

Die Mehrebenenanalyse als ein Baustein zur Analyse politischer Vielfalt

Sebastian Jäckle, Universität Freiburg

Julian Schärdel, Universität Freiburg

**Papier für das Panel Vielfalt vergleichend untersuchen?! Entwicklung von Fragen, Themen und Methoden in der vergleichenden Politikwissenschaft (Thema 29)
3-Länder-Tagung, Innsbruck (19.-21.09.2013)**

Work in progress – Bitte nur nach Absprache mit den Autoren zitieren

Kontakt:

Dr. Sebastian Jäckle
Seminar für Wiss. Politik
Werthmannstr. 12
79098 Freiburg i. Br.
e-mail: sebastian.jaeckle@politik.uni-freiburg.de

Julian Schärdel MA
Seminar für Wiss. Politik
Werthmannstr. 12
79098 Freiburg i. Br.
e-mail: julian.schaerdel@politik.uni-freiburg.de

1. Einleitung

Angesichts anhaltender Globalisierungstendenzen, einer EU-Integration, die keineswegs als abgeschlossen bezeichnet werden kann sowie einer immer stärker werdenden postmodernen Individualisierung, die sich über neue Kommunikationskanäle wie Twitter oder Facebook ihren Weg auch wieder zurück auf die gesellschaftliche Ebene hin bahnt und hierdurch Politik beeinflusst, stellt sich die Frage, wie die Politikwissenschaft mit solchen und ähnlichen Phänomenen umgehen kann. In diesem Papier wird argumentiert, dass die Mehrebenenanalyse (MEA)¹ hier oftmals eine sehr gute Wahl darstellt. Die MEA besitzt aufgrund ihrer enormen Flexibilität das Potential eine Vielzahl für unsere Disziplin relevanter Fragestellungen adäquat anzugehen. Inwiefern die Politikwissenschaft dieses Potential aktuell bereits ausschöpft soll im Folgenden untersucht werden.

Um die bisherige Verwendung sowie Chancen und Perspektiven des Einsatzes von MEA sinnvoll bewerten zu können, sollen in einem ersten Schritt die statistischen Grundlagen dieser Regressionstechnik vorgestellt werden. Dabei wird insbesondere auf die benötigte Datenstruktur, die unterschiedlichen Modellierungsstrategien aber auch mögliche Gefahrenquellen und Fallstricke eingegangen, die es bei der MEA zu beachten gilt. Im Zentrum des Papiers steht anschließend eine Meta-Analyse sämtlicher MEA-Artikel, die im Zeitraum 1990-2013 in vom Social Science Citation Index (SSCI) gerankten, politikwissenschaftlichen Zeitschriften veröffentlicht wurden. Ein in diesem Zusammenhang durchgeführter Vergleich mit unserer Nachbardisziplin der Soziologie offenbart, dass die Politikwissenschaft was die Anwendung mehrebenenanalytischer Verfahren anbelangt noch im Hintertreffen ist, obgleich sie insbesondere in den letzten zehn Jahren deutlich Boden gut gemacht hat. Gleichzeitig zeigt eine thematische Auswertung, dass die politikwissenschaftliche Forschung MEA bislang hauptsächlich für einige wenige, spezifische Fragenkomplexe heranzieht und damit das ganze Potential dieser Methode bei weitem noch nicht ausschöpft. Daher werden im letzten Kapitel unter Rückgriff auf eine qualitative Betrachtung einzelner innovativer Ansätze, die im Rahmen der Meta-Analyse aufgetaucht sind, Möglichkeiten aufgezeigt, das Instrumentarium der MEA für die Politikwissenschaft in unterschiedlichsten Bereichen und für eine Vielzahl an Fragestellungen nutzbar zu machen.

2. Statistik hinter der Mehrebenenanalyse

2.1. Grundlagen und Datenstruktur

Generell eignen sich MEA-Modelle wenn Daten in einer hierarchischen Struktur vorliegen, wie sie durch eine mehrstufige Zufallsauswahl generiert werden. Das bedeutet, dass alle zu untersuchenden Einheiten auf der untersten Ebene sich eindeutig jeweils einer einzigen Aggregateinheit auf höherer Ebene zuordnen lassen müssen, wodurch eine ineinander verschachtelte Baumstruktur entsteht. Die abhängige Variable liegt bei einer MEA immer auf der untersten Ebene (auch bezeichnet als Level-1,

¹ Eine alternative Bezeichnung, die häufig anzutreffen ist, ist die Hierarchische Lineare Modellierung (HLM).

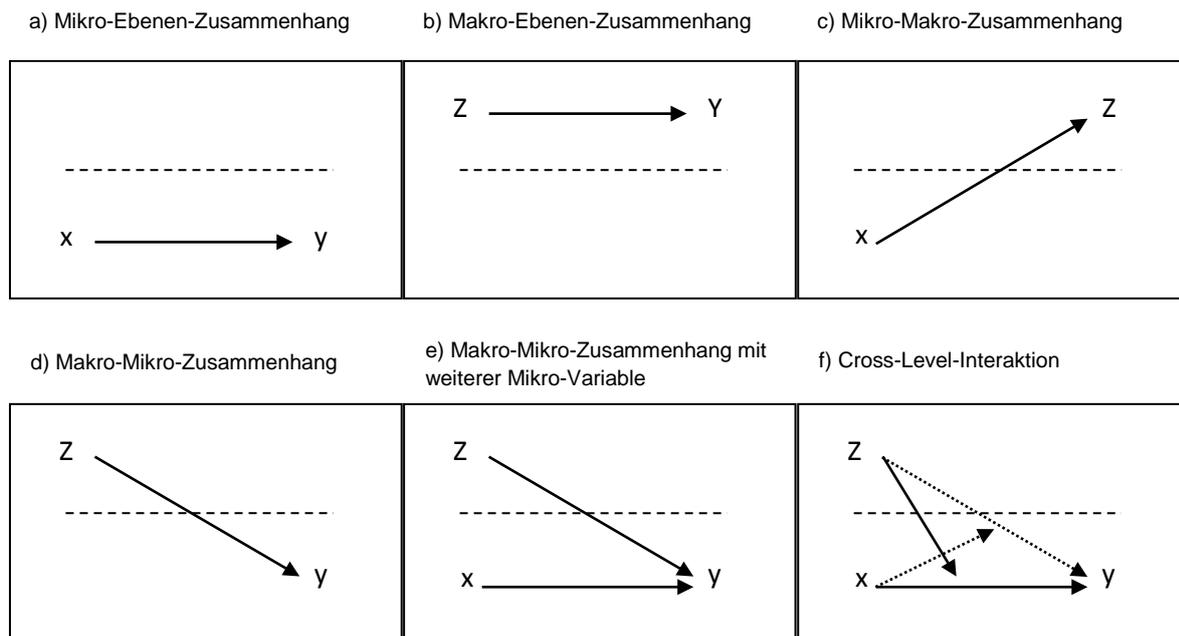
Mikro- oder Individual-Ebene), die erklärenden Variablen können dagegen sowohl auf Ebene-1 wie auch auf höheren Aggregatebenen (Level-2 oder Makroebenen) angesiedelt sein. So wird beispielsweise in der politik-soziologischen Einstellungsforschung häufig davon ausgegangen, dass Einstellungen von Personen nicht ausschließlich durch Charakteristika des Individuums selbst, sondern auch durch übergeordnete Aggregatmerkmale geprägt werden – auf eine Person also gewisse Sozialisationseffekte wirken die nicht nur sie selbst, sondern auch andere um sie herum erfahren. Bereits Ende der 1960er Jahre formulierte Allen H. Barton eine harsche Kritik an Forschungsarbeiten, die eine rein individualistische Ausrichtung verfolgten und dabei die soziale Umgebung ignorierten:

„But as usually practiced, using random sampling of individuals, the survey is a sociological meatgrinder, tearing the individual from his social context and guaranteeing that nobody in the study interacts with anyone else in it. It is a little like a biologist putting his experimental animals through a hamburger machine and looking at every hundredth cell through a microscope; anatomy and physiology get lost, structure and function disappear, and one is left with cell biology.“ (Barton, 1968: 1)

So kann sich zum Beispiel eine nationalstaatliche Kultur (Level-2) prägend auf die Einstellungen eines Individuums (Level-1) auswirken. Die betrachteten Ebene-1-Einheiten müssen dabei trotz der Bezeichnung *Individualebene* nicht zwangsläufig Individuen im Sinne von einzelnen Personen sein, sondern sie können auch selbst Aggregatniveau besitzen. Einzig relevant ist, dass es noch mindestens eine höhere Ebene gibt, in die sie sich einordnen lassen. Beispielsweise könnten die Kommunen in Deutschland auf diese Weise als Individuen begriffen werden, welche sich in die 41 Regierungsbezirke auf Aggregatebene einsortieren lassen. Generell lassen sich beliebig viele Ebenen im Rahmen einer MEA modellieren, sofern eine klare hierarchische Struktur gegeben ist. Wir werden uns in der folgenden Methodendeskription jedoch auf den einfachsten Fall einer Zwei-Ebenen-Analyse konzentrieren. Dessen grundlegende Logik lässt sich dann auch problemlos auf Modelle mit drei oder noch mehr Ebenen übertragen.

Um ein Verständnis dafür zu entwickeln, welche Kausalzusammenhänge überhaupt mittels einer Mehrebenenregression modellierbar sind, bietet es sich zunächst an, sich alle potentiell möglichen Effekte zu vergegenwärtigen, die auf bzw. zwischen Mikro- und Makro-Ebene vorkommen können. Die folgenden Schaubilder stellen diese schematisch in Anlehnung an Snijders und Bosker (2012: 10-12) dar – Großbuchstaben stehen dabei für Aggregatmerkmale, Kleinbuchstaben für Individualmerkmale.

Abbildung 1a – 1f: Zusammenhänge zwischen Mikro- und Makro-Ebene



Die ersten beiden Abbildungen (1a und 1b) zeigen reine Mikro- bzw. Makro-Ebenen Zusammenhänge. Bei ersterem spielt die Makro-Ebene nur über die Auswahl der Stichprobe mit in die Analyse hinein. Wählt man beispielsweise mittels einer Zufallsauswahl 10 europäische Länder aus, in denen man dann auf Basis von Umfragedaten den Effekt des Alters (x) auf die Einstellung zur EU (y) untersucht, so erhält man unverzerrte Ergebnisse wenn sich der Effekt, den das Alter ausübt, nicht zwischen den einzelnen Ländern unterscheidet. Wenn dem jedoch nicht so ist, würde eine reine Mikro-Modellierung immer ein gewisses Maß an unerklärtem Rauschen mit sich bringen. Nur mit einem ME-Design ließe sich dieser Varianzanteil adäquat erklären: denn auch ganz ohne einen erklärenden Faktor auf der Makroebene in das Modell einfließen zu lassen berücksichtigt man so doch die hierarchische Struktur der Daten. Bei Makro-Ebenen-Zusammenhängen ist ein ME-Design dann überflüssig, sofern es sich bei Z und Y um zwei wirkliche Aggregatmerkmale handelt. Werden die betrachteten Merkmale aber auf der Mikroebene beobachtet und erst durch Aggregation zu Aggregatmerkmalen, dann muss man die der Stichprobe eigentlich inhärente mehrstufige Zufallsauswahl beachten. Will man beispielsweise den Einfluss der Arbeitslosenrate in Städten auf das wirtschaftliche Wachstum dort untersuchen, wobei die Arbeitslosenrate als Aggregat eines auf der Mikro-Ebene gemessenen Merkmals zu betrachten ist, dann könnte eine einfache Zufallsauswahl der Städte zu verzerrten Ergebnissen führen. Abbildung 1c stellt einen Effekt der Mikro- auf die Makroebene dar. Ein solcher Effekt, bei dem die abhängige Variable auf der höheren Ebene liegt, lässt sich mit einem ME-Design nicht testen.

Anders dagegen die drei folgenden Schaubilder, die allesamt klassische Anwendungsfelder der MEA darstellen. Beim einfachsten Fall eines Makro-Mikro-Zusammenhangs (Abbildung 1d) wirkt sich eine Variable auf der Aggregatebene auf eine abhängige Variable auf Individualebene aus. Zusätzlich zu

dem direkten Effekt der Makro-Ebene können auch weitere direkte Individualeffekte (x) dazukommen – zum Beispiel Alter oder Geschlecht der Person (Abbildung 1e). Auch gibt es die Möglichkeit, dass die Z -Variable nur dann einen Einfluss ausübt, wenn die x -Variable einen bestimmten Wert annimmt. Dies wäre beispielsweise der Fall, wenn sich nur Personen mit einem geringen Einkommen in ihrer Einstellung zum Thema Atomkraft vom nationalen Strompreis beeinflussen lassen würden, bei allen anderen würde Z hier keinen Einfluss ausüben. Variablen können aber auch einen indirekten, d.h. konditionierenden Einfluss ausüben und zwar auf den Effekt, den eine weitere, auf einer anderen Ebene angesiedelte Variable hat. Dies bezeichnet man als Cross-Level-Interaktion (Abbildung 1f). Beispielsweise könnte der Effekt, den das Alter einer Person (x) auf deren Einstellung zum Thema Atomenergie (y) hat vom Postmodernismusgrad der Gesellschaft (Z) konditioniert werden. Die Cross-Level-Interaktion kann jedoch auch andersherum mit der Mikro-Variable als konditionierendem Effekt gedacht werden (gepunktete Pfeile): In diesem Fall würde der Effekt, den der Postmodernismusgrad der Nation auf die Einstellung des Individuums zur Atomenergie ausübt dadurch konditioniert, wie alt die betreffende Person ist.

