (5.543)*
Autoregressiver Term R1					0.585 (0.055)***
Konstantglied	-1.416 (3.767)	-3.485 (7.604)	-4.634 (7.744)	-0.153 (0.430)	-0.559 (0.931)
Beobachtungen	185	185	184	185	184
Angepasstes R-Quadrat	0.15	0.01	0.02	0.10 ¹	0.46
Durbin Watson	0.79	1.79	1.79		1.628
Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%					
¹ Pseudo R ²					

Die OLS-Regression zeigt, dass die Vorzeicheninterpretation der Koeffizienten eher für das Paritäts- bzw. Fluiditätsmodell sprechen. So folgt einem hohen Wert von Konzentration der Machtressourcen eine hohe Anzahl von Kriegsmonaten. Zieht man nun aber den geringen Wert von R², dem Anteil an erklärter Varianz, in Betracht, der bei 0.15 liegt, ist dieses Modell demnach wenig aussagekräftig. Zudem zeigt der geringe Wert der Durbin Watson-Statistik, dass in den Daten Autokorrelation eine große Rolle spielt. Als Faustregel gilt, dass der Durbin Watson-Koeffizient nahe bei 2.0 liegen muss, damit wir von einem Fehlen von Autokorrelation ausgehen können. Dies wird durch die beiden Verfahren Prais-Winsten und Cochrane-Orcutt eindeutig verbessert. Auch hier sprechen die Daten eher für das Paritäts- bzw. Fluiditätsmodell. Da wiederum ein sehr geringer R²-Wert vorliegt und keiner der Koeffizienten einen signifikanten Einfluss ausübt, stellt sich aber die Frage, ob das Modell richtig spezifiziert wurde. Es ist anzunehmen, dass wichtige erklärende Variablen fehlen, sodass sowohl die Effizienz und die Erwartungstreue der Resultate leiden. Aus der Sicht der heutigen Methodenforschung wäre es ferner angebracht, ein geeigneteres Modell als eine lineare Regression zu verwenden. Wie erwähnt berücksichtigen sog. Häufigkeitsmodelle (event

count), dass die abhängige Variable nicht auf einer Intervallskala gemessen wird, sondern eine Häufigkeit auszählt. Darum haben wir zusätzlich noch zwei Poisson-Modelle geschätzt. Das zweite Poisson-Modell korrigiert zudem noch für Autokorrelation, was, wie in den vorherigen Modellen deutlich wurde, eine problematische Rolle in den zugrunde liegenden Daten spielt. Die Vorzeicheninterpretation spricht auch in diesen beiden Modellen eher für das Paritäts- bzw. Fluiditätsmodell. So geht mit einem hohen Wert von Konzentration der Machtressourcen eine größere Wahrscheinlichkeit für eine hohe Anzahl von Kriegsmonaten einher. Das zweite Poisson-Modell, welches für Autokorrelation korrigiert, weist zudem als einziges Modell einen halbwegs akzeptablen Wert für das angepasste R-Quadrat auf. Trotzdem muss gesagt werden, dass bei allen Modellen, die Konfidenzintervalle so groß sind, dass eine vertrauenswürdige Interpretation der Ergebnisse nicht möglich ist. Bei einigen Koeffizienten kann aufgrund dieser „Großzügigkeit“ nicht ausgeschlossen werden, dass die Koeffizienten auch das umgekehrte Vorzeichen haben könnten.

