

Statistische Methodik zur Wählerwanderungsanalyse der Stadt München

André Klima, Thomas Schlesinger, Paul W. Thurner und Helmut Küchenhoff

Statistisches Beratungslabor und Geschwister Scholl Institut
für Politische Wissenschaft der LMU München

1. Einführung

Für die Wählerwanderungsanalyse der Stadt München wurde vom Statistischen Beratungslabor der LMU eine Studie zur statistischen Methodik durchgeführt. Ziel war es, gebräuchliche Verfahren zu evaluieren und eine klare und transparente, wissenschaftlich abgesicherte, nachvollziehbare Strategie zur Schätzung der Wählerwanderung in der Praxis zu finden. Die wesentlichen Ergebnisse der Studie und das Vorgehen sind im Folgenden kurz beschrieben. In Abschnitt 2 wird die Problemstellung dargestellt. In Abschnitt 3 finden sich die zentralen Ergebnisse zum Methodenvergleich, in Abschnitt 4 werden Details zu dem von uns verwendeten Verfahren gegeben.

2. Problemstellung und Datengrundlage

Da Befragungen am Wahltag, wie sie von verschiedenen Umfrageinstituten durchgeführt werden, aufwändig sind, wurden im Folgenden zunächst nur Strategien berücksichtigt, die als Datengrundlage die Wahlergebnisse der einzelnen Stimmbezirke verwenden.

Das Problem bei der Abschätzung der Wählerwanderung besteht dann darin, dass nur die Aggregatenergebnisse der beiden Wahlen bekannt sind. Wenn die Partei A beispielsweise 100 000 Stimmen bei der Wahl 2014 und 130 000 der Stimmen bei der Wahl 2020 bekommen hat, so kann man daraus nicht direkt ableiten, wie sich die individuellen Wähler verhalten haben. Ein (unwahrscheinlicher) Extremfall wäre, dass alle 130 000 Wähler 2014 eine andere Partei gewählt hatten und dass die 100 000 früheren Wähler zu einer anderen Partei gewechselt sind oder nicht gewählt haben. Ein anderer Extremfall wäre, dass alle 100 000 Wähler aus dem Jahr 2014 die Partei A erneut gewählt haben und dass noch 30 000 Wähler von anderen Parteien und von den Nichtwählern hinzugekommen sind.

Trotz dieser Unbestimmtheit ist es möglich, Methoden zur Schätzung der Wählerwanderung auf der Basis von Aggregatenergebnissen zu entwickeln. Die Grundidee besteht darin, die einzelnen Wahlergebnisse, die ja für alle Stimmbezirke jeweils getrennt vorliegen, in geeigneter Weise zu verwenden. Man spricht in der Literatur dann allgemein von Ökologischer Inferenz. Darunter versteht man solche Strategien, bei denen aus Daten höherer Ebene (hier die Stimmbezirke) auf individuelle Parameter (Wählerwanderung) geschlossen werden soll. Dabei ist es notwendig, dass die überprüften Methoden der speziellen Parteienlandschaft Münchens gerecht werden und Wählerwanderungen zwischen einer Vielzahl an Parteien geschätzt werden können. Dafür wurden in der wissenschaftlichen Literatur verschiedene Ansätze vorgeschlagen, die im Rahmen unsere Studie verglichen wurden.

3. Methodenvergleich

Alle Verfahren benötigen zusätzliche Annahmen, um eine Abschätzung durchführen zu können. Die einfachste, aber unrealistische Annahme ist, dass das Wählerwanderungsverhalten in allen Stimmbezirken näherungsweise gleich ist. Dies führt zu dem Verfahren der ökologischen Regression nach Goodman (1953). Ein weiterer älterer Ansatz ist der von Thomsen (1987). Dieser Ansatz ist komplexer und arbeitet mit einem Logit-Modell und der Gleichsetzung von Korrelationen auf Individualebene und auf Ebene der Gebietseinheiten. Zur Berechnung wurde von Thomsen eine entsprechende Software zur Verfügung gestellt.

Neuere Ansätze sind sog. Hierarchische Bayesianische Modelle. Diese sind in der Lage, mögliche Schwankungen in dem Wählerwanderungsverhalten zwischen den Bezirken in die Modellierung einzubeziehen. Ein Vorschlag aus dieser Modellgruppe ist das Multinomial-Dirichlet Modell von Rosen, Jiang, King and Tanner (2001). Zu dieser Strategie gibt es eine Software-Implementierung (eiPack) in dem Programmpaket R (Lau, Moore, Kellermann 2007).

Ein weiterer Vorschlag stammt von Andreadis und Chadjipadelis (2009), der von Kellermann (2011) modifiziert wurde. Dabei handelt es sich um iterative Algorithmen, die eine Anwendung von Methoden für den 2 x 2 Spezialfall der Ökologischen Inferenz ermöglichen. Die beiden iterativen Algorithmen unterscheiden sich leicht in der Reihenfolge bei der Durchführung der einzelnen Schritte.