Liegen hierarchische Daten vor, gibt es neben der MEA noch drei weitere Strategien diese zu analysieren: die Aggregation der Mikro-Daten auf Makro-Level, die Disaggregation der Makro-Level Variablen auf die Mikro-Ebene sowie die Berechnung separater Regressionsmodelle für alle Level-2-Gruppen. All diese Verfahren weisen jedoch spezifische Probleme auf. Werden Individualdaten auf eine höhere Ebene aggregiert (beispielsweise durch Mittelwertbildung) um dort dann als Aggregatdaten in eine Regression einzufließen, lassen sich erstens die aggregierten Daten nur mehr auf der Aggregatebene interpretieren (Gefahr eines ökologischen Fehlschlusses), zweitens kann die Aggregation dazu führen, dass real vorhandene Gruppenunterschiede nicht abgebildet werden und drittens sind Cross-Level-Interaktionen nicht mehr sinnvoll abprüfbar (T. Snijders & Bosker, 2012: 14-16). Durch eine Disaggregation der Makro-Level-Variablen hingegen vergrößert sich, sofern man auf Unterschiede zwischen den Gruppen testet, die Gefahr, einen Fehler erster Art zu begehen – d.h. die Nullhypothese fälschlicherweise zurückzuweisen.² Gleichzeitig führen disaggregierte Daten oftmals dazu, dass statistische Tests zu konservativ sind, wenn Unterschiede innerhalb der Gruppen im Fokus stehen. Zudem kann, selbst wenn sämtliche Variablen auf der Individualebene vorliegen, ein möglicher Klumpeneffekt dazu führen, dass Regressionen über disaggregierte Daten eine geringere Sicherheit aufweisen als dies bei einer einfachen einstufigen Zufallsauswahl der Fall wäre (T. Snijders & Bosker, 2012: 16-17). Eine dritte Option wäre es, separate Regressionen innerhalb aller Gruppen durchzuführen. Allerdings wird es umso schwieriger die geschätzten Parameter zu vergleichen, je

² Technisch gesprochen führt eine Disaggregation der Daten zu einer rein artifiziellen Vergrößerung der Observationszahl, da Einheiten, die eigentlich zusammenhängen, als komplett unabhängig gewertet werden. Die eigentliche Fallzahl ändert sich dadurch aber nicht. Bei der Schätzung der Modelle wird für die Berechnung der Standardfehler nun jedoch die künstlich erhöhte Observationszahl herangezogen. Entsprechend unterschätzen Modelle, die auf Basis von disaggregierten Daten berechnet werden generell die Standardfehler und überschätzen somit die Signifikanzen.

mehr Gruppen und damit Regressionsmodelle man hat. Zudem lassen sich so Makro- sowie Cross-Level-Effekte nicht sinnvoll testen (Pötschke, 2006: 174).

2.2. Wann ist ein Mehrebenenmodell statistisch notwendig bzw. zumindest sinnvoll?

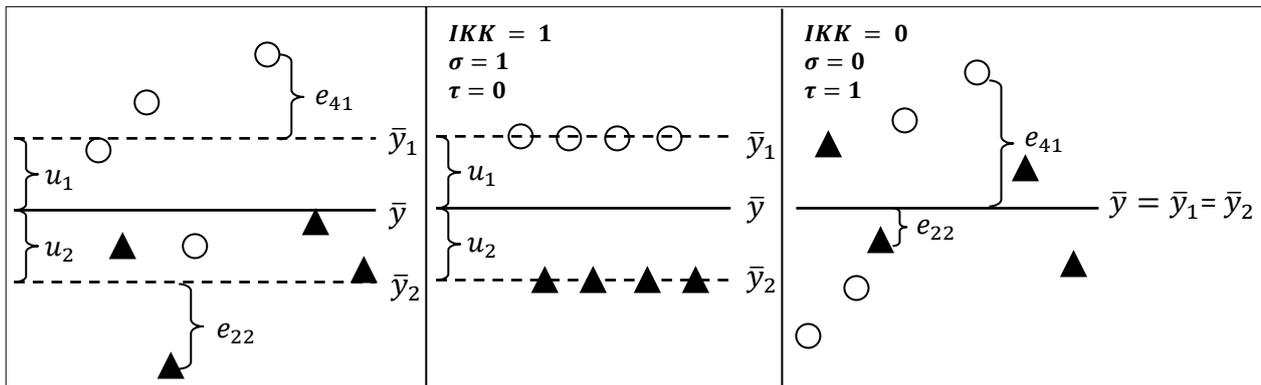
Generell lässt sich sagen, dass sobald die gesamte Varianz zwischen den betrachteten Individuen zu einem relevanten Anteil auf Unterschiede zwischen den Makro-Level-Einheiten zurückzuführen ist, es Sinn ergibt, dies aktiv über eine MEA zu modellieren.³ Der Intraklassenkorrelationskoeffizient (IKK) hilft bei der Beantwortung der Frage, wie groß dieser Anteil ist. Er berechnet sich als der Anteil der Gesamtvarianz der abhängigen Variable, welcher durch die Gruppierung erklärt wird.⁴ Abbildung 2 verdeutlicht den IKK anhand eines simplen Beispiels. Angenommen wir haben zwei Gruppen von je vier Individuen bei denen wir die Variable y betrachten. Der erste Index steht für die Nummer des Individuums in der Gruppe, der zweite Index für die Nummer der Gruppe. Die erste Abbildung zeigt, wie sich die Gesamtvarianz in einerseits die Abweichungen der Gruppenmittelwerte vom Gesamtmittelwert (u_1 und u_2) – diese Varianz zwischen den Gruppen bezeichnen wir mit σ – und andererseits die Abweichungen der Individuen vom Gruppenmittelwert (e_{i1} und e_{i2}) aufteilen lässt. Diese zweite Varianz innerhalb der Gruppen sei τ . Der IKK berechnet sich als: $IKK = \frac{\sigma}{\sigma + \tau}$. Haben nun alle Individuen einer Gruppe denselben Wert, dann geht die gesamte Varianz ausschließlich auf die Unterschiede zwischen den Gruppen zurück. In diesem Fall wäre $\sigma = 1$, $\tau = 0$ und entsprechend auch der IKK gleich 1. Umgekehrt ist der IKK gleich Null, wenn die beiden Gruppenmittelwerte auf den Gesamtmittelwert zusammenfallen und damit die Gesamtvarianz ausschließlich auf die Unterschiede innerhalb der Gruppen zurückgeht. Je nachdem ob man theoretisch von einer schwachen oder starken Intraklassenkorrelation ausgeht gibt Hox Richtwerte von 0,05 bzw. 0,3 für den IKK an, ab denen es statistisch notwendig sei ein ME-Modell zu rechnen (Hox, 2010: 244).

Da der IKK eine einfache Option darstellt, um zu testen inwieweit ein ME-Design aus statistischer Warte betrachtet notwendig, sinnvoll oder nur eine methodische Spielerei ohne nennenswerten analytischen Mehrwert ist, sollte dieser stets zu Beginn einer Analyse berechnet und auch berichtet werden.

³ Das Vorliegen von signifikanten Mittelwertunterschieden zwischen den Gruppen in der abhängigen Variable ist zwar ein Indiz dafür, dass eine MEA sinnvoll sein kann, aber als Begründung diese Methode zu verwenden doch ungeeignet. Theoretisch ist es auch möglich, dass sich diese Unterschiede zu 100 Prozent über die Charakteristika der Individuen erklären lassen. Erst wenn ein gewisser Anteil an der Gesamtvarianz nicht alleine über die Mikro-Variablen erklärt werden kann – von denen angenommen wird, dass sie in allen Gruppen exakt denselben Effekt aufweisen – ist eine statistische Begründung für die Verwendung von MEA gegeben. Aus diesem Grund sollte für die Entscheidung, ob eine MEA oder eine normale Regression berechnet wird auf den IKK zurückgegriffen werden.

⁴ Alternativ kann man den IKK auch als die Korrelation von zwei zufällig ausgewählten Individuen innerhalb einer zufällig ausgewählten Gruppe begreifen. Je höher diese Intraklassenkorrelation, desto größer ist der Anteil der Gesamtvarianz, der auf die Unterschiede zwischen den Gruppen zurückzuführen ist.

Abbildung 2: Aufteilung der Gesamtvarianz und IKK



Eigene Darstellung nach Steele (2009: 7)

2.3. Wie viele Fälle braucht man für eine MEA?

Das Vorhandensein einer ausreichenden Fallzahl stellt bei MEA wie auch bei anderen Regressionsmethoden einen limitierenden Faktor dar. Nach Snijders ist für die Analyse von Level-1-Effekten primär die Gesamtzahl der Level-1-Einheiten von Relevanz (T. A. B. Snijders, 2005). Da die in der Politikwissenschaft analysierten Datensätze oftmals vergleichsweise viele Level-1-Einheiten enthalten, dürfte es hier nur relativ selten zu Problemen kommen. Problematischer dürfte sich zumeist die Anzahl an Level-2-Gruppen gestalten, die Snijders als den am stärksten limitierenden Faktor für ME-Designs betrachtet – die durchschnittliche Gruppengröße erachtet er hingegen als nicht so wichtig für die Stärke des Modells (T. A. B. Snijders, 2005). Allerdings ist in der Forschung durchaus umstritten wie viele Level-2-Einheiten benötigt werden, um aussagekräftige Ergebnisse mittels einer MEA zu erzielen: Diverse Daumenregeln nennen Größen von 30 bis 100 Makro-Einheiten (für einen Überblick s. (Braun et al., 2010)). Allerdings zeigt ein Blick in die empirische Forschungslandschaft, dass auch mit deutlich weniger Level-2-Einheiten ME-Modelle gerechnet werden (Peffley & Rohrschneider, 2003; Rosar, 2003). Insbesondere, wenn die Schätzung mittels Restricted Maximum Likelihood (REML) durchgeführt wird, seien auch Analysen mit relativ wenigen Level-2-Gruppen möglich (Browne & Draper, 2000; Maas & Hox, 2004). Allerdings würde die Inklusion von vergleichsweise vielen auf Level-2 messenden Erklärungsvariablen bei gleichzeitig relativ wenigen Level-2-Gruppen zu einem Problem mit der Anzahl der Freiheitsgrade führen. Nach Hox erhöht sich bei einer Gruppenanzahl von 24-30 das operationale Alpha-Level, das den realen Anteil an Fehlern erster Art (Fehler für das fälschlicherweise Zurückweisen der Nullhypothese) angibt, von den eigentlich angestrebten fünf auf neun Prozent - die Standardfehler der Zufallseffekte werden demzufolge etwas kleiner geschätzt, als sie realiter sind. Die Parameter der fixen Effekte sowie deren Standardfehler werden hingegen von der niedrigen Gruppenanzahl nicht beeinflusst (Hox, 2010). In dieselbe Richtung weist auch eine jüngst veröffentlichte Arbeit von Daniel Stegmüller (2013), wobei dessen Ergebnisse noch deutlich stärker zur Vorsicht mahnen als die bisherigen Daumenregeln. Er zeigt mittels Monte-Carlo-Simulationen, dass insbesondere wenn komplexere ME-Designs getestet

werden (z.B. mit Inklusion von Level-2-Variablen und Cross-Level-Interaktionen), eine vergleichsweise geringe Gruppenanzahl (unter 20) zu verzerrten Schätzern und vor allem zu falsch ausgewiesenen Konfidenzintervallen führt. Für den Test einer Level-2-Variable fasst er zusammen:

“Estimated with 15 or 20 available countries, ML confidence intervals are almost 5% too short – in other words, researchers are more likely to obtain 90% confidence intervals rather than the 95% intervals announced by their software package.” (Stegmueller, 2013, S. 758)⁵

Was bedeuten diese Voraussetzungen nun für die Anwendung von ME-Modellen in der Politikwissenschaft? In erster Linie, dass insbesondere die Anzahl der Level-2-Einheiten als limitierender Faktor nicht übersehen werden darf. Sollten in den Daten etwa nur 20 Level-2-Einheiten vorhanden sein – wie dies in der politikwissenschaftlichen Forschung beispielsweise bei Analysen der OECD-Welt oftmals der Fall ist – sind eben nur solche ME-Modelle sinnvoll schätzbar, die nur wenige Variablen auf Aggregatebene aufweisen und die v.a. nur wenige oder gar keine Cross-Level-Interaktionen testen. Kurz gesagt: je schlanker das Modell, desto geringer sind die Probleme durch geringe Fallzahlen. Zudem kann als pragmatische Antwort auf die Analyse von Stegmueller gesagt werden, dass man bei eher geringen Fallzahlen auf Level-2 auf die angegebenen Konfidenzintervalle immer noch einen gewissen Aufschlag dazurechnen sollte, bzw. wenn man eigentlich ein Signifikanzniveau von 95% anstrebt, die Schätzung mit einem Signifikanzniveau von 99% durchführt. Gleichzeitig dürften Analysen mit weniger als zehn Level-2-Einheiten wohl in den seltensten Fällen noch sinnvolle Ergebnisse liefern. Hier sind die berichteten MLE-Konfidenzintervalle laut Stegmueller oftmals um bis zu 15 Prozentpunkte zu schmal (2013, S.753), was jegliche ernsthafte Interpretation unterminiert. Damit sich der Leser ein Bild davon machen kann, wie schwerwiegend die Fallzahlproblematik ist, sollte es selbstverständlich sein, die Fallzahlen auf jeder einzelnen Ebene und im besten Fall auch die durchschnittliche Gruppengröße (und deren Spannweite) in einer ME-publikation anzugeben.