Bias: Ein Schätzer ist dann nicht erwartungstreu, wenn die Fälle, die untersucht werden, nicht zufällig ausgewählt sind, sondern unter einem systematischen Selektionsbias leiden. In der politikwissenschaftlichen Methodenlehre ist dieses Problem früh durch Achen (1987) diskutiert worden. Unter einem Selektionsbias haben in der Internationalen Politik sowohl qualitative wie quantitative Analysen zur Effektivität der militärischen Abschreckung gelitten, wie die Aufsätze von Achen und Snidal (1989) und Fearon (1994) verdeutlichen. Bei den von diesen Forschern kritisierten Untersuchungen bestand die Verzerrung insofern, als sie sich nur auf Fälle gescheiterter Abschreckung oder auf militärische Krisen bezogen. Da die Episoden erfolgreicher Abschreckung, in denen ein potentieller Herausforderer den Status quo akzeptiert, ausgeschlossen sind, entsteht eine systematische Verzerrung. Die Berücksichtigung der gesamten Varianz ist oft nicht ausreichend, um das potentielle Problem des Selektionsbias in den Griff zu kriegen. So lässt sich etwa der Erfolg von Sanktionen nicht verlässlich analysieren, wenn als Grundlage der Untersuchung nur Fälle dienen, in denen ein Staat oder eine Staatengruppe gegenüber einem Land oder einer Regierung dieses außenpolitische Instrument gebraucht. Die Sanktionsfälle sind keine Zufallsstichprobe sämtlicher möglicher Situationen, in denen die Verhängung einer Sanktion möglich schien. Die Verzerrung rührt unter anderem daher, dass die mächtigen Schurken unter Umständen einer Sanktion entgehen, weil sie selber über ein glaubwürdiges Drohpotential verfügen, während schwache Sünder nicht die Kraft haben, eine Sanktion abzuwenden. Wenn sich nun die Analyse nur auf die tatsächlichen und nicht auf die potentiellen Sanktionen bezieht, wird die Wirkung des Instrumentes Sanktion systematisch überschätzt. Zur Analyse solcher Selektionsprozesse hat der Nobelpreisträger James Heckmann spezielle Selektionsmodelle entworfen, die vor allem bei intervallskalierten und bei binären abhängigen Variablen gebräuchlich sind. Nooruddin (2002) bietet etwa eine Studie zum Erfolg von Sanktionen, bei denen der Erfolg als Dummyvariable operationalisiert ist. Das verwendete Heckmann-Probit-Modell ist eine Spezialform von sog. Probitregressionsmodellen, bei denen die abhängige Variable – wie erwähnt – kategorial definiert ist. Bei einer einfachen Probitregression zeigt sich etwa, dass der Sanktionserfolg zunimmt, wenn die Kosten der Sanktion wachsen. Doch dieses Modell ist, wie erwähnt, der Datenstruktur nicht angemessen. Nooruddin (2002) zeigt im Vergleich des einfachen Probit- zum Heckmann-Probit-Modell,

dass sich die Wirkung der Kostenvariablen halbiert, wenn der Selektionseffekt berücksichtigt ist.

