DEUTSCHES FORSCHUNGSINSTITUT FÜR ÖFFENTLICHE VERWALTUNG SPEYER
GERMAN RESEARCH INSTITUTE FOR PUBLIC ADMINISTRATION SPEYER

Tim Jäkel

Wer vergleicht seine Leistung, wenn er hohe Schulden hat?

Empirische Evidenz aus den deutschen kreisfreien Städten

FÖV 74
Discussion Papers

ISSN 1868-971X (Print)
ISSN 1868-9728 (Internet)

Tim Jäkel

Wer vergleicht seine Leistung, wenn er hohe Schulden hat?

Empirische Evidenz aus den deutschen kreisfreien Städten

FÖV 74
Discussion Papers

Deutsches Forschungsinstitut für öffentliche Verwaltung Speyer

2013

Gefördert durch die Bundesrepublik Deutschland

Nicht im Buchhandel erhältlich
Schutzgebühr: € 5,-
Bezug: Deutsches Forschungsinstitut
für öffentliche Verwaltung Speyer
Postfach 14 09
67324 Speyer
http://www.foev-speyer.de

ISSN 1868-971X (Print)
ISSN 1868-9728 (Internet)

Dr. Tim Jäkel

Forschungsreferent im Teilprojekt 1 „Benchmarking in der öffentlichen Verwaltung: ein europäischer Ländervergleich“ des Projektverbundes „Leistungsvergleich und Benchmarking im Öffentlichen Sektor“ am Deutschen Forschungsinstituts für öffentliche Verwaltung Speyer unter der Leitung von Univ.-Prof. Dr. Sabine Kuhlmann

Inhaltsverzeichnis

Abbildungsverzeichnis VII

Zusammenfassung 1

1. Einleitung und Relevanz 2

2. Der Zusammenhang zwischen Schuldenstand und Leistungsvergleichen: Theoretische Überlegungen 4

3. Ein ökonometrisches Modell 8

4. Darstellung und Diskussion der Schätzergebnisse 14

5. Schlussfolgerungen 25

Literatur 27

Appendix 29

Abbildungsverzeichnis

Abbildung 1:	Kreisfreie Städte im Jahr 2005	5
Abbildung 2:	Lage- und Streuungsparameter der Variablen	13
Abbildung 3:	Korrelationsmatrix der erklärenden Variablen	14
Abbildung 4:	Parameter der geschätzten Modelle	14
Abbildung 5:	Marginale Veränderungen in der latenten Variable y* und Effektkoeffizienten ^	19
Abbildung 6:	Vorhergesagte Wahrscheinlichkeiten für Beobachtung Nr. 1	20
Abbildung 7:	Übersicht Predicted Probabilities für y = 1, y = 2, y = 3	21
Abbildung 8:	Predicted probabilities für Beobachtungen y = 3 (häufige Nutzung) ^	22
Abbildung 9:	Predicted probabilities für Beobachtungen y = 1 (selten Nutzung/keine Nutzung/ Nutzung in der Vergangenheit)	23
Abbildung 10:	Predicted probabilities für Beobachtungen y = 2 (gelegentliche Nutzung)	23
Abbildung 11:	Predicted und Cumulative Probabilities für zunehmende Pro-Kopf-Verschuldung	25
Abbildung 12:	brant-Teststatistik Modell Eins	29
Abbildung 13:	brant-Teststatistik Modell Drei	29
Abbildung 14:	brant-Teststatistik Modell Fünf	30
Abbildung 15:	brant-Teststatistik Modell Sechs	30

Zusammenfassung[1]

Der Beitrag zeigt für ein repräsentatives Sample von kreisfreien Städten, dass eine hohe Pro-Kopf Verschuldung negativ mit der Nutzungshäufigkeit von Leistungsvergleichen assoziiert ist. Relativ hoch verschuldete kreisfreie Städte scheuen offenbar die potentiellen politischen Kosten (in Form von Wiederwahlrisiken und/oder Reputationsverlust), die durch ein schlechtes Abschneiden bei einem Leistungsvergleich entstehen können. Diese potentiellen politischen Kosten, zusammen mit Opportunitäts- und Transaktionskosten, übersteigen den potentiellen Nutzen, den vergleichende Standortbestimmung und Erfahrungsaustausch mit sich bringen können. Lerneffekte können nicht stattfinden, mögliche Wohlfahrtssteigerungen werden nicht genutzt. Umgekehrt erhöht eine hohe Arbeitslosenquote die Wahrscheinlichkeit einer intensiven Nutzung von Leistungsvergleichen. Auf diese makroökonomische Kennzahl gibt es jedoch weniger direkte Einflussmöglichkeiten als auf die Verschuldungssituation. Entgegen den theoretischen Erwartungen hat die Größe einer Verwaltung (gemessen über die Zahl der Mitarbeiter) keinen signifikanten Einfluss auf die Wahrscheinlichkeit einer häufigen Nutzung. Die Schlussfolgerung aus den empirischen Befunden lautet, dass Maßnahmen zur kommunalen Entschuldung, insbesondere Entschuldungspakte zwischen Land und Kommunen, vorangetrieben werden sollten, denn sie erhöhen den (politischen) Spielraum zur Nutzung moderner Steuerungsinstrumente in der Kommunalverwaltung.

Schlüsselwörter: Leistungsvergleich, Benchmarking, Kommunen, Ordinal Regression Model, öffentliche Finanzen, öffentliche Verwaltung

---

[1] Der vorliegende Beitrag wurde präsentiert auf der International Graduate Conference “Crisis and Governance in Europe: Implications for State, Market and Society” in Speyer, 18.-19. April 2013. Weitere Informationen zum laufenden Forschungsprojekt „Benchmarking in der öffentlichen Verwaltung: ein europäischer Ländervergleich“ am Deutschen Forschungsinstitut für öffentliche Verwaltung Speyer (FÖV) finden Sie unter www.foev-speyer.de/benchmarking. Das Deutsche Forschungsinstitut für öffentliche Verwaltung Speyer ist Mitglied der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL).

1. Einleitung und Relevanz

Unternehmen vergleichen ihre Performanz mit der von anderen Unternehmen, um eigene Schwachstellen aufzudecken und um zu lernen. Auch öffentliche bürokratische Organisationen wie z.B. Kommunen vergleichen ihre Leistung untereinander. Es gibt zwischen ihnen zwar keinen direkten Wettbewerb wie zwischen privaten Unternehmen, aber Leistungsvergleiche sollen Wettbewerb zumindest ansatzweise simulieren und zu positiven Effekten führen (*Seckelmann* 2010, *Kuhlmann* 2011, *Jäkel/Kuhlmann* 2012, *Kuhlmann/Jäkel* 2013). Konkret verspricht man sich folgende Vorteile: Durch einen Vergleich mit anderen Einheiten kann man seine eigene Performanz (anhand von Kennzahlen) einordnen, man erkennt Schwachstellen und kann diese angehen, schlechte Leistung wird so verhindert (*Behn* 2007). Der Vergleich zeigt auch, welche Kommune besonders gute Leistung erbringt. Die anderen Kommunen können von diesen Vorreitern lernen und dadurch ihre eigene Leistung verbessern. Theoretisch verringern sich dadurch auf lange Sicht die Leistungsunterschiede in der Gruppe und das Leistungsniveau steigt an (*Meyer/Gupta* 1994).