In einer ausführlichen Simulationsstudie wurden die verschiedenen Methoden verglichen. Es stellte sich dabei heraus, dass das Multinomial-Dirichlet Modell am besten abschnitt. Die durchschnittliche Abweichung (Fehlerquote) lag allerdings immer noch bei ca. 10 - 20%-Punkten, was wohl auf die Grundproblematik der Ökologischen Inferenz zurückzuführen ist. Ähnliche Ergebnisse ergaben sich auch bei einer Simulationsstudie von Wählerwanderungen im Saarland. Daher wurde das Verfahren nach dem Multinomial-Dirichlet Modell für die Schätzung der Wählerwanderung in München bei der Oberbürgermeister*in-Wahl 2020 ausgewählt. Im folgenden Abschnitt werden einige Details dazu beschrieben.

4. Die Berechnung der Wählerwanderung in München

Zur Berechnung der Wählerwanderung mit aggregierten Daten müssen in einem ersten Schritt über die Zeit homogene Gebiete gebildet werden. Da es zu großen Veränderungen im Zuschnitt der Stimmbezirke gekommen ist, wurde statt größerer Anpassungen der verwendeten Gebietseinheiten das Ergebnis der Wahl 2014 auf die Grenzen der Wahl 2020 umgerechnet und als Input für die Wahl 2014 benutzt. Insgesamt stehen somit alle 755 Stimmbezirke in München als Gebietseinheiten für die Berechnungen zur Verfügung.

Die Briefwähler werden anteilig auf die entsprechenden Stimmbezirke aufgeteilt. Dazu wird die Zuordnung der Stimmbezirke zu den Briefwahlbezirken genutzt. Wenn vorhanden werden auch die Informationen zur Zahl der Wahlscheinbeantragungen bei der Verteilung genutzt. In der Simulationsstudie zeigte sich nämlich, dass bei Verteilung der Briefwähler bessere Ergebnisse erreicht werden können.

Veränderungen bei der Zahl der Wahlberechtigten in den Stimmbezirken werden durch eine proportionale Anpassung der Wahlergebnisse ausgeglichen. Dies bedeutet im Detail, dass das Wahlergebnis in Prozent bei der Oberbürgermeister*in-Wahl 2014 für jede Partei gewahrt wird, während die Zahl der Stimmen so angepasst wird, dass sie in der Summe der Zahl der Wahlberechtigten bei der Oberbürgermeister*in-Wahl 2020 entspricht. Bei diesem Vorgehen wird impliziert, dass sich das Wahlverhalten durch die veränderte Zahl der Wahlberechtigten nicht strukturell geändert hat. Ein Nebeneffekt dieser Anpassung ist jedoch, dass die absoluten Stimmergebnisse für eine Partei nicht mehr mit dem Wahlergebnissen übereinstimmen müssen.

Ungültige Stimmen werden bei der Datenaufbereitung den Nichtwählern zugeschlagen.

Mit den so aufbereiteten Daten wird anschließend die Wählerwanderung mit dem Multinomial-Dirichlet Modell berechnet. Die Ergebnisse können als absolute Anzahlen, Anteile der Wähler der ersten Wahl („Wohin sind die Wähler gegangen?“) und als Anteile der Wähler bei der zweiten Wahl („Woher kommen meine Wähler?“) dargestellt werden. Gebräuchlich ist auch die Darstellung der Wanderungssalden, bei der die gegenseitigen Wanderungsströme der Parteien aufgerechnet werden und nur die zahlenmäßigen Gewinne an Wählern von einer Partei oder Verluste an eine Partei angegeben werden.

Literatur

- Andreadis, I., Chadjipadelis, T. (2009). A Method for the Estimation of Voter Transition Rates. *Journal of Elections, Public Opinion and Parties*, 19, 203 - 218.
- Goodman, L.G. (1953) Ecological Regressions and Behavior of Individuals. *American Sociological Review*, 18, 663 - 664.
- Greiner, D.J., Quinn, K.M. (2010). Exit Polling and Racial Bloc Voting: Combining Individual-Level and R x C Ecological Data. *The Annals of Applied Statistics*, 4, 1774 - 1796.
- Grofman, B., Merrill, S. (2004). Ecological Regression and Ecological Inference. In: *Ecological Inference: New Methodological Strategies*. eds. King, G., Tanner, M.A., Rosen, O.. Cambridge: Cambridge University Press.
- Kellermann, T. (2011). Vom Wahlergebnis zur Wählerwanderung. *Stadtforschung und Statistik*, 1, 34 - 40. Lau, O., Moore, R.T., Kellermann, M. (2007). eiPack: R x C Ecological Inference and Higher-Dimension Data Management. *R News*, 7, 43 - 47.
- Rosen, O., Jiang, W., King, G., Tanner, M.A. (2001). Bayesian and Frequentist Inference for Ecological Inference: The R x C Case. *Statistica Neerlandica*, 55, 134 - 156.
- Thomsen, S.R. (1987). *Danish Elections 1920-79: A Logit Approach to Ecological Analysis and Inference*. Aarhus: Politica.