⁵ Laut Stegmueller stellt auch die REML Methode keine wirkliche Alternative dar. Sie generiert zwar im Fall von einfachen ME-Designs geringfügig bessere Konfidenzintervalle als die klassische, von ihm verwendete MLE. Bei komplexeren Designs hingegen ergäben sich „quite drastic noncoverage problems“ (Stegmueller, 2013, S. 751). Er empfiehlt vielmehr auf eine bayesianische Schätzung auszuweichen. Diese würde einerseits etwas weniger bias bei den Punktschätzern generieren und v.a. deutlich passendere Konfidenzintervalle erzeugen. Diese wären zudem im Gegensatz zu den von der MLE erzeugten Konfidenzintervallen eher zu lang als zu kurz, weshalb man sagen kann dass „researchers using Bayesian multilevel models put their hypotheses to more rigid tests than their colleagues relying on ML estimates!“ (Stegmueller, 2013, S.759). Da bayesianische ME-Modelle aber noch die absolute Ausnahme in der Politikwissenschaft darstellen (in unserem Datensatz finden sich gerade einmal sieben) und sich zudem ihre Schätzung deutlich von klassischen Regressionsverfahren unterscheiden, sollen diese hier nicht weiter verfolgt werden. Einen Einstieg in die Welt der Bayesianischen MEA bieten Hamaker und Klugkist (2011).

2.4. Modellierungsstrategien in Mehrebenen-Situationen

Wie die Abbildungen 1a-f gezeigt haben, können Varianzen zwischen wie innerhalb von Gruppen auf unterschiedliche Effekte zurückgehen. Die MEA bietet die Möglichkeit Level-1-, Level-2- sowie Cross-Level-Interaktionseffekte zu modellieren.

In Bezug auf Level-1-Effekte besteht der Vorteil der MEA gegenüber einer normalen OLS darin, dass je nach Gruppe ein unterschiedlicher Effekt angenommen werden kann und man entsprechend nicht davon ausgehen muss, dass eine Variable über alle Gruppen hinweg denselben Effekt auf die in der Gruppe enthaltenen Individuen ausübt. Dabei können zwei unterschiedlich flexible Modelle unterschieden werden: das Random-Intercept und das Random-Slope-Modell. Bei ersterem (formal: $y_{ij} = b_0 + b_1 x_{ij} + u_j + e_{ij}$)⁶ wird ausschließlich der Achsenabschnitt (Intercept) als variabel angenommen. Dieser variiert für jedes Land um u_j um den mittleren Intercept aller Gruppen b_0 herum (vgl. Abbildung 3a). Die Steigung b_1 hingegen wird für alle Länder als gleich angenommen. Insofern unterscheiden sich im Random-Intercept-Modell die Effekte der Level-1-Variablen ausschließlich entsprechend ihres Niveaus, nicht jedoch in ihrer Steigung.

Abbildung 3a Random-Intercept-Modell

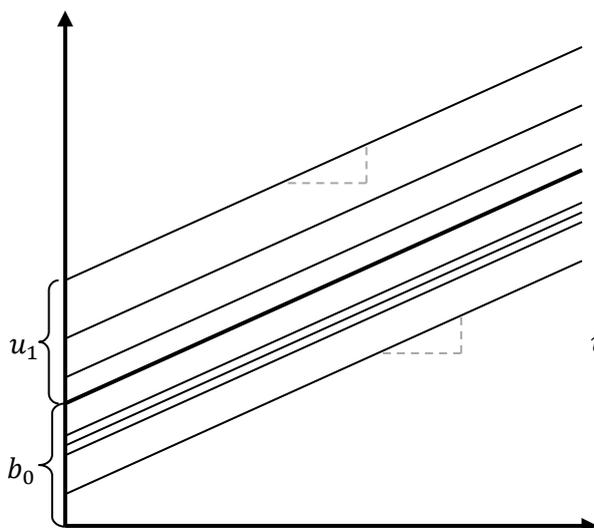
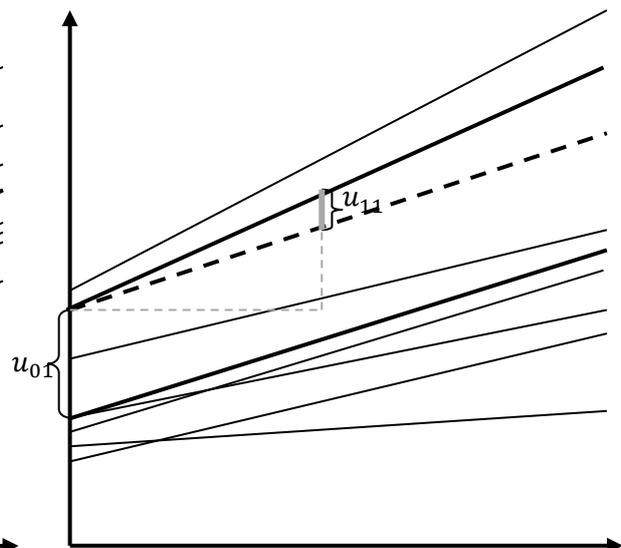


Abbildung 3b Random-Slope-Modell



Aus Gründen der besseren Interpretierbarkeit empfiehlt es sich oftmals Variablen am Gesamtmittelwert zu zentrieren (*grand mean centering*). Von einer Zentrierung am Gruppenmittelwert (*group mean centering*), wie sie Raudenbush (1989) noch generell empfohlen hat, raten heute hingegen die meisten Experten ab, da diese Form der Zentrierung das Basis-Modell so verändert, dass es nicht mehr äquivalent zum nicht-zentrierten Modell ist (Kreft et al., 1995). Hierdurch kann es vorkommen, dass man allein durch das Zentrieren am Gruppenmittelwert ein anderes Modell schätzt,

⁶ Der Index i steht für ein Individuum in der Gruppe j .

als man eigentlich im Auge hatte (Paccagnella, 2006).⁷ Generell sollten nur Variablen zentriert werden, bei denen der Mittelwert eine sinnvolle Aussage erlaubt. Anders gesagt, sollten primär metrische Variablen zentriert werden. Ordinale können am Median, Dummies, wie das Geschlecht, sollten hingegen nicht zentriert werden.

Nimmt man an, dass sich die Level-1-Effekte nicht nur im Niveau zwischen den Gruppen unterscheiden, sondern, dass eine Level-1-Variable wirklich unterschiedliche Effekte, d.h. unterschiedliche Steigungen in unterschiedlichen Gruppen aufweist, lässt sich dies über ein Random-Slope-Modell schätzen (formal: $y_{ij} = b_0 + b_1x_{ij} + u_{0j} + u_{1j}x_{ij} + e_{ij}$). Die neu hinzugekommenen Elemente der Gleichung lassen sich anschaulich in Abbildung 3b betrachten. Der Effekt, den eine Level-1-Variable hat, wird relativ zum mittleren Effekt über alle Gruppen betrachtet. Der Unterschied zum vorherigen Random-Intercept-Modell ist die Annahme, dass sich die Gruppen in Bezug auf einen Level-1-Faktor nicht nur in ihren Achsenabschnitten u_{0j} , sondern auch in ihren Steigungen u_{1j} unterscheiden. Anschaulich sieht das folgendermaßen aus: Vergleicht man die Gerade für Gruppe 1 mit der mittleren Gerade über alle Individuen (die beiden dicken Linien), zeigt sich für dieses Land eine größere Steigung als im Mittel (u_{1j}). Dies ist auch der Term, der zusätzlich zum Random-Intercept-Modell in die Formel hinzukommt.

Neben den Level-1-Effekten, die sich wie gesehen auch von Level-2-Einheit zu Level-2-Einheit unterscheiden können, kann es auch noch Effekte geben, die direkt von Level-2 aus auf die abhängige Variable wirken. So könnte sich beispielsweise die Arbeitslosenquote in einem Land auf die Einstellung der Bevölkerung zum Thema Arbeitslosengelderhöhung auswirken und dies unabhängig davon, ob die betreffende Person selbst arbeitslos ist oder nicht. Um solche Sozialisations- oder Kontexteffekte adäquat zu modellieren ist ein Rückgriff auf ME-Modelle notwendig, da die einzige Alternative – eine Inklusion von Level-2-Variablen in ein Level-1-Modell (= künstliche Disaggregation der Level-2-Einheiten) – wie beschrieben die Standardfehler dieser Variablen systematisch unterschätzen würde.

Neben den Haupteffekten können Level-1- und Level-2-Variablen auch in Form von Interaktionseffekten auf die abhängige Variable einwirken. Besonders interessant in dieser Hinsicht ist der Fall von Cross-Level-Interaktionen zwischen einer Level-1- und einer Level-2-Variable (vgl. Abbildung 1f). Die um den Interaktionseffekt erweiterte Modellgleichung schreibt sich folgendermaßen: $y_{ij} = b_0 + b_1x_{1ij} + b_2X_{2j} + b_3x_{1ij}*X_{2j} + u_{0j} + u_{1j}x_{1ij} + e_{ij}$. Die Darstellung der Cross-Level-Interaktionseffekte sollte sinnvollerweise – wie auch bei anderen Interaktionen üblich – entweder über vorhergesagte Werte (dabei auf einen sinnvoll zu interpretierenden Wertebereich achten) oder über marginale Effekte erfolgen (Preacher et al., 2006).

⁷ Nur bei spezifischen Fragestellungen böte sich eine Zentrierung am Gruppenmittel an: Beispielsweise um den sogenannten „frog-pond“-Effekt nachzuweisen, bei dem die relative Positionierung eines Falles im Vergleich zu den anderen Fällen seiner Gruppe die Ausprägung der abhängigen Variable mit determiniert (Hox, 2010, S. 69).

Auch wenn die Flexibilität, welche Random-Slope-Modelle und Cross-Level-Interaktionen bieten, verführerisch sein kann, sollten Forscher stets bedenken, dass jede Erweiterung des Modells durch zusätzliche Parameter immer auch auf Kosten der Modellschätzung geht. Insbesondere bei MEA gilt deshalb, dass die Modelle zwar so genau wie nötig aber eben auch so sparsam wie möglich gehalten sein sollten. Deshalb sollten Forscher zunächst stets testen, ob die Freisetzung einer Level-1-Varianz oder die Inklusion einer Cross-Level-Interaktion die Modellschätzung signifikant verbessern und damit wirklich sinnvoll sind. Hierzu wird zumeist auf Devianz-Tests (Likelihood-Ratio-Chi-Quadrat-Tests) zurückgegriffen. Alternativ können, um die Güte von zwei Modellen zu vergleichen, auch das Akaike Information Criterion (AIC) oder das Bayesian Information Criterion (BIC) herangezogen werden. Im Gegensatz zur Devianz bestrafen diese beiden eher die Aufnahme weiterer Variablen, weshalb sie im Sinne einer möglichst sparsamen Modellbildung einem Devianztest vorzuziehen sind. Das BIC bestraft zusätzlich auch noch eine hohe Fallzahl und stellt damit in der Regel einen noch härteren Test als das AIC dar (Hox, 2010, S.50). Ein Vorteil dieser Testmaße ist, dass sie nicht nur auf ineinander geschachtelte Modelle anwendbar sind – wie beim Devianztest der Fall – sondern mit ihrer Hilfe beispielsweise auch Modelle mit komplett unterschiedlichen erklärenden Variablen miteinander verglichen werden können. Einzig die Fallzahl der beiden Modelle muss identisch sein.