Zahlreiche Untersuchungen haben allerdings gezeigt, dass das Arbeiten mit Kennzahlen, Vergleichsdaten und Ranglisten verschiedene unerwünschte Nebenwirkungen hat, in Unternehmen, aber auch in öffentlichen Verwaltungen. Wenn das Abschneiden bei einem Vergleich an Sanktionen oder Belohnungen gekoppelt ist, gibt es für die Verantwortlichen starke Anreize, sich strategisch zu verhalten, d.h. sein Verhalten möglichst an den Spielregeln des Vergleichs auszurichten. Die „Spielregeln“, das sind die Kennzahlen, die vorab festgelegt werden, wie und welche Bereiche von Leistung man misst und dann vergleicht[2]. Zu den Spielregeln gehört auch, ob es am Ende eine Rangliste und eine Veröffentlichung geben soll und wie aus den verschiedenen Indikatoren und Daten dann die Rangplätze berechnet werden. Eine Folge von strategischem Verhalten wäre, dass eine Organisation ihre selbst gesetzten Ziele zu Gunsten anderer Prioritäten zurückstellt,

---

2 Die wesentliche Aufgabe von öffentlichen Verwaltungen, in unserem Fall Kommunen, ist es, öffentliche Güter und Dienstleistungen für ihre Einwohner bereitzustellen bzw. zu erbringen. Dabei spielen die Kosten pro Einheit (Effizienz), die Wirksamkeit mit Blick auf bestimmte Ziele, die man erreichen will (Effektivität des Mitteleinsatzes), aber auch die Qualität der Güter und Dienstleistungen eine Rolle. Die verschiedenen Dimensionen von Leistung lassen sich in der Regel nicht mit einer einzigen Kennzahl abbilden.

die sich aus den externen Benchmarks des Vergleichs ergeben.[3] Andere extremere Formen von strategischem Verhalten sind Datenselektion (cream skimming) bis hin zur Datenmanipulation (*de Bruijn* 2002, *Van Dooren* et al. 2010). Falls es keine staatlichen Sanktionen gibt, die Ranglisten mit Vergleichsergebnissen aber veröffentlicht werden, wie das z. B. bei Rankings von Zeitungen über Lebensqualität in Städten oder die Preise im öffentlichen Nahverkehr der Fall ist, gibt es für gewählte Entscheidungsträger wegen der öffentlicher Prangerwirkung solcher Ranglisten je nach Bedeutung des Themas ebenfalls Anreize, sich so zu verhalten, dass man zumindest nicht auf den letzten Plätzen landet.

Die Art und Weise, wie die Leistung von Kommunen gemessen und verglichen wird, variiert sehr stark von Land zu Land. In zentralistischen Ländern wie England gibt es verpflichtende Inspektionen und Leistungsvergleiche, die alle lokalen Verwaltungseinheiten abdecken. In Deutschland dominieren dagegen freiwillige Leistungsvergleiche der Kommunen, z. B. unter dem Dach von Kommunalverbänden wie der KGSt (*Kuhlmann/Jäkel* 2013). Die Verbreitung von Benchmarking hängt hier von der Bereitschaft jeder einzelnen Kommune ab, die Möglichkeiten dieses Instruments gegen die Kosten des Einsatzes abzuwägen.

Qualitative und quantitative Projekt- und/oder Länderfallstudien haben eine Reihe von kritischen Erfolgsfaktoren herausgearbeitet, die die Teilnahmebereitschaft an Leistungsvergleichen beeinflussen. Johansson und Siverbo (2006) zeigen, dass Budgetdefizite die Anwendungshäufigkeit von Vergleichsinstrumenten in schwedischen Kommunen erhöhen, während linke Mehrheiten im Rat sowie ein starker Parteienwettbewerb die Wahrscheinlichkeit von Vergleichsarbeit in einer Kommune negativ beeinflussen. Studien über kommunale Benchmarking-Projekte in den USA (*Dalehite* 2008, *Ammons/Rivenbark* 2008, *Julnes/Holzer* 2001) zeigen darüber hinaus, dass die Ausgestaltung der technischen Details eines konkreten Projektes ebenfalls über Erfolg oder Nichterfolg mitentscheidet.

---

3 Falls die ursprünglichen Prioritäten einer Kommune von den extern gesetzten Standards abweichen, geht eine solche Verschiebung von lokalen Prioritäten zu Lasten der Wähler der jeweiligen Gebietskörperschaft und ihrer Präferenzen; was in der Logik des Fiskalföderalismus ein klarer unerwünschter Nebeneffekt wäre.

2. Der Zusammenhang zwischen Schuldenstand und Leistungsvergleichen: Theoretische Überlegungen

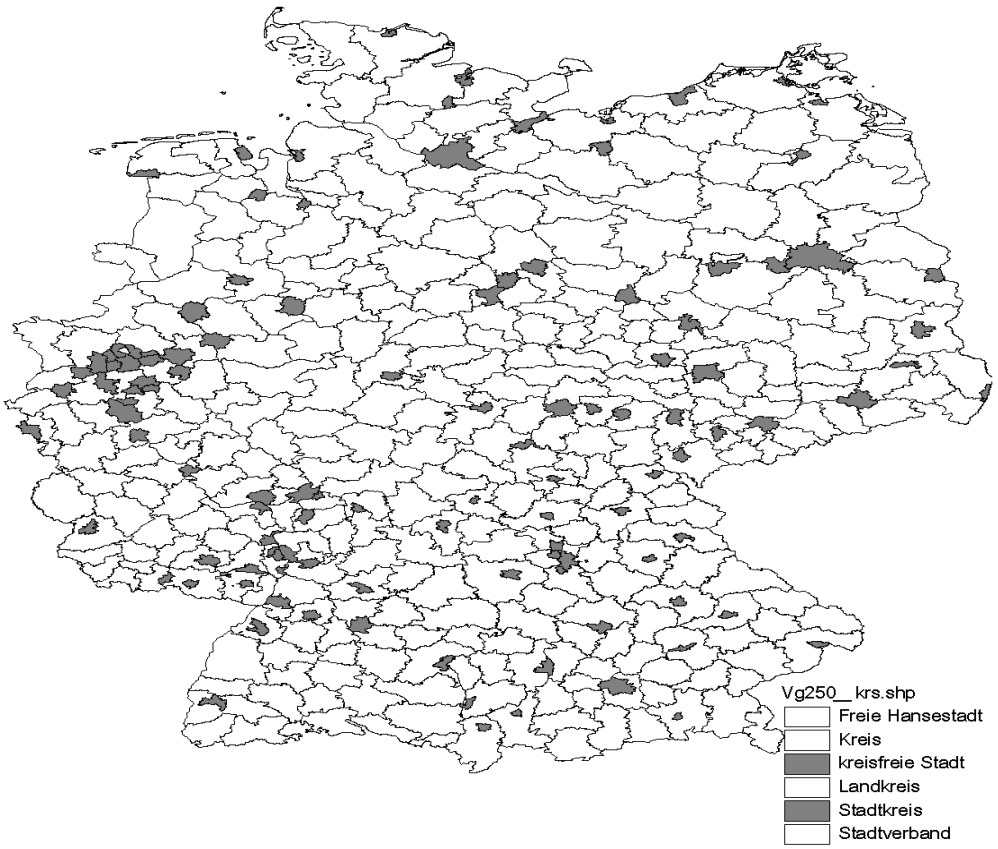
Der vorliegende Beitrag untersucht den Zusammenhang zwischen dem Verschuldungsniveau einer Kommune und der Bereitschaft zur Nutzung von Leistungsvergleichen. Weil es in Deutschland zahlreiche verschiedene Benchmarking-Projekte unterschiedlicher Dienstleister mit wechselnden Teilnehmern gibt, ist es nicht möglich, exakt anzugeben, welche der mehr als 11 000 Kommunen in Deutschland Leistungsvergleiche durchführt bzw. an ihnen teilnimmt. Die Nutzungshäufigkeit wird daher über Umfragedaten aus einer Studie von Bogumil et al. aus dem Jahr 2005 (*Bogumil* et al. 2008, *Kuhlmann/Bogumil/Grohs* 2008) abgebildet[4]. Betrachtet werden Leistungsvergleiche in der Gruppe der kreisfreien Städte (nicht kreisangehörige Städte und Gemeinden sowie Landkreise). Das Sample umfasst 73 der insgesamt 115 kreisfreien Städte, die es im Jahr 2005 gab (Abbildung 1)[5].

---

4 In der Umfrage wurden mehr als 800 Kommunen schriftlich befragt, welche Verwaltungsreformen sie seit Mitte der 1990er Jahre umgesetzt haben. Dabei wurde auch abgefragt, ob und wenn ja wie häufig Kommunen ihre Leistung mit anderen Kommunen vergleichen.