3 Aktueller Forschungsstand und weitere Entwicklungen

Die Herausforderungen in der Methodendiskussion in der Internationalen Politik bestehen darin, bessere Schätzer zu entwickeln. Wichtig ist dabei auch immer, dass die Grundgesamtheit der Fälle richtig definiert ist. Viele Studien haben gerade in der Konfliktforschung in den letzten Jahrzehnten das sog. Dyadenjahr als Untersuchungseinheit für ihre Explorationen zum „Demokratischen Frieden“ verwendet. Staatenpaare statt einzelne Länder in den Vordergrund der Analyse zu rücken bietet den Vorteil, dass sich die Fallzahl erhöht. Dennoch besteht die Gefahr, dass das Konfliktrisiko in Zeiten der zwischenstaatlichen Anspannung eher unterschätzt wird, weil sich die Analyse auch auf Fälle wie das Staatenpaar Neuseeland-Österreich erstreckt, für das die Kriegswahrscheinlichkeit wohl seit Dekaden 0 ist. Russett und Oneal (z.B. 2001) haben dies insofern erkannt, als sie nur sog. politisch relevante Dyaden in die Analyse einbeziehen – das sind benachbarte Staaten oder Staatenpaare, in denen mindestens ein Staat eine Großmacht ist. Der Vorteil einer solchen Beschränkung kommt aber mit einem gewaltigen Nachteil daher – es werden unter anderem Dyadenjahre ausgeschlossen, in denen ein militärischer Konflikt auftrat. Dass so Bias entsteht, ist offensichtlich. Bennett und Stam (2003) beziehen ihre vergleichende Modellevaluation deshalb auf sämtliche Staatenpaare. Konsequenz dieser Vorgehensweise sind sehr große Datensätze, die sich zur Überprüfung der bestehenden Forschung eignen, kaum jedoch zusätzliche Variablen berücksichtigen können, da die Kodierungskosten immens wären. Den Einbezug von zusätzlichen Erklärungsfaktoren erlauben nach Ansicht von King und Zeng (z.B. 2001) viel mehr Forschungsdesigns, in denen die Kriegsfälle mit einer relativ kleinen, dafür aber sorgfältig konstruierten Kontrollgruppe von Nicht-Kriegsfällen verglichen werden. Eine weitere methodische Innovation ergibt sich über das Erfordernis, die Modelle auch praktisch für Prognosen zu verwenden, wie das in den USA unter anderem über das staatlich finanzierte Projekt zur Vorhersage von Staatskollapsen (state failure) geschehen ist. Konventionelle Ansätze verwenden zur Erklärung von solchen Phänomenen zumeist logistische Regressionen. Dies wirft zum einen das Problem auf, dass der Standardansatz sich nicht unbedingt zur Analyse von seltenen Ereignissen eignet, zu denen der Zerfall von Staaten oder große Kriege gehören (King/Zeng 2001). Bezogen auf die Prognosegüte ergibt sich ferner, dass einfache logistische Modelle nur das für die Beratung weniger relevante Phänomen vorhersagen, die Jahre nämlich, in dem Frieden in einer Dyade herrschte. Konventionelle Verfahren haben umgekehrt Mühe, die eigentlich interessierenden Kriegsjahre zu prognostizieren. Hier sind sie deutlich anderen Verfahren unterlegen wie den neuronalen Netzwerken, wie Beck, King und Zeng (2000, siehe auch 2004) deutlich zeigen. Neuronale Netzwerke sind Techniken, die in der Künstlichen Intelligenz entwickelt wurden. In der Datenanalyse finden sie vor allem beim induktivistischen „*data mining*“ Verwendung, d.h. der Suche nach Struktur mit Hilfe von komplexen Algorithmen bei einer Fülle von Informationen. Der Vorteil von neuronalen Netzwerken ist es, dass sie flexibler sind als Standardregressionsverfahren, die für die Beziehung zwischen den unabhängigen Variablen und der

abhängigen Variablen eine feste Funktion vorgeben – im Falle von Logit ist das eine Funktion, die s-förmig verläuft. Dies bedeutet aber, dass bei einer äquivalenten Änderung eines Kovariats wie Demokratie die Kriegsgefahr in so unterschiedlichen Dyaden wie Indien-Pakistan und Deutschland-Frankreich im Jahr 2004 ähnlich groß ausfällt, obgleich anzunehmen ist, dass eine Veränderung im Demokratiegrad in Deutschland oder Frankreich heute wohl kaum das Kriegsrisiko ähnlich beeinflussen würde wie die gleiche Änderung im asiatischen Staatenpaar. Neuronale Netzwerke sind nun als Prognoseverfahren konventionellen Logit- oder Probitansätzen insofern überlegen, als sie die simultane Schätzung unterschiedlicher Beziehungen gleichzeitig zulassen.