2.5. Gütemaße

Als erstes Gütemaß für MEA kann die bereits angesprochene Devianz dienen, welche sich als $2 \cdot$ der Log-Likelihood-Wert berechnet. Je näher dieser Wert an Null ist, desto besser passt das Modell auf die Daten. Allerdings ist der Absolutwert der Devianz abhängig von der Fallzahl und lässt sich dadurch nicht direkt interpretieren. Erst im direkten Vergleich zweier ineinander geschachtelter Modelle sind die Devianzwerte aussagekräftig: Ist ihre Differenz signifikant (Test mittels Chi-Quadrat-Verteilung), so kann dasjenige Modell mit dem absolut betrachtet niedrigeren Devianzwert als signifikant besser angepasst gelten als das andere (T. Snijders & Bosker, 2012). Die Devianz lässt damit nur Aussagen über die relative Güte eines ME-Modells im Vergleich zu einem, i.d.R. sparsamer ausgestalteten Modell zu. Bei AIC und BIC, die auch als Gütemaße verwendet werden können, weisen niedrige Werte auf einen besseren Fit des Modells hin (Raftery, 1995: 134). Auch hier lässt sich allerdings der Absolutwert nicht sinnvoll interpretieren, sondern ausschließlich die Differenz der AIC- bzw. BIC-Werte zweier Modelle M_0 und M_1 . Differenzwerte größer zehn weisen dabei sehr stark darauf hin, dass M_1 einen besseren Fit aufweist als M_0 (Raftery, 1995: 140).

Devianz-, AIC und BIC-Tests eignen sich um im direkten Vergleich mit anderen Modellen die am besten passende Modellspezifikation zu ermitteln, ein absolutes Maß für die Erklärungskraft eines ME-Modells stellen sie jedoch nicht dar. Hierfür schlagen Snijders und Bosker für ein Zwei-Ebenen-Modell zwei getrennt zu berechnende R^2 vor, ein R^2_{mikro} und ein R^2_{makro} . Insbesondere das erste, welches die „proportionale Fehlerreduktion für ein individuelles Outcome“ (T. Snijders & Bosker, 2012: 111 [eigene Übersetzung]) angibt, ist wichtig und dürfte für die meisten

politikwissenschaftlichen Analysen sinnvoll zu berichten sein. Das R^2 auf der Makroebene gibt hingegen an, um wie viel Prozent sich die Varianz zwischen den Level-2-Gruppen, d.h. zwischen den Gruppenmittelwerten, reduzieren lässt. Dies ist in den meisten Fällen von geringerem Interesse als das R^2_{mikro} (T. Snijders & Bosker, 2012: 111). Für die Berechnung greift man auf die Varianzen der Residuen zurück, die zwischen (τ) sowie innerhalb der Level-2-Gruppen (σ) vorliegen und vergleicht dabei das interessierende Modell (M_1) stets mit dem ME-Nullmodell (M_0). Formal sieht dies folgendermaßen aus (T. Snijders & Bosker, 1999: 102-103):

$$R^2_{mikro} = 1 - \frac{\sigma_{M_1} + \tau_{M_1}}{\sigma_{M_0} + \tau_{M_0}}; \quad R^2_{makro} = 1 - \frac{\tau_{M_1}}{\tau_{M_0}};$$

Diese Formeln sind für Random-Intercept-Modelle geeignet. Bei Random-Slope-Modellen ist die Berechnung komplizierter (T. Snijders & Bosker, 1994). Da sich die R^2 -Werte aber meist nur marginal ändern, sobald Zufallsparameter in das Modell aufgenommen werden, plädieren Snijders und Bosker dafür ihr R^2 stets auf Basis des entsprechenden Random-Intercept-Modells zu berechnen.⁸

Es gibt also durchaus Möglichkeiten die Güte eines ME-Modells entweder für sich genommen oder im Vergleich zu anderen Modellen in den Blick zu nehmen. Entsprechende Angaben zur Modellgüte sollten deshalb Standard in allen Veröffentlichungen sein.

3. Meta-Analyse politikwissenschaftlicher Mehrebenenartikel

3.1. Die Bedeutung der MEA in der politikwissenschaftlichen Forschungspraxis

Um einen möglichst repräsentativen Überblick der Anwendung der MEA geben zu können, greifen wir auf den Social Science Citation Index (SSCI) zurück. Dort werden die Bereiche *Political Analysis*, *Public Administration* und *International Relations* auf die Stichworte *multilevel model*, *multilevel analysis*, *Mehrebenenanalyse* sowie *hierarchical linear model* hin durchsucht.⁹ Um einen Vergleichsmaßstab zu erhalten führen wir dieselbe Abfrage auch im Bereich *Sociology* durch. Abbildung 4 zeigt die Entwicklung der Publikationszahlen pro Jahr seit Anfang der 1990er Jahre als die ersten MEA veröffentlicht wurden.¹⁰ Bei den auf diese Weise identifizierten Artikeln dürfte es sich zum überwiegenden Teil wirklich um ME-Artikel im von uns intendierten statistischen Sinn handeln. Für die Artikel aus dem Bereich Politikwissenschaft fanden wir im Rahmen der weiter unten noch

⁸ Sofern Kovarianzen zwischen den unabhängigen Variablen in dem Modell angenommen werden, lassen sich die R^2 -Werte nach Snijders und Bosker nicht mehr sinnvoll interpretieren. Pötschke (2006: 174) schlägt in diesen Fällen die Verwendung des Pseudo- R^2 nach Maddala vor, welches grundsätzlich bei allen nach dem Maximum Likelihood Verfahren geschätzten Modellen berechnet werden kann:

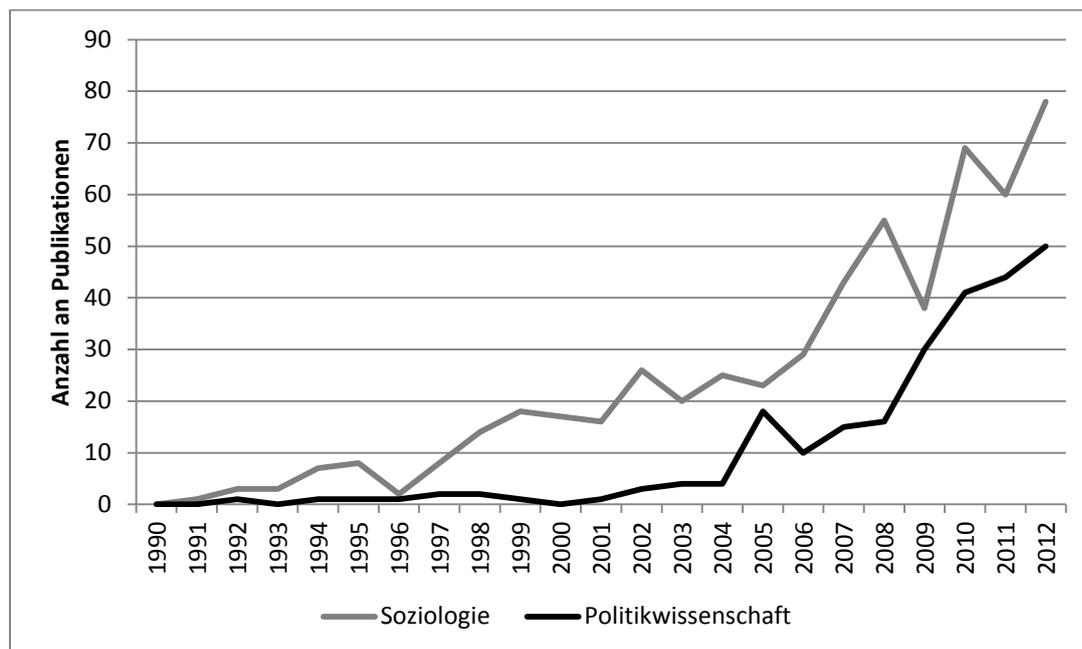
$Maddala R^2 = 1 - \exp\left(\frac{-2 \log Likelihood_{M_1} - (-2 \log Likelihood_{M_0})}{n}\right)$. Der sich ergebende Wert gibt an wie viel Prozent der Varianz die im Vergleich zum Nullmodell hinzugefügten Variablen erklären können, wobei diese Erklärungskraft sich sowohl auf „Varianzen der unabhängigen Variablen als auch [...] Kovarianzen zwischen unabhängigen Variablen“ (Pötschke, 2006: 174) beziehen kann.

⁹ Nach alternativen Schreibweisen wie *multi-level analysis* oder *multi-level model* wurde ebenfalls gesucht.

¹⁰ Auch wenn unser Datensatz alle Artikel enthält, die bis zum 03.07.2013 veröffentlichten wurden, lassen wir für die quantitativen Analysen das unvollständige Jahr 2013 außen vor um eine vergleichbare Datenbasis zu erhalten. In den qualitativen Analyseschritten werden die 2013 veröffentlichten Artikel dann berücksichtigt.

genauer beschriebenen qualitativen Analyse nur in elf der über 240 Fälle Artikel, die nicht in unser Untersuchungsraaster passen (z.B. rein qualitative Analysen unterschiedlicher Ebenen). Da anzunehmen ist, dass wir es in der Soziologie mit einer ähnlich niedrigen Fehlerquote zu tun haben, können wir davon ausgehen, dass die beiden folgenden Kurven das Vorkommen von MEA tendenziell nur sehr wenig überschätzen. Zudem ist natürlich auch denkbar, dass wir einige MEA über die Stichwortsuche nicht identifizieren. Insgesamt dürften die Ergebnisse aber eine ausreichend reliable und valide Basis für Aussagen zur Ausbreitung dieser statistischen Methode bieten.

Abbildung 4: Mehrebenenpublikationen in der Politikwissenschaft und Soziologie



Offensichtlich ist zunächst ein Anstieg an ME-Artikeln innerhalb des von uns betrachteten Zeitraums zu verzeichnen. Gleichzeitig sieht man aber auch, dass in der Soziologie diese Entwicklung bereits früher als in der Politikwissenschaft eingesetzt hat. Bereits Mitte der 1990er Jahre finden sich in den betrachteten soziologischen Zeitschriften jährlich stets eine gute Handvoll von ME-Artikeln und seit Ende dieses Jahrzehnts stieg deren Zahl dann relativ kontinuierlich bis auf aktuell fast 80 pro Jahr. In der Politikwissenschaft gab es hingegen bis Mitte der 2000er Jahre nur sehr vereinzelt MEA. Dann jedoch steigt die Kurve nahezu parallel zu der in der Soziologie – insbesondere die letzten fünf Jahre zeigen dabei ein sehr starkes Wachstum. Letztlich kann somit konstatiert werden, dass die MEA als statistisches Verfahren in der Breite der politikwissenschaftlichen Forschung angekommen zu sein scheint, auch wenn wir im Vergleich zur Soziologie etwa fünf bis zehn Jahre aufzuholen haben.

Neben den Publikationen kann auch die Anzahl an Zitationen als Maß für die Bedeutsamkeit eines bestimmten Forschungsbereichs oder in unserem Fall für die Relevanz einer bestimmten Methode innerhalb einer Wissenschaftsdisziplin herangezogen werden. Denn nur wenn die veröffentlichten MEA von der politikwissenschaftlichen Community auch wahrgenommen werden, können sie einerseits dazu beitragen unsere Disziplin inhaltlich voran zu bringen (ansonsten wären Sie

methodische Spielereien ohne wirklichen Mehrwert) und andererseits sozusagen als Multiplikatoren den Lesern die Methode der MEA sowie die Möglichkeiten, die diese bietet, vorstellen und damit helfen dieses Verfahren in der Politikwissenschaft zu verbreiten.¹¹

Abbildung 5: Zitationen der Mehrebenenpublikationen in Politikwissenschaft und Soziologie