5 Die mittlere Einwohnerzahl in der Stichprobe ($\bar{x}$=191 504) ist kleiner als in der Grundgesamtheit (μ0=223 335), die Einwohnerzahlen (Standardabweichung $\sqrt{s^2}$ = 177 320) streuen weniger stark als in der Grundgesamtheit ($\sqrt{\sigma^2}$ = 380 503). Der Median im Sample ($\tilde{x}_0$=121 314) bzw. in der Grundgesamtheit ($\tilde{x}$ = 119 356) unterscheiden sich kaum. Getestet wird die Nullhypothese H0: μ = μ0 = 223‘335. Standardabweichung σ bzw. Varianz σ^2 des Merkmals Einwohnerzahl in der Grundgesamtheit sind bekannt, der Standardfehler des Mittelwertes kann berechnet werden: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$=44 534 (Wenn der Mittelwert des Merkmals Einwohnerzahl in der Grundgesamtheit 223 335 und der Standardfehler 44 534 betragen, wie wahrscheinlich ist dann ein Mittelwert des Merkmals von 191 504 in der Stichprobe?). Die Verteilung der Prüfgröße (z-Werte) folgt für H0 der Standardnormalverteilung. Für die Prüfgröße $z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} =$ -0,7148 ergibt sich ein p-Wert von p=0,4778 bzw. eine Irrtumswahrscheinlichkeit von 47%, die Nullhypothese kann nicht abgelehnt werden. Das spricht dafür, dass größere oder kleinere Städte im Sample im Vergleich zu der Gesamtheit der 115 Städte nicht über- oder unterrepräsentiert sind (was die Übertragbarkeit der empirischen Ergebnisse auf alle kreisfreien Städte einschränken würde).

Abbildung 1: Kreisfreie Städte im Jahr 2005



Anmerkung: Datenquelle: Bundesamt für Kartographie und Geodäsie.

Tabelle 1: Kreisfreie Städte: Teilnahme an Leistungsvergleichen

Teilnahme an Leistungsvergleichen[6]	Anzahl	Häufigkeit in %	Kumulierte Häufigkeiten
Regelmäßig	20	27,4%	27,4%
Gelegentlich	38	52,1%	79,5
Selten/Nein/Teilnahme in der Vergangenheit	15	20,55%	100,0
Summe	73	100%	-

Eigene Darstellung. Datenquelle: *Bogumil* et al. (2008).

6 Frage: „Beteiligt sich Ihre Kommune an interkommunalen Leistungsvergleichen, Vergleichsringen oder Wettbewerben?“.

Leistungsvergleiche sind fester Bestandteil der städtischen Verwaltung: 3 von 4 Städten (79,5%) vergleichen ihre Leistung zumindest gelegentlich mit anderen Kommunen, eine von vier Städten tut dies sogar regelmäßig (27,4%). Nur jede fünfte Stadt stellt sich selten oder gar nicht dem Vergleich (Tabelle 1).

In diesem Beitrag gilt das Interesse den Gründen für die unterschiedlich starke Nutzung von Leistungsvergleichen und insbesondere der Frage, ob es einen statistischen Zusammenhang gibt zwischen der Verschuldung einer Stadt und dem Ausmaß der Nutzung von Leistungsvergleichen gibt.

Um den Zusammenhang zwischen individueller Nutzungshäufigkeit und Leistung zu modellieren, wird ein klassisches Kosten-Nutzen-Modell verwendet. Die Teilnahme an einem Leistungsvergleich bringt für eine Kommune einen gewissen Nutzen: Erfahrungsaustausch von Praxiswissen, Identifikation von Optimierungsmöglichkeiten und Risikobereiche aus den Vergleichsdaten sowie Lerneffekte, die wiederum Kostentransparenz – und/oder Kosteneffizienz und -effektivität erhöhen (ökonomischer Nutzen). Das steigert wiederum die demokratische Akzeptanz der Verwaltung und der politischen Entscheidungsträger (politischer Nutzen).

Dem Nutzen stehen verschiedene Kosten gegenüber: direkte Transaktionskosten (Teilnahmegebühren), Opportunitätskosten für Datenerhebungen, Arbeitssitzungen etc. sowie politische Kosten. Die politischen Kosten bestehen in dem Risiko, in dem Vergleich schlecht abzuschneiden; d.h. nicht zur Mittel- oder zur Spitzengruppe zu gehören. So haben James und John (*James/John* 2007) für die englischen Kommunen gezeigt, dass sich eine schlechte Benotung einer Kommune in den alljährlichen Kommunalinspektionen negativ auf die Wiederwahlwahrscheinlichkeit der amtierenden Kommunalregierung auswirkt. Diese wird für eine schlechte Note an der Urne abgestraft, die Wiederwahlchance sinkt. Einen umgekehrten Belohnungseffekt für ein gutes Abschneiden gibt es jedoch nicht. Dieser Befund, wenn auch aus einem anderen kommunalen Kontext, deutet darauf hin, dass aus Vergleichsergebnissen konkrete politische Risiken resultieren können. Ich übertrage dieses Argument auf den deutschen Kontext.

Anstelle der tatsächlichen Kosten und Nutzen wird dabei die individuelle angegebene Nutzungsintensität beobachtet (Regelmäßig, gelegentlich oder selten/keine Nutzung/in der Vergangenheit). Das Leistungsniveau einer Kommune lässt sich nicht wie in England nähe-

rungsweise über standardisierte Leistungskennzahlen bzw. Scores aus verpflichtenden Inspektionen abbilden. Ich verwende den Schuldenstand je Einwohner (Mittelwert der Jahre 2000 bis 2004) als Annäherung an das mittelfristige Leistungsniveau einer Kommune. Diese einfache Kennzahl bietet gegenüber konstruierten/zusammengesetzten Leistungskennzahlen verschiedene Vorteile:

– es ist eine über Bundesländer hinweg standardisierte Finanzkennzahl, Werte für Kommunen aus verschiedenen Bundesländern sind vergleichbar,

– die Kennzahl und ihre Interpretation (der Unterschied zwischen hohen und niedrigen Werten) sind auch in der breiten Öffentlichkeit geläufig. Bei vielen Zeitungsberichten über den Problemdruck von NRW-Kommunen sind die Schulden je Einwohner in den großen Städten aufgelistet worden. Wähler sind daher sehr wahrscheinlich sensibilisiert für das Thema Schulden.

Im Zeitraum 2000 bis 2004 gab es ganz erhebliche Unterschiede im Verschuldungsniveau der kreisfreien Städte. Die tiefste Verschuldung lag bei rund 61 Euro pro Kopf (im Zeitraum 2000 bis 2004), die höchste Pro-Kopf-Verschuldung betrug rund 2 800 Euro[7] (Abbildung 2).

Theoretisch sind nun zwei Zusammenhänge denkbar zwischen der Nutzungsintensität und dem Verschuldungsniveau, das auch als fiskalischer Problemdruck interpretiert werden könnte:

1. Das Verschuldungsniveau hat einen negativen Effekt auf die Nutzungsintensität von interkommunalen Leistungsvergleichen: Städte mit Schwachstellen in der Leistungserbringung nutzen Leistungsvergleiche besonders intensiv, um Ursachen für Leistungsdefizite zu identifizieren und sich über Verbesserungsmöglichkeiten auszutauschen. Eine geringe Pro-Kopf-Verschuldung impliziert dagegen geringen Verbesserungsbedarf und damit geringe Nutzungsintensität.

---

7 Lage- und Streuungsparameter Schulden pro Kopf in Euro (2000 bis 2004): $\bar{x}$ = 1 338, Standardabweichung 539 Euro. Um Verzerrungen durch Ausreißer zu nivellieren, verwende ich in den Schätzungen die logarithmierten Pro-Kopf-Werte (2000 bis 2004). Eine Alternative zum fünf-jährigen Mittel wäre das zehn-jährige Mittel 1995 bis 2004. Einige Kommunen haben ihren Schuldenstand seit dem Jahr 2000 gesenkt, d.h. ihr Leistungsniveau verbessert. Insgesamt sind die Merkmalsrealisationen jedoch sehr stark korreliert (rp = 0,93).

2. Das Verschuldungsniveau hat einen positiven Effekt auf die Nutzungsintensität: Mit guten Vergleichsergebnissen lassen sich politische Gewinne (gegenüber den Wählern) und/oder Reputationsgewinne (gegenüber der Fachcommunity, wenn die Vergleichsergebnisse nicht veröffentlicht werden) realisieren. Umgekehrt besteht für Städte mit vergleichsweise hohem Schuldenstand das Risiko, durch einen Vergleich in der Öffentlichkeit an Reputation zu verlieren (oder innerhalb der lokalen Community, wenn die Vergleichsergebnisse nicht veröffentlicht werden). Diese politischen Kosten überwiegen möglicherweise den potentiellen Nutzen durch Lern- und Austauscheffekte.