Es mag der Einwand aufkommen, dass Prognose allein nicht das Ziel der empirischen Sozialwissenschaft sein kann. Aus diesem Grund sind auch die Anstrengungen wichtig, die es erlauben, rigorose theoretische Modelle direkt empirisch zu schätzen und damit die Relevanz von innovativen Erklärungen zu prüfen. Die übliche Vorgehensweise außerhalb der experimentellen Forschung ist es, dass die aus der formalen Theorie abgeleiteten Modelle mit Standardverfahren geschätzt werden, wie dies etwa Fearon (1994) für sein einflussreiches Krisenverhandlungsmodell getan hat. Dies ist aus verschiedenen Gründen fragwürdig: Erstens ist ein statistisches Modell wiederum an zusätzliche Annahmen geknüpft, die mit den theoretischen Aussagen darüber, wie die Daten entstanden sind, durchaus in Widerspruch stehen können (Morton 1999). Zweitens lassen sich aus den spieltheoretischen Modellen oft deterministische Prognosen ableiten. Im Falle des Gefangenendilemma läuft das auf die Vorhersage hinaus, dass sich die Akteure mit Wahrscheinlichkeit 1 nicht-kooperativ verhalten werden. Zu einem Test dieser Hypothese passen aber nicht die probabilistischen Modelle, die üblicherweise Verwendung finden. Drittens ist bei der Überprüfung eines strategischen Modells damit zu rechnen, dass die Handlungen der Akteure sich gegenseitig bedingen und dass die Untersuchung diese Interdependenz berücksichtigen muss. Ein einfaches Logit- oder Probit-Modell beschränkt die Analyse viertens auf den letzten Ast des Spielbaums, die Entscheidung für oder gegen die Kriegsoption. Nicht berücksichtigt werden dabei die Züge, die vorangegangen sind. Signorino (1999, 2003, siehe auch Lewis/Schultz 2003) zeigt, dass die Standardverfahren nicht effizient sind und einen Bias aufweisen. Signorino benutzt ein statisches Verfahren, das die Spielstruktur einer militärischen Krise berücksichtigt. Zu beachten ist allerdings hier, dass natürlich nicht für die Analyse jedes Phänomens ein neues Schätzverfahren gesucht werden sollte, da dies die Transparenz des Forschungsprozesses deutlich verringern und das Gebot unterminieren würde, Theorie und Schätzansatz auseinander zu halten. Zugleich ist es wohl nicht sinnvoll, barocke Modelle, die nicht dem Grundsatz des „*non-fat modeling*“ genügen wollen, 1 zu 1 testen zu wollen. Achen (2002) empfiehlt als Regel sogar, die Zahl der erklärenden Variablen auf drei zu reduzieren – dies aber immer auf der Basis eines klaren, möglichst mathematisierten Hypothesenfundaments. Die klare Verbindung zwischen theoretischen und statistischen Modellen wird es in Zukunft auch erlauben, vermehrt vergleichende Tests von theoretischen Modellen vorzunehmen, wie dies jüngst Bennett und Stam (2003) für die Kriegsursachenforschung sowie Thomson et. al. (2004) und Schneider et al. (2004) für die Analyse von Entscheidungsprozessen in der Europäischen Union getan haben.

In der qualitativen Politikforschung ist interessanterweise eine ähnliche Bewegung hin zur direkten Überprüfung der Theorie zu beobachten. Ein Problem besteht hier allerdings

darin, dass die Theorien oft komplex sind. Eine Möglichkeit zur Explizierung solcher Zusammenhänge besteht darin, mit Hilfe der Booleschen Logik die Variablen logisch miteinander zu verknüpfen. Braumoeller (2003) hat dazu einen Schätzer entwickelt, der bei quantitativen Untersuchungen mit geringem n zum Einsatz kommen kann. Zugleich sind die Theorien aber oft auch auf einer Makroebene angesiedelt und deterministisch. Um die Hypothesen zu widerlegen, genügt deshalb bereits eine einzelne Fallstudie mit divergierenden Ergebnissen (Lieberson 1991). Eine subjektivistische Art und Weise der Theorieprüfung bieten bayesianische Modelle, die zunehmend den Weg in die Politikwissenschaft finden (Western 1996, Gill 2004). Dieser Ansatz verbindet die unbeobachtbaren Daten mit einer a priori-Wahrscheinlichkeit, die einer substantiellen Theorie entnommen sein können oder schlicht auch auf der Erfahrung des Forscher – sprich: seinem Vorurteil – beruhen können. Dieses „Wissen“ wird dann mit Hilfe der Regel von Bayes und aufgrund der beobachteten Daten in eine posteriore Einschätzung darüber verwandelt, wie der Datengenerierungsprozess tatsächlich verlaufen ist. Der Reiz dieses Ansatzes besteht in den Sensitivitätsanalysen. Sie erlauben es, die subjektive Einschätzung der Wirklichkeit gezielt mit der Realität zu vergleichen. Damit ist das „Fata Morgana“-Problem natürlich nicht gelöst, aber der Weg hin zur Etablierung von Zusammenhängen wird transparenter.