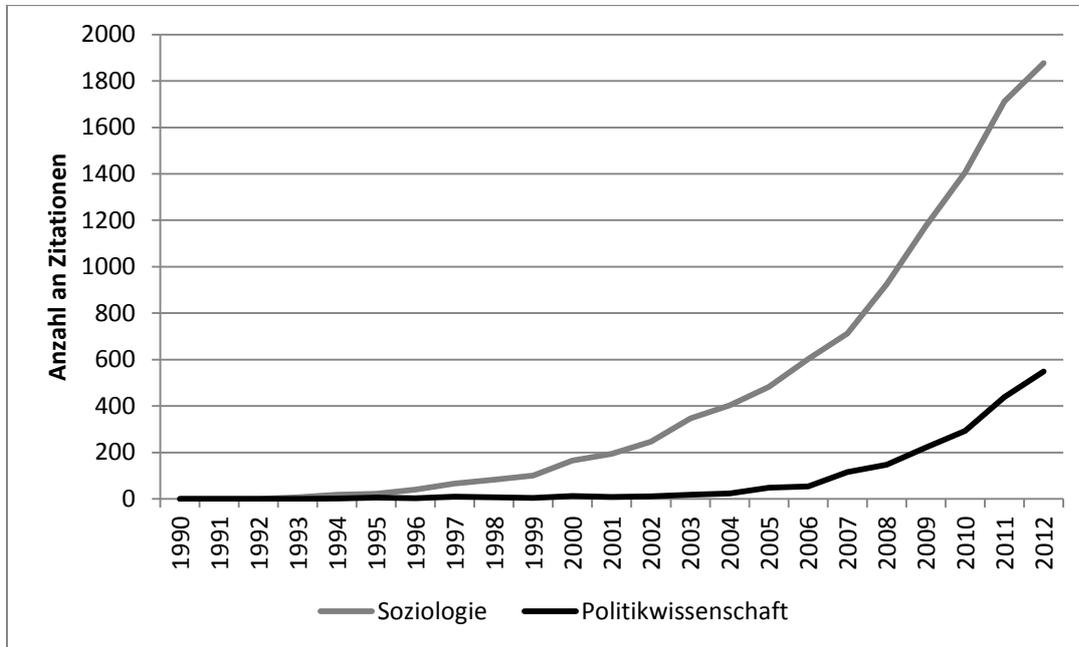


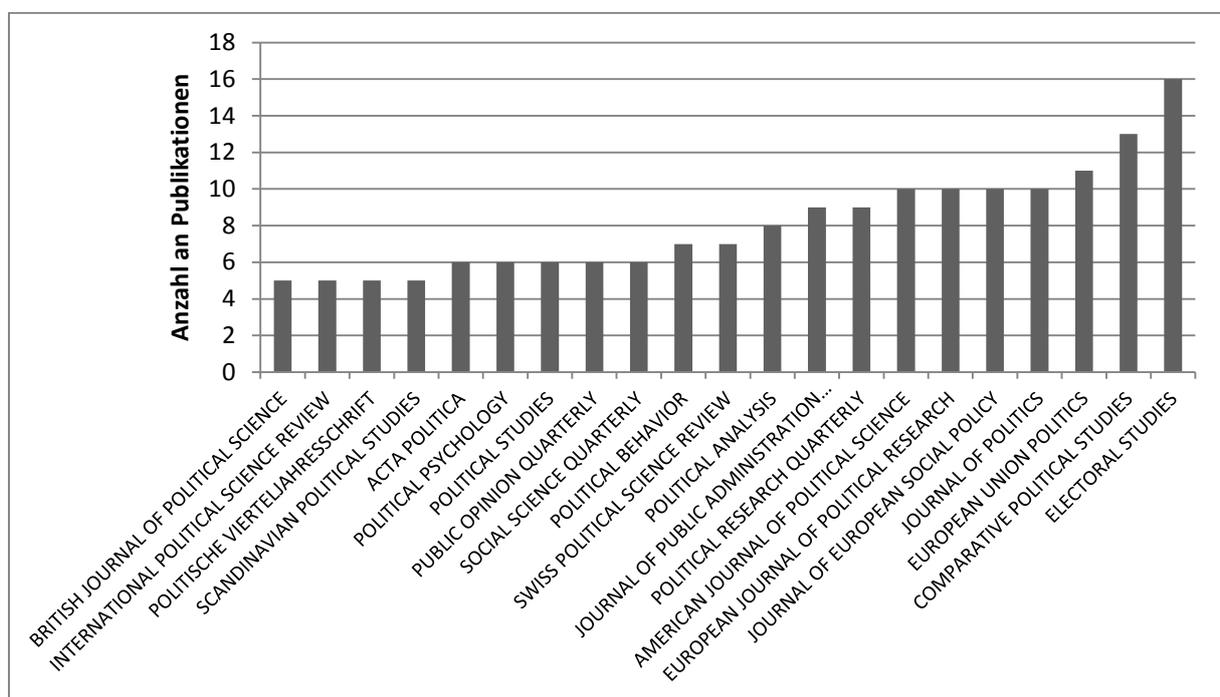
Abbildung 5 trägt die Anzahl an Zitationen der zuvor betrachteten ME-Publikationen über die Zeit ab. Wie nicht anders zu erwarten zeigt sich auch hier ein deutlicher Anstieg, der in der Soziologie bereits Mitte der 1990er Jahre beginnt, in der Politikwissenschaft in etwa 10 Jahre versetzt. Das Niveau, welches die Soziologie hier erreicht ist durchaus beeindruckend: Die insgesamt 563 ME-Artikel, die dort bis 2012 inklusive erschienen, wurden zuletzt insgesamt 1877 mal in anderen SSCI-gerankten Zeitschriften zitiert. Entsprechend werden im Durchschnitt die bislang erschienenen soziologischen ME-Arbeiten 3,3 mal zitiert. In der Politikwissenschaft liegt dieser Wert noch bei 2,2. Betrachtet man allerdings die Entwicklung der Publikationen, bei denen die Politikwissenschaft ähnliche Wachstumsraten verzeichnet wie die Soziologie, ist davon auszugehen, dass auch die wissenschaftliche Perzeption dieser Arbeiten und damit die durchschnittliche Anzahl an Zitationen sich auf ähnlich hohe Werte wie in der Soziologie zubewegen wird.

Eine Aufgliederung der ME-Artikel nach Zeitschrift ist ebenfalls aufschlussreich. In Abbildung 6 sind alle Zeitschriften enthalten, in denen zwischen 1990 und 2013 insgesamt mindestens fünf ME-Artikel veröffentlicht wurden. Spitzenreiter ist *Electoral Studies*, gefolgt von einer Reihe an Journals, die entweder dem Teilbereich der vergleichende Politikwissenschaft (*Comparative Political Studies*, *European Union Politics*) zuzuordnen sind, oder als *general purpose* Journals eine breite Palette an

¹¹ Einen ähnlichen, wahrscheinlich sogar noch stärkeren Effekt dürften die Veröffentlichungen der ersten anwendungsorientiert geschriebenen Methodenlehrbücher zum Thema MEA gehabt haben. Dies waren die auch heute noch in fast jeder ME-Publikation zitierten Werke von Snijders und Boskers (1999), Hox (1995, 2002) sowie Raudenbush und Bryk (2002).

Themen beinhalten (z.B. Journal of Politics, American Journal of Political Science). Daneben finden sich auch acht Artikel in Political Analysis, einer Zeitschrift, die sich insbesondere der Weiterentwicklung quantitativer Methoden verschrieben hat. Während mit Political Behavior, Public Opinion Quarterly und Political Psychology drei Zeitschriften mit einem Profil im Bereich der politischen Einstellungsforschung vertreten sind, ist nur eine im Bereich Policy Analysis (Journal of European Social Policy) und keine einzige aus dem Bereich der internationalen Beziehungen unter den 20 Zeitschriften, die die meisten ME-Artikel publizieren. Somit zeigt sich bereits auf der Ebene der Zeitschriften eine leichte Schiefelage zugunsten bestimmter Themen, wenngleich sich auf diese Weise die vielen in *general purpose* Zeitschriften erschienenen Artikel noch nicht verorten lassen.

Abbildung 6: Mehrebenenpublikationen in der Politikwissenschaft nach Zeitschrift

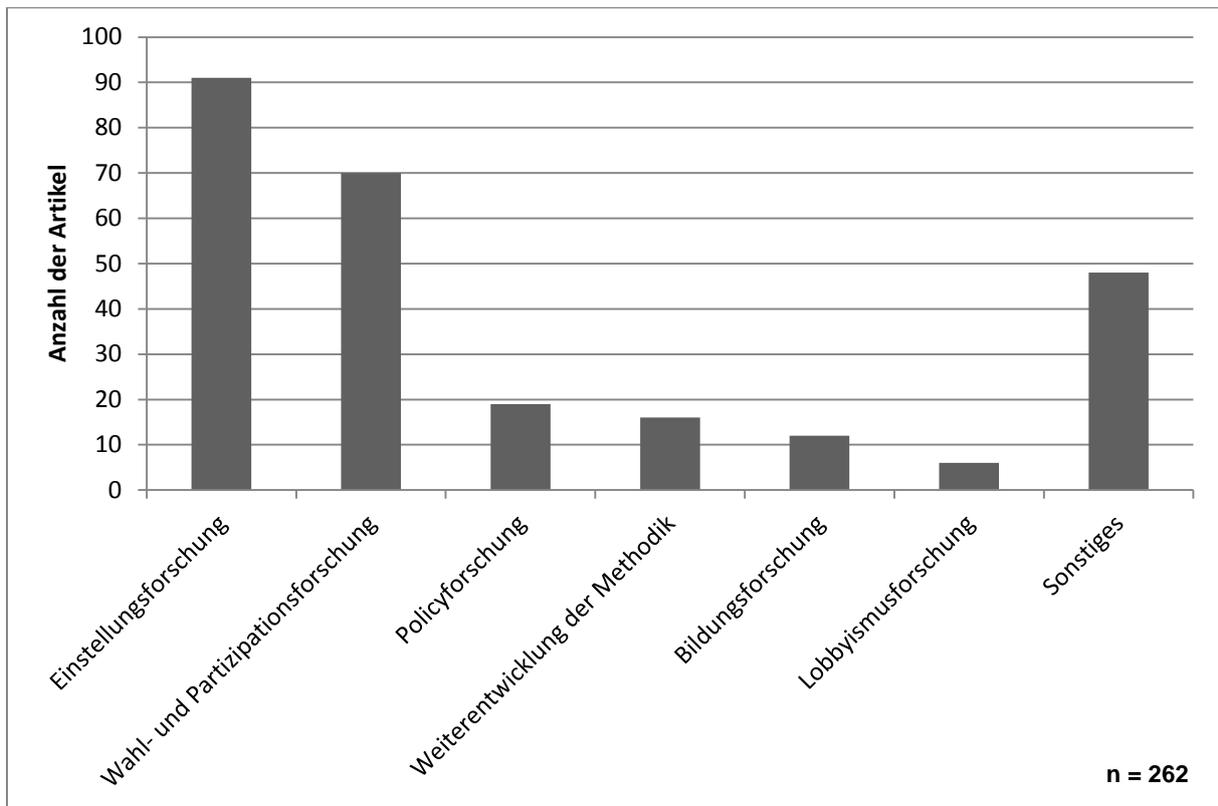


Anwendungsgebiete mehrebenenanalytischer Untersuchungen

Um ein genaueres Bild geben zu können, für welche politikwissenschaftlichen Fragestellungen die MEA eingesetzt wird, wurden die Artikel von uns in einem zweiten Schritt qualitativ ausgewertet. Hierfür wurden alle insgesamt 261 politikwissenschaftlichen Artikel, die wir unter den oben genannten Stichworten für den Zeitraum von 1990 bis Mai 2013 über den SSCI ausfindig machen konnten in Themenkategorien eingeteilt.

Der Themenkatalog wurde im Sinne der *grounded theory* induktiv erstellt und bei der genaueren Untersuchung der Artikel weiter modifiziert bzw. durch neue Kategorien ergänzt (Mey & Mruck 2011: 23f). Jeder Artikel wurde mindestens einem Themenbereich zugeordnet, wobei auch die Möglichkeit bestand, einem Artikel im Zweifelsfall mehrere Themengebiete zuzuschreiben. Abbildung 7 zeigt die sechs häufigsten Kategorien. Alle Themenbereiche, die nach Abschluss der Analyse weniger als fünf Artikel umfassten wurden unter der Kategorie *Sonstiges* zusammengefasst.

Abbildung 7: Mehrebenenpublikationen in der Politikwissenschaft nach Themen¹²



Die Analyse zeigt, dass der überwiegende Teil der Artikel im Bereich der Einstellungsforschung zu verorten ist. Zwar kommen die ersten politikwissenschaftlichen Artikel aus den neunziger Jahren aus der Wahl- und Methodenforschung, seit Mitte des letzten Jahrzehnts gewinnt die MEA in der Einstellungsforschung jedoch kontinuierlich an Bedeutung, sodass allein für das Jahr 2012 bereits 23 Artikel in diesen Themenbereich fallen. Mit nun insgesamt 91 Artikeln macht die Einstellungsforschung damit mehr als ein Drittel aller politikwissenschaftlichen Publikationen aus. Aufgrund der großen Fallzahl und der Breite des Forschungsgebietes wird diese Themenkategorie in Abbildung 8 nochmals aufgegliedert. Hier bietet sich ein relativ ausgewogenes Bild, dominiert von drei größeren Themenblöcken: Eine erste Gruppe an Arbeiten untersucht die Einstellung der Bürger zu bestimmten Policies, wobei hier allein im letzten Jahr acht Publikationen erschienen sind. Der Fokus liegt dabei auf der Sozialpolitik, es finden sich aber auch Arbeiten zur Unterstützung der Bevölkerung von Maßnahmen in den Bereichen der Wirtschafts-, Umwelt- oder Entwicklungshilfepolitik. Der zweite größere Block besteht aus Publikationen, die sich mit der Einstellung der Bürger zur Europäischen Integration befassen. Dies ist insofern wenig überraschend, als dass die zunehmende Kompetenzübertragung auf supranationale Entscheidungsgremien deutlich zugenommen hat, was die Europäische Union als Gebilde mit mehreren Analyseebenen auf den ersten

¹² Bei der Analyse fanden sich insgesamt 16 Artikel, die keinen direkten Bezug zur statistischen Methode der MEA aufwiesen, sodass letztlich 245 Artikel in die Untersuchung aufgenommen wurden. 17 Artikeln wurden mehr als einem Themenbereich zugeordnet und finden sich entsprechend doppelt in der Grafik sodass die Einteilung auf 262 Observationen basiert.