3. Ein ökonometrisches Modell

Die Zielvariable „Nutzungshäufigkeit von Leistungsvergleichen“ hat drei Kategorien mit einer inhaltlich logischen Rangfolge (Nutzung: Regelmäßig > gelegentlich > selten/nein/in der Vergangenheit). Das Ausmaß der Nutzung ist eine Funktion der Pro-Kopf-Verschuldung sowie einer Reihe weiterer Kontrollvariablen. Spezifiziert wird ein *ordered logit*-Modell (Long/Freese 2006, S. 183-222). *Ordered logit*-Modelle gehen von einer latenten, unbeobachteten Variable (y*) aus, mit Werten zwischen $-\infty$ und $+\infty$ [8]:

$$y_i^* = \mathbf{x}_i'\beta + \epsilon_i$$

Die latente Variable (y*) wird durch J-1 unbeobachtete Schwellenwerte in die J Kategorien der beobachteten Zielvariable geteilt:

$$y_i = \tau_{m-1} \leq y_i^* < \tau_m$$

mit m=1,..., 3.

τ_1 bis τ_{J-1}, $\tau 0 = -\infty$, $\tau_J = \infty$

---

8 Man könnte sich die latente Variable y* z. B. als den erhofften Gewinn aus einem Leistungsvergleiche vorstellen, wobei Gewinn gleich Nutzen (durch Lerneffekte, Kosteneinsparungen etc., Wählerstimmenmaximierung) abzüglich Kosten (Transaktions- und Opportunitätskosten, politische Kosten). Nutzen und Kosten für eine Gemeinde sind unbekannt.

Statt der latenten Variable (y*) wird die Variable Nutzungshäufigkeit mit J ordinalen Kategorien beobachtet. Das Model besagt, dass die Wahrscheinlichkeit (Pr), dass y für gegebene Merkmalsrealisationen der erklärenden Variablen (**x**) den Wert m annimmt (Pr[y_i = m | **x**]), der Wahrscheinlichkeit entspricht, mit der die latente, unbeobachtete Variable zwischen die zwei entsprechenden Schwellenwerte (τ_{m-1}, τ_m) fällt.

$$Pr(y = m|\mathbf{x}) = Pr(\tau_{m-1} \leq y^* < \tau_m|\mathbf{x})$$

Y* wird durch **x'ß** + ε ersetzt, daraus ergibt sich die Standardformel für die vorhergesagten Wahrscheinlichkeiten im ordinalen logistischen Modell:

$$Pr(y = m|\mathbf{x}) = F(\tau_1 - \mathbf{x}\beta) - F(\tau_{m-1} - \mathbf{x}\beta)$$

F steht dabei für die kumulative Verteilungsfunktion, bei ordered logit ist das die logistische Verteilungsfunktion (mit einem Erwartungswert von 0 und einer Varianz $\sigma^2 = \pi^2/3$). Geschätzt wird ein Set von zu den erklärenden Variablen korrespondieren Parametern (**ß**) sowie die Schwellenwerte (τ 's), die Anzahl der Schwellenwerte ist dabei die Anzahl der möglichen Ausprägungen von Y minus 1.

Entsprechend ergeben sich die Wahrscheinlichkeiten für y = 1, y = 2 und y = 3:

$$\begin{aligned}
Pr(y = 1|\mathbf{x}) &= F(\tau_1 - \mathbf{x}\beta) - F(\tau_0 - \mathbf{x}\beta) \\
&= F(\tau_1 - \mathbf{x}\beta) - F(-\infty - \mathbf{x}\beta) \\
&= F(\tau_1 - \mathbf{x}\beta) - F(0 - \mathbf{x}\beta) \\
&= F(\tau_1 - \mathbf{x}\beta) \\
Pr(y = 2|\mathbf{x}) &= F(\tau_2 - \mathbf{x}\beta) - F(\tau_1 - \mathbf{x}\beta) \\
Pr(y = 3|\mathbf{x}) &= F(\tau_3 - \mathbf{x}\beta) - F(\tau_2 - \mathbf{x}\beta) \\
&= F(\infty - \mathbf{x}\beta) - F(\tau_2 - \mathbf{x}\beta) \\
&= 1 - F(\tau_2 - \mathbf{x}\beta)
\end{aligned}$$

Die Wahrscheinlichkeit für y = 1 (Pr[y = 1 | **x**]) berechnet sich wie folgt, (für y = 2 entsprechend mit τ_2) (*Long/Freese* 2006, S. 183-222):

$$F(\tau_1 - \mathbf{x}'\beta) = \frac{exp(\tau_1 - \mathbf{x}'\beta)}{1 + exp(\tau_1 - \mathbf{x}'\beta)}$$

Kontrolliert wird für den Einfluss einer Reihe weiterer Einflussfaktoren. Insgesamt werden sechs Modelle gefittet:

$$\mathbf{x}'\beta = \beta_{schlog00}\, schlog00 + \beta_{logbes}\, logbes + \beta_{spd}\, spd \quad (1)$$
$$\mathbf{x}'\beta = \beta_{schlog00}\, schlog00 + \beta_{logbes}\, logbes + \beta_{fdp}\, fdp \quad (2)$$
$$\mathbf{x}'\beta = \beta_{schlog00}\, schlog00 + \beta_{logbes}\, logbes + \beta_{spd}\, spd + \beta_{alq}\, alq \quad (3)$$
$$\mathbf{x}'\beta = \beta_{schlog00}\, schlog00 + \beta_{logbes}\, logbes + \beta_{fdp}\, fdp + \beta_{alq}\, alq \quad (4)$$
$$\mathbf{x}'\beta = \beta_{schlog00}\, schlog00 + \beta_{logbes}\, logbes + \beta_{soz}\, soz \quad (5)$$
$$\mathbf{x}'\beta = \beta_{schlog00}\, schlog00 + \beta_{logbes}\, logbes + \beta_{alq}\, alq \quad (6)$$

mit

teilnahme	Teilnahme an Leistungsvergleichen: 1 = in der Vergangenheit/nein/selten, 2 = gelegentlich, 3 = häufig
schlog00	Logarithmus Schulden pro Kopf Mittelwert 2000 bis 2004
alq	Arbeitslosenquote bezogen auf alle Erwerbspersonen in Prozent, Mittelwert der Werte für 2001-2004
soz	Empfänger(innen) von Sozialhilfe je 1 000 Einwohner, Mittelwert der Werte für 2001 bis 2004
logbes	Logarithmus Beschäftigte der Gemeinde (Vollzeit und Teilzeitbeschäftigte mit mehr als 50 % der regulären Wochenarbeitszeit), Mittelwert der Jahre 2000 bis 2004
spd	SPD-Oberbürgermeister: 1 = Ja, 0 = Nein
fdp	FDP-Oberbürgermeister: 1 = Ja, 0 = Nein

Die Arbeitslosenquote und die Zahl der Sozialhilfeempfänger pro Kopf (die Surveydaten stammen aus dem Jahr 2005, die Umstellung der Sozialhilfe fällt nicht in die Daten) sind jeweils zusätzliche Indikatoren für Problemdruck. Hohe Belastungen könnten entweder das Interesse erhöhen, durch Vergleiche Verbesserungen in diesen Bereichen zu er-

zielen. Hohe Belastungen könnten aber auch dazu führen, dass Städte resignieren, sich auf eine Opferrolle zurückziehen und statt aus eigener Kraft innovative Lösungen zu suchen, Hilfe von Bund und Land fordern, um hohe Arbeitslosigkeit zu bewältigen. Verwendet werden bei den Sozialhilfeempfängern pro 1 000 Einwohner-Werte, da größere Städte rein mengenmäßig i.d.R. auch mehr Sozialhilfeempfänger haben.