Literatur

- Achen, Christopher H. 1987: *The Statistical Analysis of Quasi-Experiments*, Berkeley, CA.
- Achen, Christopher H./Snidal, Duncan 1989: Rational Deterrence Theory and Comparative Case Studies, in: *World Politics* 41, 144-169.
- Achen, Christopher H. 2002: Toward a New Political Methodology: Microfoundations and ART, in: *Annual Review of Political Science* 5: 1, 423-450.
- Beck, Nathaniel/King, Gary /Zeng, Langche 2000: Improving Quantitative Studies of International Conflict: A Conjecture, in: *American Political Science Review* 94: 1, 21-36
- Beck, Nathaniel 2001: Time-Series Cross-Section Data: What Have We Learned in the Past Few Years?, in: *Annual Review of Political Science* 4: 1, 271-293.
- Beck, Nathaniel/Katz, Jonathan 2001: Throwing Out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon, in: *International Organization* 55: 2, 487-498.
- Beck, Nathaniel/King Gary/Zeng, Langche 2004: Theory and Evidence in International Conflict: A Response to de Marchi, Gelpi, and Grynaviski, in: *American Political Science Review* 98: 2, 379-389.
- Bennett, D. Scott/Stam, Allan C. 2003: *The Behavioral Origins of War*, Ann Arbor, MI.
- Bueno de Mesquita, Bruce/Gleditsch, Nils/Petter, James Patrick/King, Gary/ Metelits, Claire/ Ray, James Lee/Russett, Bruce/ Strand, Håvard/, Brandon; Valeriano 2003: Symposium on Replication in International Studies Research, in: *International Studies Perspectives* 4: 1, 72-107.
- Box-Steffensmeier, Janet M./Jones, Bradford S. 2004: *Event History Modeling. A Guide for Social Scientists*, Cambridge.
- Braumoeller, Bear F. 2003: Causal Complexity and the Study of Politics, in: *Political Analysis* 11: 3, 209-233.
- Fearon, James 1994: Signalling versus the Balance of Power and Interests: an Empirical Test of a Crisis Bargaining Model, in: *Journal of Conflict Resolution* 38: 2, 236-269.
- Gill, Jett 2004: Special Issue on Bayesian Methods, in: *Political Analysis* 12: 4, 323-443.

- Hoff, Peter D./Ward, Michael D. 2004: Modeling Dependencies in International Relations Networks, in: *Political Analysis* 12: 2, 160-175.
- 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, NJ.
- King, Gary/Langche, Zeng 2001: Explaining Rare Events in International Relations, in: *International Organization* 55: 3, 693-715.
- Leblang, David/Mukherjee, Bumba 2004: Presidential Elections and the Stock Market: Comparing Markov-switching and Fractionally Integrated GARCH Models of Volatility, in: *Political Analysis* 12: 3, 296-322.
- Lewis, Jeffrey B./Schultz, Kenneth A. 2003: Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information, in: *Political Analysis* 11: 4, 345-367.
- Liebersohn, Stanley 1991: Small Ns and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases, in: *Social Forces* 70, 307-320.
- Morton, Rebecca 1999: *Methods and Models: a Guide to the Empirical Analysis of Formal Models in Political Science*, Cambridge.
- Russett, Bruce M./Oneal, John R. 2001: *Triangulating Peace: Democracy, Interdependence, and International Organizations*, New York, NY.
- Nooruddin, Irfan 2002: Modeling Selection Bias in Studies of Sanctions Efficacy, in: *International Interactions* 28: 1, 59-75.
- Schneider, Gerald/Finke, Daniel/Bailer, Stefanie 2004: *Bargaining Power in the European Union: An Evaluation of Competing Game-Theoretic Models*, unv. Ms.
- Schneider, Gerald/Tröger, Vera 2004: *The Winners and Losers of War: Stock Market Effects of Armed Conflict, 1990-2000* (Working Paper des Juan March Institute), Madrid.
- Signorino, Curtis S. 1999: Strategic Interaction and the Statistical Analysis of International Conflict, in: *American Political Science Review* 93: 2, 279-297.
- Signorino, Curtis S. 2003: Structure and Uncertainty in Discrete Choice Models, in: *Political Analysis* 11: 4, 316-344.
- Singer, J. David/Bremer, Stuart/Stuckey, John 1972: Capability Distribution, Uncertainty, and Major Power War, 1820-1965, in: Russett, Bruce (Hrsg.): *Peace, War, and Numbers*, 1. Auflage, London, 19-48.
- Thomson, Robert/Frans N. Stokman/Christopher H. Achen/Thomas König 2004: *The European Union Decides*, Cambridge.
- Western, Bruce 1998: Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach, in: *American Journal of Political Science* 42: 4, 1233-1259.