Blick für die MEA geeignet erscheinen lässt. Der letzte der drei großen Blöcke innerhalb der Einstellungsforschung, besteht aus Arbeiten, die sich entweder mit dem Vertrauen in die politischen Institutionen und Entscheidungsträger oder dem soziale Vertrauen in die Mitmenschen beschäftigen. Daneben finden sich Artikel zu den Einstellungen der Bevölkerung gegenüber ethnischen Minderheiten und Immigranten oder solche, die nach Faktoren suchen, welche den Einfluss auf die Befürwortung allgemeiner demokratischer Werte determinieren.

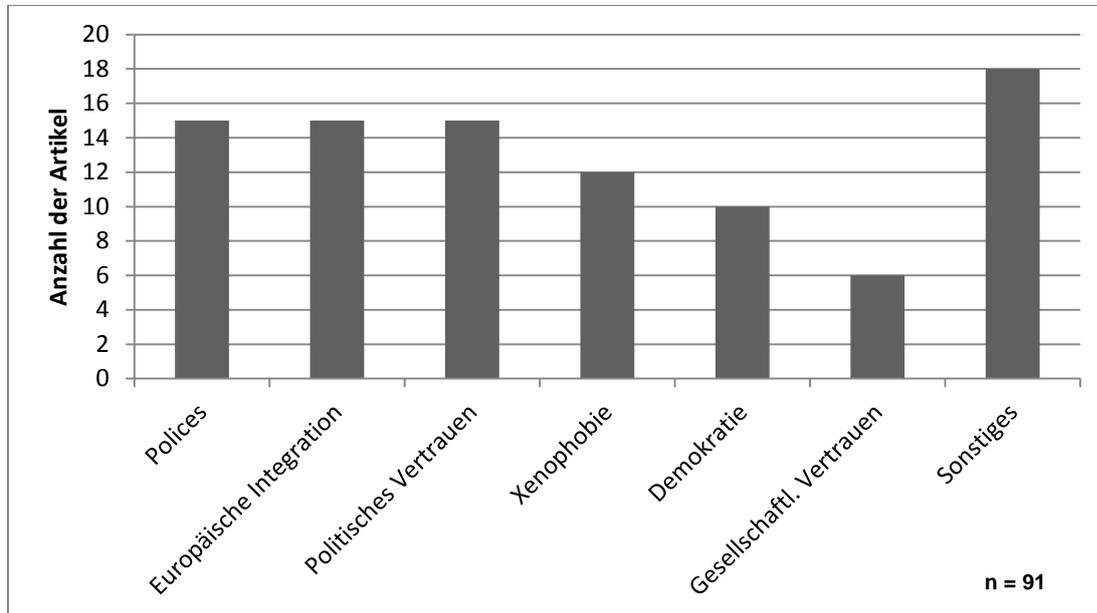
Ein anderer, mit der Einstellungsforschung eng verwandter Forschungsbereich, der mit 70 Artikeln rund ein Viertel aller Publikationen ausmacht, ist die Wahl- und Partizipationsforschung. Hierbei fokussiert eine kleine Gruppe an Artikeln auf die Motivationen und Voraussetzungen für die Teilnahme an alternativen demokratischen Partizipationsformen (Volksentscheide u.ä.), während sich ein Großteil der Arbeiten auf die Analyse von Wahlergebnissen konzentriert. Mit Daten aus verschiedenen Wahlumfragen wird zum einen untersucht, welchen Einfluss Mikro- und Makrofaktoren auf die Wahlentscheidung für eine bestimmte Partei aufweisen und zum anderen, welche Faktoren für eine hohe oder niedrige Wahlbeteiligung ursächlich sind. Interessant erscheinen in diesem Zusammenhang auch einige Artikel, die das ME-Modell als Alternative zur gepoolten Zeitreihenanalyse verwenden und so versuchen Veränderungen der Wahlbeteiligung über die Zeit zu erklären (s. Kap. 4).

Nicht nur die Einstellung der Bevölkerung gegenüber politischen Konzepten sind Gegenstand von ME-Untersuchungen, auch die Auswirkungen und Erfolge bestimmter Policies wurden mit dieser Methode bereits analysiert – auch wenn es sich hierbei noch um vergleichsweise wenige Artikel handelt. In den meisten Fällen wurden dazu Experteninterviews ausgewertet und mit Individualdaten und Makrofaktoren in Verbindung gesetzt. Die Politikfelder sind hier ähnlich wie im Bereich der Einstellungsforschung breit gestreut, wobei sich ein leichter Schwerpunkt im Bereich der Sozial- und Familienpolitik feststellen lässt. Wie bereits angemerkt, finden sich unter den Zeitschriften mit vergleichsweise vielen MEA auch methodenorientierte Journals wie *Political Analysis*. Allein die Hälfte aller Artikel, die von uns in die Kategorie *Weiterentwicklung der Methodik* eingeordnet wurden, sind in eben dieser Zeitschrift erschienen. Zwar geben auch einige der anderen Artikel kurze Einführungen in die statistischen Grundlagen der MEA, die Veröffentlichungen, die wir in dieser Kategorie zusammengefasst haben, beschäftigen sich jedoch ausschließlich mit der Methode und setzen sich intensiv mit ihren Vor- und Nachteilen auseinander. Da historisch betrachtet die Bildungsforschung einer der Vorreiter in der Anwendung von MEA war,¹³ verwundert es auch nicht, dass sich auch in den von uns untersuchten politikwissenschaftlichen Journals immerhin 11 Artikel aus dieser Fachrichtung finden lassen, obgleich die dort untersuchten Fragestellungen sicherlich nicht zum Kernbereich politikwissenschaftlicher Themen zu zählen sind. Zu guter Letzt enthält das Sample

¹³ Hier werden primär die Leistungen von Schülern in Abhängigkeit von deren individuellen Charakteristika sowie Merkmalen der Schulklasse (z.B. Qualität des Lehrers) oder der Schulform modelliert.

einige Artikel, die sich mit dem Einfluss von Interessengruppen auf politische Entscheidungen befassen. Da diese bis auf eine Ausnahme alle 2012 erschienen sind, bleibt jedoch noch abzuwarten, ob sich die MEA in diesem Bereich der Lobbyismusforschung tatsächlich als Analysetool etabliert.

Abbildung 8: Themengebiete von MEAs im Bereich der Einstellungsforschung



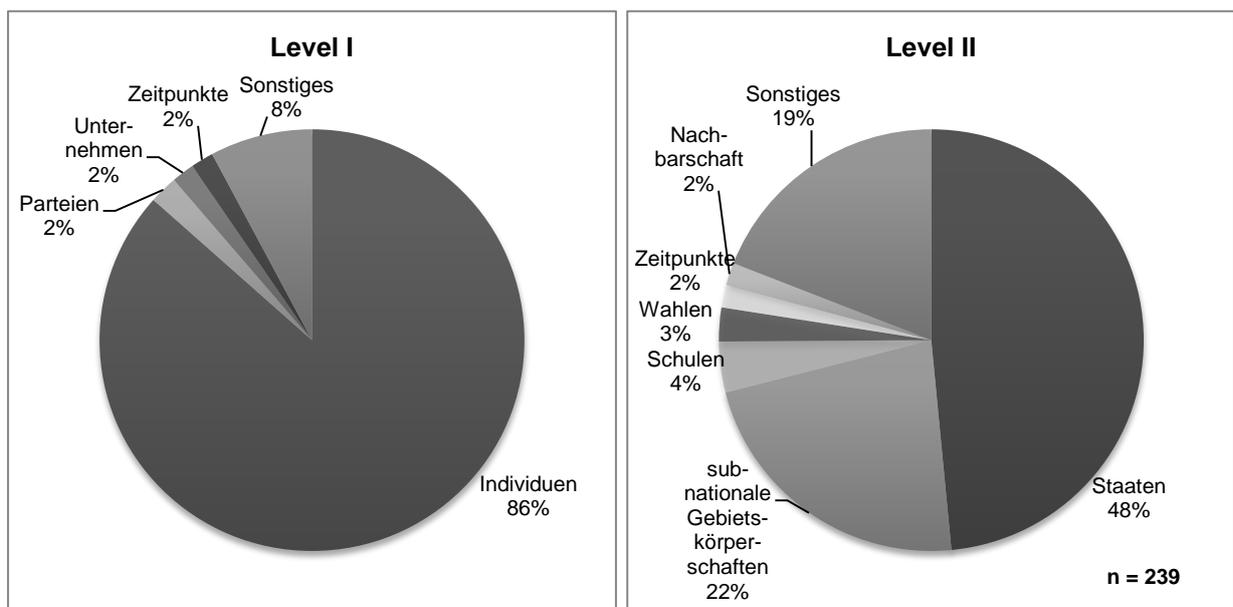
3.2 Analyseebenen und statistische Voraussetzungen

Neben der thematischen Einordnung interessiert uns zudem, auf welchen Ebenen sich die Forschungsarbeiten bewegen, und inwiefern sie die oben beschriebenen statistischen Voraussetzungen erfüllen. Wir haben hierfür aus den Artikeln – sofern diese die entsprechenden Informationen enthielten – die Art der Untersuchungseinheiten, die unterschiedlichen Ebenen, die Fallzahlen sowie die angegebenen IKK-Werte entnommen.

Auch wenn die MEA generell die Einbindung mehrerer Analyseebenen gestattet, beschränken sich die von uns erhobenen Artikel zu einem großen Teil auf die Arbeit mit nur einem übergeordneten Level. Nur 27 Artikel arbeiten zusätzlich mit einer dritten Ebene, eine einzige Arbeit zieht sogar eine vierte Ebene heran. So beschränken auch wir uns bei der Auswertung der Analyseeinheiten auf die ersten zwei Ebenen. Abbildung 9 zeigt die prozentuale Verteilung der Analyseeinheiten auf den ersten beiden Ebenen. Betrachtet man die thematische Auswertung und den starken Fokus auf die Bereiche Einstellungs- und Wahlforschung überrascht es nicht, dass sich fast neunzig Prozent der Artikel bei der Analyse auf der ersten Ebene auf wirkliche Individual- d.h. Personendaten stützen. Die am häufigsten herangezogenen Datensätze sind mit 21 Fällen der European Social Survey und mit 19 Fällen der World Values Survey, gefolgt von Daten aus den Eurobarometer-Umfragen (13 Artikel) und dem International Social Survey Programme (9 Artikel). Die durchschnittliche Fallzahl auf Level-1 liegt bei den Publikationen, die mit Individualdaten arbeiten bei rund 35.000 (Min: 120 Max:

455.000). Wie erwartet lassen sich also hier, keine besonderen Probleme mit zu geringen Fallzahlen auf der Individualebene feststellen. Werden andere Analyseeinheiten als Individuen untersucht, sind die Fallzahlen allerdings mitunter deutlich geringer. Zwar finden sich auch hier im Fall von Paaren, Haushalten oder Unternehmen als Level-1-Einheiten N im fünfstelligen Bereich, manche Artikel arbeiten aber auch mit Fallzahlen um die 100 Level-1-Einheiten, was die Aussagekraft dieser Arbeiten stark einschränkt. Die Forschungsarbeiten, die bestimmte zeitabhängige Variablen auf der ersten Ebene ansiedeln, weisen in der Regel etwas höhere Fallzahlen auf, wobei dies stark davon abhängig ist, ob die zeitabhängigen Faktoren auf der Individualebene liegen (z.B. bei Untersuchungen von individuellem Wahlverhalten) oder bereits auf Makroebene (z.B. bei Untersuchungen zur Entwicklung der Wahlbeteiligung). An dieser Stelle muss kritisch angemerkt werden, dass über zehn Prozent der Autoren in den Publikationen für ihre Berechnungen keine genauen Fallzahlen ausweisen oder auch die einzelnen von ihnen verwendeten Ebenen für ihr ME-Design nicht benennen, auch wenn diese Angaben essentiell sind, um die Arbeiten adäquat bewerten zu können.