Für Städte mit großer Verwaltung ergeben sich durch Leistungsvergleiche zusätzliche Vorteile. Vergleichsinformationen aus anderen Städten reduzieren Informationsasymmetrien zu Gunsten der Führungsebene[9]. Je größer die Anzahl der Beschäftigten in einer Verwaltung umso eher entstehen Ineffizienzen durch Informationsasymmetrien. Umso größer ist damit der Anreiz für die Leitungsebene, sich an Vergleichen zu beteiligen, um diese Informationsasymmetrien zu reduzieren. Ein weiteres Argument für einen positiven Effekt der Mitarbeiterzahl ist, dass große Verwaltungen eher als kleine Kommunen über den nötigen personellen Spielraum für Vergleichsarbeiten, Datenerhebung und systematische Auswertung der Vergleichsergebnisse etc. verfügen[10]. Die Beschäftigtenzahlen variieren sehr stark, der Median liegt

---

9 Zwischen den Führungskräften und den Mitarbeitern der operativen Ebene besteht eine Prinzipal-Agenten-Beziehung, die Dezernats-, Abteilungsleiter etc. sind die Prinzipale und delegieren Aufgaben an die Agenten (middle level und die street-level bureacrats oder auch front-liner, weil sie an der vordersten Front zum Bürger stehen, z. B. als Sozialarbeiter oder als Sachbearbeiter im Bürgerbüro). Die Agenten verfügen über einen Informationsvorsprung gegenüber dem Prinzipal, d.h. der Dezernatsleiter kann nicht ohne weiteres beurteilen, ob Person A gut oder schlecht arbeitet. Diesen Informationsvorsprung könnten die Agenten zu ihren Gunsten ausnutzen, indem sie langsamer arbeiten als sie könnten, also ihre Produktivität künstlich reduzieren; sie realisieren dadurch ökonomische Renten (im Sinne des englischen Begriffs rent-seeking). Erhält die Führungsebene nun Vergleichsinformationen etwa über die Fälle pro Sachbearbeiter aus anderen Städten, kann sie die Leistung ihrer Mitarbeiter besser beurteilen. Durch die Vergleichsinformationen werden bestehende Informationsasymmetrien zu Gunsten der Prinzipale reduziert.

10 Vor 2006 liegen keine Vollzeitäquivalente vor, daher werden Vollzeitbeschäftigte (VZ) und Teilzeitbeschäftigte mit mehr als 50 Prozent der regelmäßigen Wochenarbeitszeit (TZ1) addiert. Nicht berücksichtigt werden Teilzeitbeschäftigte mit weniger als 50 Prozent der regelmäßigen Wochenarbeitszeit (TZ2). Vergleiche mit Daten nach 2006 zeigen, dass diese Summe etwa ein Drittel über den Vollzeitäquivalenten liegt. Da ich mit der Anzahl der Beschäftigten argumentieren und nicht mit dem Beschäftigungsumfang in VZÄ,

bei 3 200 Beschäftigten, in einigen großen Städten arbeiten aber auch mehr als 15 000 Beschäftigte. Um Verzerrungen durch Ausreißer zu vermeiden, verwende ich die logarithmierten Werte.

Um zu prüfen, inwieweit Unterschiede in der Nutzungshäufigkeit parteipolitisch motiviert sind, werden zwei jeweils dichotome Variablen verwendet. In den kreisfreien Städten gab es 2005 überwiegend SPD- und Unionsbürgermeister, dazu drei mit FDP-Parteibuch und drei Parteilose. Leistungsvergleiche kommen aus der Privatwirtschaft und sollen Wettbewerb simulieren. Oberbürgermeister der SPD haben daher möglicherweise eine geringere Bereitschaft als übrige Parteien, den Wettbewerbsmechanismus in die öffentliche Verwaltung zu importieren. Umgekehrt könnte man von FDP-Oberbürgermeistern eine höhere Bereitschaft erwarten.

Abbildung 2 zeigt die Lage- und Streuungsparameter der einzelnen Variablen. Es gibt keine Hinweise auf mögliche Multikollinearität durch stark korrelierte erklärende Variablen (

---

stellt diese systematisch Überschätzung des Beschäftigungsumfanges kein Problem für die inhaltliche Interpretation dar.

Abbildung 3).

Abbildung 2: Lage- und Streuungsparameter der Variablen

```
. tab teilnahme
```

teilnahme	Freq.	Percent	Cum.
1	15	20.55	20.55
2	38	52.05	72.60
3	20	27.40	100.00
Total	73	100.00	

```
. . sum  schlog00  alq soz besch logbes spd fdp
```

Variable	Obs	Mean	Std. Dev.	Min	Max
schlog00	72	3.078959	.239703	1.788623	3.448925
alq	73	11.56062	4.002445	6.15	21.375
soz	73	46.89833	18.3271	19.49239	121.1961
besch	72	4435.817	4139.203	748	22500.2
logbes	72	3.517423	.3212384	2.873902	4.352186
spd	73	.5205479	.5030349	0	1
fdp	73	.0410959	.1998858	0	1

Abbildung 3: Korrelationsmatrix der erklärenden Variablen

```
. pwcorr  schlog00 alq soz logbes
```

	schlog00	alq	soz	logbes
schlog00	1.0000			
alq	-0.0963	1.0000		
soz	0.1824	0.3199	1.0000	
logbes	0.3161	-0.1125	0.2948	1.0000

4. Darstellung und Diskussion der Schätzergebnisse

Abbildung 4 zeigt die Paramater der geschätzten Modelle (**ß's** und Schwellenwert, die sog. cuts).

Abbildung 4: Parameter der geschätzten Modelle

Variable	Eins	Zwei	Drei	Vier	Fünf	Sechs
teilnahme						
schlog00	-3.574**	-3.519**	-3.462**	-3.381**	-3.868**	-3.507**
logbes	0.565	0.719	0.660	0.793	0.473	0.670
spd	-0.061		-0.119			
fdp		1.837		1.690		
alq			0.119*	0.113*		0.118*
soz					0.012	
cut1						
_cons	-10.554**	-9.764**	-8.586*	-7.814*	-11.226**	-8.636*
cut2						
_cons	-7.996*	-7.154*	-5.940	-5.119	-8.655**	-5.992
Statistics						
N	72	72	72	72	72	72
ll	-68.844	-67.707	-67.030	-66.053	-68.555	-67.061
chi2	8.565	10.839	12.194	14.147	9.142	12.130
aic	147.688	145.414	146.059	144.106	147.111	144.123

legend: * p<.1; ** p<.05; *** p<.001

Ordered Logit Modelle (ORM) basieren auf der Annahme einer latenten Variable (y*, siehe oben). Es werden brant-Tests durchgeführt, um auf Verletzungen der Parallel Regression Assumption (PRA) zu prüfen[11].

11 Ein ORM ist äquivalent zu J-1 binären Regressionen unter der Annahme, dass die geschätzten Koeffizienten x über die einzelnen Gruppen identisch sind. Diese Annahme wird PRA genannt. Um zu prüfen, ob die PRA verletzt ist, muss man die Koeffizienten für J-1 binäre Regressionen schätzen, vergleichen und testen, ob die Koeffizienten signifikant voneinander abweichen. Ein brant-Test ist eine Möglichkeit, diese zu tun, ein approximativer LR Test eine andere (*Long/Freese* 2006, S. 197-200). Weichen die Koeffizienten

Drei der vier geprüften Modelle[12] absolvieren erfolgreich einen brant-Test auf Verletzung der PRA, auch wenn einzelne Variablen den Test nicht schaffen (Abbildung 12 bis Abbildung 15). Für Modell 5 muss die PRA verworfen werden (chi2 = 6,83, p = 0,078), Modell 1 schafft den Test nur knapp (chi2 = 6,13, p = 0,106). Problematisch sind dabei in den Modellen die beiden Parteienvariablen sowie die Variable Sozialhilfeempfänger pro Kopf. Die brant-Teststatistiken zeigen insbesondere an, dass Modell 5 nicht als orm-Modell spezifiziert bzw. gefittet werden kann, ohne die Annahmen auf Proportional Odds zu verletzen. Da die brant-Tests keine signifikante Verletzung der PRA durch die beiden Problemdruckvariablen Pro-Kopf-Verschuldung sowie Arbeitslosenquote anzeigen, wähle ich den Weg, nur die Modelle ohne Verletzungen der PRA weiter zu betrachten und die parteipolitischen Variablen sowie die Sozialhilfeempfänger-Variable nicht näher zu betrachten. Nicht weiter verfolgt wird die Alternative, die betroffenen Modelle als *generalized ordered logit models* zu spezifizieren und zu fitten, weil in erster Linie der statistische Effekt der Verschuldungssituation interessiert.

Im Folgenden wird nur das Modell Sechs weitergehend betrachtet. Getestet werden die Nullhypothesen $ß_{schlog00} = 0$, $ß_{logbes} = 0$ und $ß_{alq} = 0$. Wald-Tests implizieren, dass die log. Pro-Kopf Verschuldung (chi2 = 5,63, df = 1, p < 0,05) sowie die Arbeitslosenquote (chi2 = 3,44, df = 1, p < 0,05) die Vergleichshäufigkeit signifikant beeinflussen. Dagegen kann die Hypothese $ß_{logbes} = 0$ nicht verworfen werden (chi2 = 0,75, df = 1). LR-Tests liefern inhaltlich identische Ergebnisse[13].

---

der J-1 binären Regressionen signifikant voneinander ab, wird die PRA verletzt. Das Model muss als generalisiertes Modell oder ggf. als multinominales Modell spezifiziert werden; jedenfalls als Modell, in dem die Annahme der Proportional odds gelockert ist. Teststatistik ist die chi2-Verteilung. Eine signifikante Abweichung für eine Variable bedeutet eine Verletzung der PRA. Brant-Tests für die Modelle Eins bis Sechs zeigen, dass die PRA für die Parteien-Variablen spd sowie für soz (die Zahl der Sozialhilfeempfänger pro Kopf) abgelehnt werden muss.

12 Für die Modelle mit fdp-Dummy kann der brant-Test nicht durchgeführt werden, da nicht in jeder Antwortkategorie y = m, mit m = 1, 2, 3, eine Kommune mit FDP-Oberbürgermeister vertreten ist.

13 Um die einzelnen Koeffizienten zu testen, werden dabei die log likelihood des vollen Modells (hier Modell 6, x'ß = ßschlog00*schlog00 + ßlogbes*logbes + ßalq * alq) jeweils verglichen mit Modellen, in denen die interessierende erklärende Variable ausgeschlossen ist. Schlog00 ist demnach

Eine hohe Pro-Kopf Verschuldung wirkt sich negativ auf die Nutzung von Leistungsvergleichen aus. Eine hohe Arbeitslosenquote ist dagegen mit einer stärkeren Nutzung von Vergleichen assoziiert. Der Personalumfang der Verwaltung wirkt sich nicht signifikant auf die Vergleichsaktivität aus. Das negative Vorzeichen des Koeffizienten entspricht dabei zumindest den theoretischen Erwartungen.

Die inhaltliche Interpretation der marginalen Effekte ist etwas schwieriger, weil die Pro-Kopf-Verschuldung und die Beschäftigtenzahlen als logarithmierte Werte in die Modelle eingegangen sind (

signifikant auf 0.05-Niveau (LRchi2 = 7,81, df = 1, p < 0,05), alq auf 0.1-Niveau (LRchi2 = 3,58, df = 1, p < 0,1); logbes hingegen nicht (LRchi2 = 0,75, df = 1).

Abbildung 5): Eine Erhöhung der Variable log. Schuldenstand pro Kopf (schlog00) um eine Standardabweichung (*SDofx* = 0,2397, entspricht $10^{0,2397}$ = 1,7366 Euro pro Kopf) verändert die Odds von Wahrscheinlichkeit einer häufigen Nutzung (y = 3) zu Gegenwahrscheinlichkeit (gelegentliche oder seltene/keine Nutzung, y < 3) um den Faktor exp(bStdX) = exp($ß_{schlog00}$ * $Std._{schlog00}$)= exp(-3,50675 * 0,2397) = 0,4315, (d.h. sie sinkt, da der Faktor < 1), sofern die log. Beschäftigtenzahl (logbes) und die Arbeitslosenquote (alq) unverändert bleiben. Mit anderen Worten, die Wahrscheinlichkeit einer häufigen Nutzung sinkt mit wachsender Pro-Kopf-Verschuldung (

Abbildung 5).

Abbildung 5: Marginale Veränderungen in der latenten Variable y* und Effektkoeffizienten

```
. listcoef, std help

ologit (N=72): Unstandardized and Standardized Estimates

 Observed SD: .68974151
   Latent SD: 2.0493284
```

teilnahme	b	z	P>|z|	bStdX	bStdY	bStdXY	SDofX
schlog00	-3.50675	-2.372	0.018	-0.8406	-1.7112	-0.4102	0.2397
logbes	0.66989	0.864	0.388	0.2152	0.3269	0.1050	0.3212
alq	0.11770	1.854	0.064	0.4672	0.0574	0.2280	3.9697

```
      b = raw coefficient
      z = z-score for test of b=0
  P>|z| = p-value for z-test
  bStdX = x-standardized coefficient
  bStdY = y-standardized coefficient
 bStdXY = fully standardized coefficient
  SDofX = standard deviation of X

. listcoef, help

ologit (N=72): Factor Change in Odds

  Odds of: >m vs <=m
```

teilnahme	b	z	P>|z|	e^b	e^bStdX	SDofX
schlog00	-3.50675	-2.372	0.018	0.0300	0.4315	0.2397
logbes	0.66989	0.864	0.388	1.9540	1.2401	0.3212
alq	0.11770	1.854	0.064	1.1249	1.5956	3.9697

```
      b = raw coefficient
      z = z-score for test of b=0
  P>|z| = p-value for z-test
    e^b = exp(b) = factor change in odds for unit increase in X
e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
  SDofX = standard deviation of X
```

Aufschlussreicher als die marginalen Effekte sind die vorhergesagten Eintrittswahrscheinlichkeiten für die Ergebnisse y = 1, y = 2 und y = 3; sie berechnen sich mit den im Modell Sechs geschätzten Parametern wie folgt:

$$F(\tau_1 - \mathbf{x}'\beta) = \frac{exp(\tau_1 - \mathbf{x}'\beta)}{1 + exp(\tau_1 - \mathbf{x}'\beta)}$$

Ich berechne hier exemplarisch die Werte für die Beobachtung Nr. 1, mit folgenden drei Merkmalsrealisationen für die erklärenden Variablen:

schlog00 = 3,332 (entspricht Pro-Kopf-Verschuldung von $10^{3,332}$ = 2 147 Euro)

logbes = 3,689 (entspr. abs. Mitarbeiterzahl von $10^{3,689}$ = 4 891)

alq = 10,65 (% der Erwerbsbevölkerung).

τ_1 = -8,636

τ_2 = -5,992

Pr (y = 1 | **x**) = exp(τ_1 -**x'ß**) / 1 + exp(τ_1 -**x'ß**) = exp(τ_1 – ($ß_{schlog00}$ * schlog00 + $ß_{logbes}$* logbes + $ß_{alq}$ * alq)) / 1 + exp(τ_1 – $ß_{schlog00}$ * schlog00 + $ß_{logbes}$* logbes + $ß_{alq}$ * alq) = exp[-8,636 – (-7,8733)] / 1 + exp[-8,636 – (-7,8733)] = exp(-0,7627) / 1 + exp(-0,7627) = 0,46 / 1,46 = 0,318

Pr (y = 2 | **x**) = F(τ_2 – **x'ß)** – F(τ_1 – **x'ß**) = exp[-5,992 – (-7,8733)] / 1 + exp[-5,992 – (-7,8733)] – 0,318 = 0,8676 – 0,318 = 0,5498

Pr(y =3 | **x**) = 1 – [Pr(y = 1) + Pr(y=2)] = 1 – [0,318 + 0,5498] = 0,1322.

Die Berechnung mittels des Stata-Post-Regression Kommandos predict ergibt die Wahrscheinlichkeiten 0,34, 0,54 und 0.12 (Abbildung 6).

Abbildung 6: Vorhergesagte Wahrscheinlichkeiten für Beobachtung Nr. 1

	selten	gelegent	häufig
1.	.3371297	.5402841	.1225863

Die durch das Modell vorhergesagten Wahrscheinlichkeiten werden entsprechend für alle Beobachtungen berechnet. In Abbildung 7 sind die Häufigkeiten dieser vorhergesagten Wahrscheinlichkeiten für die einzelnen Kategorien y = 1, y = 2, und y = 3 abgetragen. Abbildung 7 zeigt, dass die vorhergesagte Wahrscheinlichkeiten für die Ergebnisse seltene/keine Nutzung sowie häufige Nutzung nur sehr selten Werte größer als 0,5 annehmen, d. h. diese Ereignisse werden nur sehr selten als am wahrscheinlichsten vorhergesagt. Deutlich am häufigsten wird die „Mittelkategorie" einer gelegentlichen Nutzung von Leistungsvergleichen vorhergesagt. Das Ereignis y = 2 ist im Sample zwar am häufigsten vertreten (52,1%; siehe Tabelle 1), aber es gibt auch eine

bedeutsamen Anteil sowohl an häufiger als auch an seltener/keiner Nutzung, der auf jeden Fall größer ist als die vorhergesagten Werte in Abbildung 7. Die Erklärungskraft des Gesamtmodells mit Blick auf die „Extremkategorien“ ist beschränkt. Der Vergleich der Gruppe der Fälle mit einer häufigen Nutzung mit den vorhergesagten Wahrscheinlichkeiten macht ebenfalls deutlich, dass zahlreiche Fälle fehlklassifiziert werden.