Abbildung 9: Einheiten auf den unterschiedlichen Analyseebenen



Auch auf der zweiten Ebene lässt sich ein deutlicher Fokus auf bestimmte Analyseeinheiten ausmachen, auch wenn dieser nicht ganz so deutlich ist, wie auf Level-1. Fasst man alle politisch-geographischen Einheiten zusammen, machen diese rund 75% der Level-2-Einheiten aus. Den größten Block stellen dabei Staaten dar, die in knapp der Hälfte aller politikwissenschaftlichen ME-Artikel die zweite Analyseebene bilden. In 22% der Fälle werden Kontextfaktoren auf der Ebene von subnationalen Gebietskörperschaften wie den amerikanischen Bundesstaaten oder den Schweizer Kantonen in die Untersuchung einbezogen - auch MEA mit Daten auf regionaler oder kommunaler Ebene finden sich unter den erhobenen Publikationen. Wie bereits vermutet stellen die Fallzahlen auf Level-2 offensichtlich die größere statistische Herausforderung dar. Gerade in den Fällen in denen staatliche Kontextfaktoren wie z.B. das Bruttoinlandsprodukt, Wahl- und Parteiensysteme oder auch

auf Staatsebene aggregierte Individualdaten wie Arbeitslosenzahlen in die Analyse aufgenommen werden, sind die Fallzahlen in vielen der analysierten Artikel relativ klein. Da die nötigen Individualdaten oft nur für OECD-Staaten oder die Mitgliedstaaten der Europäischen Union vorliegen und von diesen Datenkorpussen zumeist auch noch weitere Fälle für die Analyse ausgeschlossen werden müssen, liegt der Durchschnitt an staatlichen Level-2-Einheiten nur bei rund 27. Auffällig ist, dass ungefähr die Hälfte der „Staaten-Artikel“ sogar mit weniger als 20 Fällen auf der Kontextebene arbeitet und somit nach den oben beschriebenen Ergebnissen Stegmüllers zu verzerrten Schätzern und zu falsch ausgewiesenen Konfidenzintervallen kommen dürften. Eine deutlich weitere Spannweite was die Fallzahlen betrifft, findet sich bei den Gebietskörperschaften, je nach Größe der Einheiten. Hier lässt sich festhalten, dass je kleiner das Gebiet, das als zweite Ebene herangezogen wird, desto größer im Schnitt die Fallzahlen. Erwartungsgemäß noch größere und somit statistisch unbedenkliche Fallzahlen liefern andere Level-2-Einheiten wie Bildungseinrichtungen oder die Nachbarschaft.

Neben den teilweise relativ kleinen Fallzahlen auf Level-2 zeigt sich bei der Analyse der politikwissenschaftlichen ME-Artikel ein weiteres methodisches Problem. Wie oben erläutert, stellt der IKK eine wichtige Größe dar, um noch vor der Durchführung die Sinnhaftigkeit des Einsatzes einer MEA zu bewerten. Fällt die durch die Kontextebene zu erklärende Varianz der abhängigen Variable sehr gering aus, so bringt auch der Einbezug der hierarchischen Struktur welche unterschiedliche Level-1-Effekte je nach Gruppe ermöglicht sowie die Hinzunahme von Kontextvariablen in das Modell nur einen geringen Mehrwert im Vergleich zu einer normalen Regression. Eine zufällige Stichprobe von einem Drittel der Artikel ergab allerdings, dass der IKK in rund der Hälfte der Forschungsarbeiten gar nicht berechnet – oder zumindest nicht im Artikel ausgewiesen wird. Dies stellt für die Bewertung der Ergebnisse ein nicht zu unterschätzendes Manko dar. In den Fällen, in denen der IKK angegeben wurde liegt dieser in einem Bereich zwischen 0,002 und 0,82 wobei sich die meisten Werte zwischen 0,05 und 0,3 und somit eher an der von Hox angegebenen Untergrenze für den sinnvollen Einsatz der MEA bewegen. Für die Politikwissenschaft besonders bemerkenswert ist, dass die Erklärungskraft von staatlichen Kontextfaktoren im Vergleich zu anderen Kontextvariablen relativ gering zu sein scheint. Die größten Werte für den IKK finden sich in den Forschungsarbeiten, die Kontextfaktoren auf Unternehmens- oder Nachbarschaftsebene verwenden. Dieses Ergebnis lässt sich durch zwei einfache Überlegungen erklären: Zum einen lässt sich in kleineren Gruppen auch eine größere Homogenität der Mitglieder, bzw. ein stärkerer sozialisierender Effekt auf die Gruppenmitglieder erwarten. Die Varianz zwischen den Individuen innerhalb einer Gruppe ist dementsprechend geringer, während die Varianz zwischen den Gruppen tendenziell an Bedeutung gewinnt. Zum anderen liegen, wie oben bereits erläutert, bei geographisch großen Kontexteinheiten oftmals auch relativ kleine Fallzahlen vor. Wenn es gleichzeitig eine große Zahl an Beobachtungen auf der Individualebene gibt, wird klar, dass die Varianz der abhängigen Variable schon allein aufgrund der Fallzahlen hauptsächlich durch die erste Ebene zu erklären ist.

4. Innovative Herangehensweisen und Perspektiven

In diesem letzten Kapitel sollen die Ergebnisse der Meta-Analyse durch die genauere Beschreibung einiger innovativer Beispiele aus unserem Datensatz ergänzt und so mögliche Perspektiven und Voraussetzungen für den gewinnbringenden Einsatz der MEA in der politikwissenschaftlichen Forschung aufgezeigt werden. Die Meta-Analyse zeigt, dass die ME-Methode in der Politikwissenschaft hauptsächlich dafür angewendet wird, den Einfluss von Makrofaktoren der nationalen oder subnationalen Ebene auf individuelle Einstellungen oder Verhaltensmuster messbar zu machen. Das erscheint insofern logisch und sinnvoll, als dass individuelles Denken und Verhalten nur selten losgelöst von kontextuellen Bedingungen erfolgt und entsprechend auch nur in Kombination mit Makrofaktoren adäquat analysiert werden kann. Gleichzeitig konnte unsere Meta-Analyse aber optimistische Einschätzungen wie von van Deth nicht bestätigen. Dieser hatte postuliert, dass die Bedeutung des Staats als Erklärungsfaktor für die Varianzaufklärung individueller Einstellungen in manchen Fällen sogar über die individueller Variablen hinaus (van Deth, 1995: 457). Daraus soll nicht der Schluss gezogen werden, dass der Einsatz von MEA für die Messung des Einflusses staatlicher Faktoren auf individuelle Einstellungen generell unangebracht wäre. Im Gegenteil finden sich in unserem Datensatz auch Forschungsarbeiten, die ca. 30 Prozent der Gesamtvarianz auf der Ebene des Nationalstaats verorten. In diesen Fällen ist es nicht nur sinnvoll, sondern sogar unbedingt notwendig mehrebenenanalytische vorzugehen, um den Einfluss nationaler Spezifika sinnvoll modellieren zu können. In den Fällen allerdings, in denen die Varianzaufteilung ergibt, dass weniger als 5% der Varianz durch die staatliche Aggregatebene zu erklären sind, sollte man sich fragen, ob diese die angemessen Untersuchungsebene darstellt, oder nicht vielleicht die Konzentration auf kleinere Einheiten fruchtbarere Ergebnisse produzieren könnte. So findet sich in unserem Datensatz beispielsweise ein Artikel aus der Zeitschrift *Electoral Studies* (Johnston et al., 2005), welcher mithilfe von Daten des British Household Panel Survey (BHPS) das Wahlverhalten britischer Wähler auf individueller und auf Haushaltsebene untersuchen. Ausgehend von der Annahme, dass das engste soziale Umfeld auch den stärksten Einfluss auf individuelles Verhalten ausübt, untersuchen die Autoren die Wahlentscheidung von knapp 15.000 Bürgern, gruppiert in rund 8.500 Haushalten. Dabei kommen Sie zu dem Ergebnis, dass der überwiegende Anteil der Varianz im Wahlverhalten durch Variablen auf der Haushaltsebene zu erklären ist (Johnston et al., 2005: 217). Der enorme Varianzanteil von 82%, den die Autoren auf der Kontextebene ausmachen, muss hier allerdings insofern relativiert werden, als dass auch Ein-Personen-Haushalte in die Analyse aufgenommen wurden und so die zwei Ebenen, die eigentlich getrennt betrachtet werden sollen – nämlich Individuum und Haushalt – letztlich doch wieder vermischt wurden. Auch wenn in diesem Fall also die Level-2-Ebene teilweise mit der Individualebene zusammenfällt, bieten kleinere Einheiten, wie die Haushalt-, Nachbarschafts- oder auch Kommunalebene für politikwissenschaftliche MEA im Prinzip einen interessanten, bislang weitgehend vernachlässigten Forschungsrahmen. Zum einen ist die Homogenität innerhalb kleinerer Gruppen zumeist deutlich größer als dies bei so großen

Kontexteinheiten wie dem Nationalstaat der Fall ist. Zum anderen sind zwar die IKKs der Artikel in unserem Datensatz, die subnationale Gebietskörperschaften als Kontextebene heranziehen, nicht bedeutend größer als die der Artikel mit nationalen Level-2-Einheiten, jedoch stellt sich bei diesen kleineren Gruppen seltener das Problem zu geringer Fallzahlen auf der zweiten Ebene. Da die Auswahl der Untersuchungseinheiten erfahrungsgemäß auch immer mit der Zugänglichkeit der Daten im Zusammenhang steht, erscheint es an dieser Stelle sinnvoll, auch dafür zu werben, sich von dem starken Fokus der Politikwissenschaft auf die Ebene des Nationalstaats zumindest ein Stück weit zu lösen und verstärkt Datensätze für subnationale Einheiten zu erstellen.

Neben der Verhaltens- und Einstellungsforschung gibt es aber durchaus auch andere Bereiche, in denen die MEA gewinnbringend angewendet werden kann. Wie wir in Kapitel 2 dargelegt haben, ist eine MEA immer dann sinnvoll, wenn sich die Einheiten auf der ersten Ebene eindeutig jeweils einer einzigen Aggregateinheit auf höherer Ebene zuordnen lassen, wobei hierfür nicht zwangsläufig Individuen auf der ersten Ebene notwendig sind. Zum einen ist es auch möglich, Aggregateinheiten (z.B. subnationale Gebietskörperschaften) als Level-1 zu betrachten, solange sie sich in einer höheren Ebene (z.B. dem Nationalstaat) gruppieren lassen. Zum anderen besteht aber auch die Möglichkeit, sich gänzlich von Personen oder Gruppen von Personen als Bezugseinheit zu lösen. Ein besonders innovatives Beispiel aus unserem Datensatz, das in diese Richtung arbeitet, stellt das Paper von Cranmer (2011) dar, die die populistischen Tendenzen in der politischen Kommunikation in der Schweiz untersucht. Dafür wurden 1174 öffentliche Reden von eidgenössischen Politikern, die in Bezug auf zwei Abstimmungen zur Asyl- und Einwanderungspolitik gehalten wurden, auf populistische Äußerungen durchsucht. Cranmer erstellt einen Populismusindex von 0 bis 13, der als abhängige Variable dient. Die einzelnen Reden stellen somit die erste Ebene ihrer Analyse dar, auch wenn einige individuelle Merkmale der Redner wie die Parteizugehörigkeit sowie die Kontrollvariablen Bildungsgrad, politische Erfahrung, Alter und Geschlecht als erklärende Variablen auf Level-1 dienen.¹⁴ In der Annahme, dass sich die Reden innerhalb einer Parlaments-, Ausschuss- oder Talkshowdebatte aufeinander beziehen und sich somit stark gegenseitig beeinflussen, gruppiert Cranmer die Reden in die 204 Debatten, in denen Sie gehalten wurden. Die Berechnung des IKK ergibt einen Wert von 0,1. Der Einbezug von Debattenmerkmalen kann also ungefähr 10% der Varianz des Populismus in den Reden erklären. Dies ist zwar nicht besonders viel, lässt jedoch eine MEA auch nicht völlig unangebracht erscheinen. Die Autorin geht weiterhin davon aus, dass der Grad des Populismus in politischen Reden unter anderem davon abhängt, in welchem Kontext die Rede gehalten wurde. Als Level II-Variablen nimmt Cranmer daher den Öffentlichkeitsgrad der Debatte, die zeitliche Distanz der Debatte zum Abstimmungstermin sowie einen Dummy, der angibt, ob es sich um eine „entrance debate“ handelt, in ihre Analyse auf. Sie kommt zu dem Schluss, dass insbesondere der

¹⁴ Dieses Vorgehen erscheint in diesem Fall insofern gerechtfertigt, als dass vermutlich die meisten der Reden von unterschiedlichen Politikern gehalten wurden. Gesetzt dem Fall, dass nur sehr wenige Politiker sehr viele Reden gehalten hätten, wäre diese Disaggregation von individuellen Eigenschaften weniger Politiker auf viele Reden nicht zulässig.

Öffentlichkeitsgrad der Debatte einen starken Einfluss auf die Verwendung populistischer Äußerungen von Politikern hat, wobei eine große Öffentlichkeit populistische Tendenzen erwartungsgemäß befördert. Durch den Einbezug von Interaktionseffekten in die MEA wird außerdem nachgewiesen, dass dieser Effekt insbesondere für die Schweizer Volkspartei zutrifft, auch wenn sich die Parteizugehörigkeit des Redners allein nicht als signifikante Variable gezeigt hatte.

Ein weiterer interessanter Einsatz der MEA findet sich in einem Artikel von Tor Midtbø, der 2011 in der Zeitschrift *Scandinavian Political Studies* veröffentlicht wurde. Darin untersucht er für ein Sample von 153 norwegischen Abgeordneten, welche Politiker in den Medien besonders häufig und besonders positive Berichterstattung erhalten, bzw. welche Politiker eher schlechte oder gar keine Presse bekommen. Dafür analysiert er 18.144 Artikel aus 14 norwegischen Zeitungen im Zeitraum von 2001-2005. Die Artikel stellen hier die Einheiten der ersten Ebene dar, individuelle Eigenschaften der Abgeordneten sind auf Level-2 angesiedelt. Dahinter steht die Vermutung, dass sich die Berichterstattung über einen Abgeordneten in verschiedenen Artikeln tendenziell ähnelt und die Artikel daher sinnvoll über die Individuen gruppiert werden können. Auf Level-1 wurden der Zeitpunkt der Veröffentlichung des Artikels (Wahljahr oder kein Wahljahr) und die Art des Artikels (Boulevardzeitung oder Qualitätszeitung) als unabhängige Variablen verwendet. Auf Level-2 finden sich – ähnlich wie im vorigen Beispiel auf der ersten Ebene – Individualmerkmale der Abgeordneten wie Alter, Geschlecht und politische Erfahrung, sowie die Parteimitgliedschaft und die Position innerhalb dieser Partei. Auch hier weist der IKK nur Werte von 0,09 bzw. 0,13 auf, was darauf hindeutet, dass insgesamt relativ ausgeglichen über die unterschiedlichen Personen berichtet wurde. Auch wenn die Erklärungskraft der zweiten Ebene somit auch in diesem Fall nicht besonders beeindruckend ist, deckt die Analyse doch einige interessante Tendenzen auf. So wird in den norwegischen Medien beispielsweise über weibliche Abgeordnete zwar weniger, dafür aber positiver berichtet. Das Gegenteil lässt sich über Parteivorsitzende konstatieren. Auf Level-1 erweist sich die Zeitungsart als besonders einflussreich, wobei dieses Ergebnis aufgrund der durchgeführten Disaggregation der Zeitungsmerkmale auf Artikelebene mit Vorsicht zu genießen ist.¹⁵

Zuletzt soll hier noch kurz darauf eingegangen werden, dass, wie wir oben gesehen haben, auch die Zeit als Analyseeinheit genutzt werden kann. Diese Anwendung der MEA kam bislang vor allem in der Wahlforschung zum Einsatz. Zum einen besteht die Möglichkeit, die Zeit als Level-2-Einheit zu modellieren und beispielsweise das Wahlverhalten von Individuen über bestimmte Jahre hinweg zu gruppieren. In diesem Fall stellt die Wahlentscheidung und eventuell weitere Individualvariablen die erste Ebene dar, während die Wahljahre mit ihren jeweiligen Umständen (wie z.B. die staatliche Wirtschaftslage, besondere politische Ereignisse ect.) die zweite Ebene bilden.

¹⁵ Da davon auszugehen ist, dass sich die Berichterstattung bezüglich eines Abgeordneten innerhalb einer Zeitung ähnelt und es gleichzeitig Unterschiede hierin zwischen den Zeitungen gibt, wäre es an dieser Stelle sinnvoll gewesen die Zeitung als weitere Ebene zwischen Artikel und Parlamentarier einzuziehen. Hierdurch würde man eine künstliche Erhöhung der Fallzahl mit den oben geschilderten negativen Konsequenzen vermeiden.

Zum anderen ist es aber auch möglich, die MEA als Alternative zur gepoolten Zeitreihenanalysen zu verwenden. Wenn beispielsweise Paneldaten für das Wahlverhalten bestimmter Bürger vorliegen, können die einzelnen Beobachtungszeitpunkte als erste Ebene verwendet werden. Wie im obigen Beispiel bei den Zeitungsartikeln fungieren in diesem Fall die Individuen als „Gruppen“ auf Level-2, denen die einzelnen Beobachtungen zugeordnet werden können. Ein anderes Beispiel für den Einsatz der MEA zur Analyse von TSCS-Daten bietet ein Artikel von LaFree und Andromachi (2013). Die beiden untersuchen in ihrer Arbeit den Zusammenhang zwischen Mordraten und Demokratie. Überprüft wird für 44 Staaten, ob sich die Mordrate mit zunehmender Demokratisierung absenkt, ob das Gegenteil der Fall ist, oder ob die These zutrifft, dass die Mordrate in der Transitionsphase stark ansteigt, um sich dann über die Jahre abzusenken. Die jährliche Mordrate zum Zeitpunkt $t = 1, 2, 3 \dots$ nach der ersten Aufzeichnung bildet die abhängige Variable auf Level I. Da davon auszugehen ist, dass die Mordrate zum Zeitpunkt $t = 1$ in einem Land mit der Mordrate von $t = 2$ in demselben Land zusammenhängt, werden die Beobachtungen über die jeweiligen Nationalstaaten gruppiert. Die MEA wird hier somit als Werkzeug zur Handhabung von Autokorrelationsproblemen verwendet. Auf der ersten Ebene wird zusätzlich zum Demokratieniveau auch auf den Anteil der jungen Leute in der jeweiligen Bevölkerung kontrolliert, auf der Aggregatebene werden regionale Dummyvariablen und wirtschaftliche Eigenschaften des Landes als erklärende Variablen überprüft. Die durchgeführte MEA bestätigt schließlich die Modernisierungsthese, die einen Gewaltanstieg für den begrenzten Zeitraum der Transitionsphase vermutete.

Im Vergleich zur herkömmlichen Zeitreihenanalyse liegt ein großer Vorteil der Mehrebenenanalyse darin, dass sich flexibler ist im Umgang mit unausgeglichenen und unvollständigen Datensätzen (Snijders & Bosker, 1999; Skrondal & Rabe-Hesketh, 2008). Inwiefern sie der gepoolten Zeitreihenanalyse aber tatsächlich methodisch überlegen ist, wird sich erst mit der zunehmenden Anwendung und Weiterentwicklung der Methode in diesem Bereich in den nächsten Jahren abschließend beurteilen lassen. Denn dem methodisch vergleichsweise ausgereiften Instrumentarium – zu einem großen Teil aus der Ökonometrie übernommen – welches die Politikwissenschaft augenblicklich zur Analyse von TSCS-Daten anwendet, steht bislang von MEA-Seite aus noch zu wenig entgegen.¹⁶

Fazit

Unsere Metaanalyse zeigt, dass die Methode der MEA in der Politikwissenschaft zwar in den letzten Jahren deutlich an Bedeutung gewonnen hat, jedoch im Vergleich zur Nachbardisziplin Soziologie sowohl was die Zahl der Veröffentlichungen, als auch was die Zahl der Zitationen angeht noch einige Jahre zurück liegt. Betrachtet man die kontinuierlichen Zuwachsraten an politikwissenschaftlichen

¹⁶ Eine positive und daher hier zu berichtende Ausnahme stellt das Buch von Hox (2010) dar. In diesem werden insbesondere die Anknüpfungspunkte der MEA zu anderen statistischen Techniken wie der Survival-Analyse oder Strukturgleichungsmodellen aufgezeigt.

ME-Artikeln in den letzten fünf Jahren und die noch vorhandenen Möglichkeiten ihrer Anwendung, kann jedoch davon ausgegangen werden, dass sich die MEA auch in der Politikwissenschaft zu einer der quantitativen Standardmethoden entwickeln wird, die im Werkzeugkoffer eines jeden Politologen vorzufinden sein sollte. Ein weiterer interessanter Befund der Meta-Studie ist, dass rund drei Viertel der Autoren die MEA in vergleichsweise klassischer Weise nutzen, um den Einfluss insbesondere staatlicher Kontextfaktoren auf Einstellung und Verhalten von Individuen, d.h. Personen unter die Lupe zu nehmen. Obgleich sich die MEA für diese Fragen zweifellos sehr anbietet, so zeigen einzelne innovative Arbeiten doch auch wie flexibel diese Methode auf ganz andere Fragestellungen angewendet werden kann, bei denen die Datenstruktur weit ab von dieser klassischen ME-Struktur liegt. Solche Analysen sind nicht komplizierter als klassische Regressionsanlagen, einzig die Herangehensweise an unsere Fragestellungen müssen wir überdenken. Die Flexibilität, welche die MEA bietet, erfordert eben auch vom Forscher eine gewisse Flexibilität wenn es darum geht Fragestellungen zu überlegen, Datensätze zu erstellen und Hypothesen zu operationalisieren. Der Appell an die Politikwissenschaft kann daher nur lauten noch stärker als bisher die ausgetretenen methodischen Pfade zu verlassen, quer zu denken und hierdurch unsere Fragestellungen aus einer anderen Perspektive wahrzunehmen. Als Analysewerkzeug kann die MEA dann von großem Nutzen sein.

Literatur

- Barton, Allen H. (1968). Bringing Society Back In. *Survey Research and Macro-Methodology. American Behavioral Scientist*, 12(2), XXX.
- Braun, D., Seher, N., Tausendpfund, M., & Wolsing, A. (2010). Einstellungen gegenüber Immigranten und die Zustimmung zur Europäischen Integration - eine Mehrebenenanalyse, *Working Paper* (Vol. 136). Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.
- Browne, W. J., & Draper, D. (2000). Implementation and Performance Issues in the Bayesian and Likelihood Fitting of Multilevel Models. *Computational Statistics*, 15(3), 391-420.
- Cranmer, M. (2011). Populist Communication and Publicity: An Empirical Study of Contextual Differences in Switzerland. *Swiss Political Science Review*, 17(3), 286-307.
- van Deth, Jan W. (1995). Comparative politics and the decline of the nation-state in Western Europe. *European Journal of Political Research*, X(4), XXX-XXX.
- Hamaker, E. L., & Klugkist, I. (2011). Bayesian Estimation of Multilevel Models. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of Advanced Multilevel Analysis* (pp. 137-162). New York: Taylor and Francis.
- Hox, J. J. (1995). *Applied Multilevel Analysis*: TT-Publikaties.
- Hox, J. J. (2002). *Multilevel Analysis - Techniques and Applications*. New York: Taylor & Francis.
- Hox, J. J. (2010). *Multilevel analysis: techniques and applications*. New York: Routledge.

- Johnston, R. et al. (2005). A missing level in the analyses of British voting behavior: the household as context as shown by analyses of a 1992-1997 longitudinal survey. *Electoral Studies*, 24(2), 201-255.
- Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The Effect of Different Forms of Centering in Hierarchical Linear Models. *Multivariate Behavioral Research*, 30(1), 1.
- LaFree, G. & Tseloni, Andromachi (2006). Democracy and Crime: A Multilevel Analyses of Homicide Trends in Forty-Four Countries, 1950-2000. *Annals of the American Academy of Political and Social Science*, 605(1), 26-49.
- Maas, C. J. M. & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, 58(2), 127-137.
- Mey, G. & Mruck, K. (2011). Grounded Theory Methodologie: Entwicklung, Stand, Perspektiven. In G. Mey & K. Mruck (Eds.), *Grounded Theory Reader* (pp. 11-48). Wiesbaden: Springer.
- Midtbø, T. (2011). Explaining Media Attention for Norwegian MPs: A New Modelling Approach. *Scandinavian Political Studies*, 34(3), 226-249.
- Paccagnella, O. (2006). Centering or Not Centering in Multilevel Models? The Role of the Group Mean and the Assessment of Group Effects. *Evaluation Review*, 30(1), 66-85.
- Peffley, M., & Rohrschneider, R. (2003). Democratization and Political Tolerance in Seventeen Countries: A Multi-level Model of Democratic Learning. *Political Research Quarterly*, 56(3), 243-257.
- Pötschke, M. (2006). Mehrebenenanalyse. In J. Behnke, T. Gschwend, D. Schindler & K.-U. Schnapp (Eds.), *Methoden der Politikwissenschaft. Neuere qualitative und quantitative Analyseverfahren* (pp. 167-179). Baden-Baden: Nomos.
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational Tools for Probing Interactions in Multiple Linear Regression, Multilevel Modeling, and Latent Curve Analysis. *Journal of Educational and Behavioral Statistics*, 31(4), 437-448.
- Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, 25, 111-163.
- Raudenbush, S. W. (1989). "Centering" predictors in multilevel analysis: Choices and consequences. *Multilevel Modelling Newsletter*, 1(2), 10-12.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*: SAGE Publications.
- Rosar, U. (2003). Die Einstellungen der Europäer zum Euro. Ein Anwendungsbeispiel der Mehrebenenanalyse als Instrument komparativer Umfrageforschung. In S. Pickel, G. Pickel, H.-J. Lauth & D. Jahn (Eds.), *Vergleichende politikwissenschaftliche Methoden: neue Entwicklungen und Diskussionen* (pp. 221-245). Wiesbaden: Westdeutscher Verlag.
- Snijders, T., & Bosker, R. (1994). Modeled Variance in Two-Level Models. *Sociological Methods & Research*, 22(3), 342-363.
- Snijders, T., & Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.
- Snijders, T., & Bosker, R. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.
- Snijders, T. A. B. (2005). Power and sample size in multilevel modeling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Vol. 3, pp. 1570-1573). Chichester: Wiley.
- Skrondal, A. & Rabe-Hesketh, S. (2008). Multilevel and Related Models for Longitudinal Data. In J. de Leeuw & E. Meijer (Eds.), *Handbook of Multilevel Analysis* (pp. 275-299). Wiesbaden: Springer.
- Steele, F. (2009). Module 5: Introduction to Multilevel Modelling Concepts: Centre for Multilevel Modelling.
- Stegmueller, D. (2013). How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches. *American Journal of Political Science*, 57(3), 748-761.