Abbildung 7: Übersicht Predicted Probabilities für y = 1, y = 2, y = 3

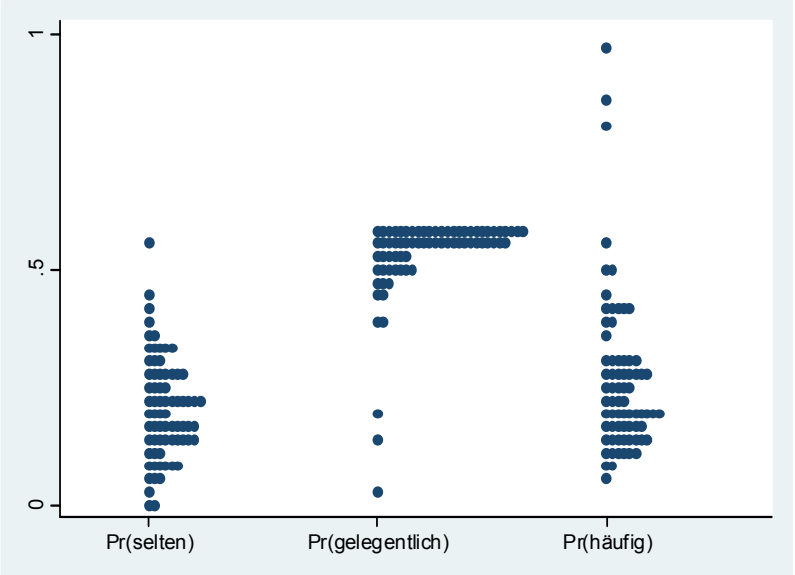


Abbildung 8: Predicted probabilities für Beobachtungen y = 3 (häufige Nutzung)

	teilna~e	selten	gelegent	häufig
5.	3	.3346246	.5415761	.1237993
11.	3	.0174398	.1824266	.8001336
12.	3	.1433909	.558615	.297994
13.	3	.	.	.
15.	3	.143699	.5588309	.2974701
20.	3	.0110685	.1250103	.8639212
25.	3	.1455718	.560112	.2943162
26.	3	.2078567	.5790504	.2130928
32.	3	.1581267	.5673997	.2744735
47.	3	.1366113	.5534809	.3099078
53.	3	.2273661	.578134	.1944999
54.	3	.0951104	.5015336	.4033559
56.	3	.1324384	.5499364	.3176253
57.	3	.0872749	.4864105	.4263146
60.	3	.170708	.5726809	.2566111
64.	3	.190426	.5775702	.2320038
65.	3	.2324972	.577503	.1899997
67.	3	.1743625	.5738791	.2517583
69.	3	.0020122	.0255798	.9724081
71.	3	.1486977	.5621333	.289169

Von den 20 Fällen mit einer häufigen Nutzung werden durch das Modell nur drei (Fall 11, 20 und 69, entspricht 15 %) korrekt vorhergesagt/klassifiziert (siehe Abbildung 8). Für die übrigen wird ein Ereignis y = 2 vorhergesagt, eine krasse Fehlspezifikation mit der höchsten Eintrittswahrscheinlichkeit für y = 1 tritt zumindest nicht auf.

Von den 15 Fällen mit einer seltenen/keinen Teilnahme wird auf der Basis von Modell Sechs nur für einen einzigen Fall eine Wahrscheinlichkeit Pr(y = 1) größer 0,5 vorhergesagt (entspricht 6,7 %, Abbildung 9). Von den 38 Fällen mit y = 2 (gelegentliche Nutzung) werden 35 Fälle korrekt (in dem Sinne, dass die vorhergesagte Wahrscheinlichkeit ihr Maximum für Pr[y = 2] annimmt) vorhergesagt (Abbildung 10). Insgesamt werden damit 54,2 % der Fälle korrekt klassifiziert.

Abbildung 9: Predicted probabilities für Beobachtungen y = 1 (selten Nutzung/keine Nutzung/Nutzung in der Vergangenheit)

	teilna~e	selten	gelegent	häufig
1.	1	.3371297	.5402841	.1225863
2.	1	.2833789	.5643003	.1523209
6.	1	.0918616	.4955225	.4126159
19.	1	.2247886	.5783931	.1968182
21.	1	.4395155	.4773993	.0830853
28.	1	.1747203	.5739889	.2512908
34.	1	.1509546	.5635048	.2855406
44.	1	.0990842	.5084201	.3924958
45.	1	.2382631	.5766189	.185118
46.	1	.3793698	.5164905	.1041397
58.	1	.3322209	.5428019	.1249772
62.	1	.2296208	.5778753	.1925039
63.	1	.5561496	.3901849	.0536655
72.	1	.3391616	.5392253	.121613
73.	1	.0788512	.4675689	.4535798

Abbildung 10: Predicted probabilities für Beobachtungen y = 2 (gelegentliche Nutzung)

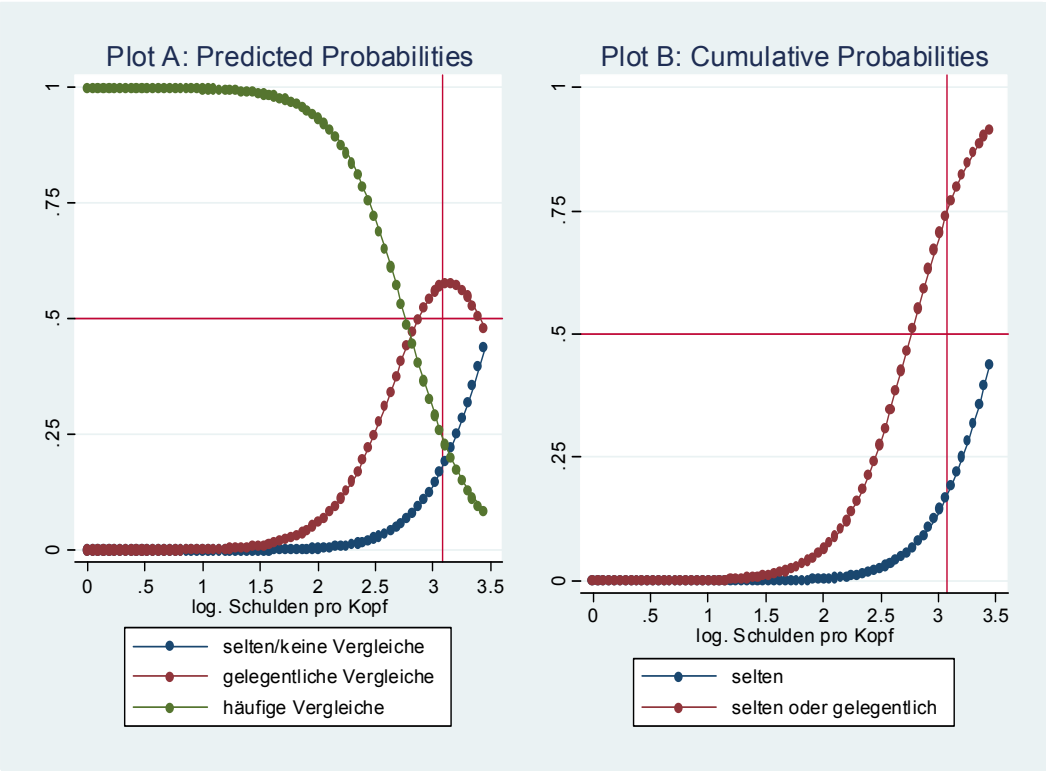
	teilna~e	selten	gelegent	häufig
3.	2	.2313438	.5776579	.1909983
4.	2	.165899	.5708812	.2632198
7.	2	.2892017	.5621208	.1486775
8.	2	.1172715	.5342536	.3484749
9.	2	.2753316	.5671139	.1575545
10.	2	.1634246	.5698519	.2667235
14.	2	.2112162	.5790724	.2097114
16.	2	.2930869	.5606027	.1463104
17.	2	.1594002	.5680209	.2725789
18.	2	.2700724	.5688214	.1611062
22.	2	.2696319	.5689595	.1614086
23.	2	.0936479	.4988714	.4074807
24.	2	.0661924	.4332003	.5006073
27.	2	.0865908	.4849854	.4284238
29.	2	.327166	.5453338	.1275002
30.	2	.2951029	.5597954	.1451017
31.	2	.1834119	.5762588	.2403293
33.	2	.1523539	.5643192	.2833268
35.	2	.409247	.4977238	.0930292
36.	2	.2909506	.5614436	.1476058
37.	2	.0971381	.5051097	.3977522
38.	2	.1584837	.567576	.2739403
39.	2	.2770103	.5665466	.1564431
40.	2	.3537956	.5313289	.1148755
41.	2	.0525855	.3859714	.5614431
42.	2	.218594	.5788503	.2025557
43.	2	.2156931	.5789807	.2053263
48.	2	.2465244	.5750484	.1784272
49.	2	.22112	.5786927	.2001873
50.	2	.2591877	.5720018	.1688105
51.	2	.295458	.5596519	.1448901
52.	2	.1441123	.5591183	.2967694
55.	2	.190473	.5775775	.2319495
59.	2	.0675823	.4373776	.4950401
61.	2	.3546426	.5308578	.1144996
66.	2	.2189225	.578832	.2022454
68.	2	.2513217	.5739816	.1746966
70.	2	.1650421	.5705329	.264425

Im Folgenden wird gezeigt, wie sich eine zunehmende Pro-Kopf-Verschuldung auf die Nutzungsintensität von Leistungsvergleichen auswirkt (Abbildung 11). Dazu werden die Eintrittswahrscheinlichkeiten für alle drei Ereignisse y = 1, y = 2 und y = 3, also eine seltene/keine, gelegentliche sowie häufige Nutzung von Leistungsvergleichen, für verschiedene Pro-Kopf-Verschuldungswerte vorhergesagt. Nur der Wert der Variable schlog00 wird variiert, vom Maximalwert (schlog00 = 2,4489, entspricht $10^{3,4489}$ = 2 811 Euro pro Kopf) im Sample, über den Minimalwert hinaus bis auf einen fiktiven Wert von 0 [14]; die Werte für logbes und alq werden konstant gehalten und zwar jeweils beim Mittelwert im Sample ($\overline{alq}$ = 11,56, $\overline{logbes}$ = 3,517, siehe Abbildung 2).

Demnach würde jede Kommune mit einer Pro-Kopf-Verschuldung unter 180 Euro Leistungsvergleiche mit sehr hoher Wahrscheinlichkeit häufig durchführen. Ab einer Pro-Kopf-Verschuldung von rund 560 Euro (entspricht einem Wert schlog00 = 2,75) sinkt die Wahrscheinlichkeit für eine intensive Nutzung von Leistungsvergleichen unter 0,5, und eine nur noch gelegentliche Nutzung wird wahrscheinlicher. Gleichzeitig steigt auch die Wahrscheinlichkeit rapide an, dass das Instrument gar nicht genutzt wird (Abbildung 11).

14 Im Jahr 2005 gab es im Sample keine schuldenfreien kreisfreien Städte. Es gibt jedoch eine Reihe von Kommunen in Deutschland ohne Schulden, so dass es sich um einen realistischen Zustand handelt.

Abbildung 11: Predicted und Cumulative Probabilities für zunehmende Pro-Kopf-Verschuldung



Im Jahr 2008, das letzte, für das bundesweite Schuldendaten vorliegen, lag die durchschnittliche Pro-Kopf Verschuldung der kreisfreien Städte bei 1 139 Euro, und damit nur geringfügig unter dem Mittelwert der Jahre 2000 bis 2004. Bei gleichbleibenden makroökonomischen Rahmenbedingungen (Arbeitslosenquote) dürften sich die Nutzung von Vergleichen damit leicht erhöht haben.

5. Schlussfolgerungen

Die empirischen Ergebnisse zeigen, dass Kommunen mit hoher Pro-Kopf-Verschuldung Leistungsvergleiche zurückhaltender nutzen als weniger verschuldete Kommunen. Eine steigende Schuldenlast wirkt sich signifikant negativ auf die Wahrscheinlichkeit aus, dass Leistungsvergleiche häufig genutzt werden. Ich interpretiere diesen Befund wie folgt: Relativ (bezogen auf die Gesamtheit der kreisfreien Städte) hoch verschuldete Städte scheuen die potentiellen politischen Kosten (in Form von Wiederwahlrisiken und/oder Reputationsverlust), die

durch ein schlechtes Abschneiden bei einem Leistungsvergleich entstehen können. Diese potentiellen Kosten, zusammen mit Opportunitäts- und Transaktionskosten, übersteigen für diese Städte den potentiellen Nutzen, den eine vergleichende Standortbestimmung und Erfahrungsaustausch mit sich bringen können. Lerneffekte können nicht stattfinden, mögliche Wohlfahrtssteigerungen werden nicht genutzt.

Umgekehrt erhöht eine hohe Arbeitslosenquote die Wahrscheinlichkeit einer intensiven Nutzung von Leistungsvergleichen. Allerdings unterliegt die Arbeitslosenquote sehr viel weniger als die Pro-Kopf-Verschuldung den Möglichkeiten einer direkten Einflussnahme durch die politischen und administrativen Entscheidungsträger.

Mit Blick auf die angespannte Finanzlage vieler kreisfreien Städte ergibt sich daraus folgende Schlussfolgerung: Zahlreiche Gründe sprechen für Entschuldungspakte von Land und seinen Kommunen zur Reduktion der Schuldenlast, insbesondere der Kassenkredite. Eine Senkung der Pro-Kopf-Verschuldung würde nach den vorliegenden empirischen Ergebnissen auch den Spielraum für die Nutzung moderner Managementinstrumente, konkret von vergleichender Kennzahlenarbeit und Leistungsvergleichen erhöhen. Denn sie würde die potentiellen politischen Kosten eines schlechten Abschneidens reduzieren.

Literatur

Ammons, D.N. and W.C. Rivenbark 2008. 'Factors Influencing the Use of Performance Data to Improve Municipal Services: Evidence from the North Carolina Benchmarking Project', Public Administration Review, 68, 304-318. doi: 10.1111/j.1540-6210.2007.00864.x.

Behn, R.D. 2003. 'Why Measure Performance? Different Purposes Require Different Measures', Public Administration Review, 63, 586-606. doi: 10.1111/1540-6210.00322.

Bogumil, J., S. Grohs, S. Kuhlmann und A.K. Ohm 2008. Zehn Jahre neues Steuerungsmodell: eine Bilanz kommunaler Verwaltungsmodernisierung. 2 edn. Berlin: Edition Sigma.

Kuhlmann, S., J. Bogumil und S. Grohs 2008. 'Evaluating Administrative Modernization in German Local Governments: Success or Failure of the "New Steering Model"?' Public Administration Review, 68, 851-863. doi: 10.1111/j.1540-6210.2008.00927.x.

Dalehite, E.G. 2008. 'Determinants of Performance Measurement: An Investigation into the Decision to Conduct Citizen Surveys', Public Administration Review, 68, 891-907. doi: 10.1111/j.1540-6210.2008.00930.x.

de Bruijn, H. 2002. Managing performance in the public sector. London ; New York: Routledge.

Jäkel, T. und S. Kuhlmann 2012. 'Kann man von der Schweiz lernen? Leistungsvergleiche kommunaler Finanzkennzahlen in föderalen Ländern', Verwaltung & Management, 18, 131-135.

James, O. und P. John 2007. 'Public Management at the Ballot Box: Performance Information and Electoral Support for Incumbent English Local Governments', Journal of Public Administration Research and Theory, 17, 567-580. doi: 10.1093/jopart/mul020.

Julnes, P.d.L. und M. Holzer 2001. 'Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation', Public Administration Review, 61, 693-708. doi: 10.1111/0033-3352.00140.

Meyer, M.W. und V. Gupta 1994. 'The Performance Paradox', Research in Organizational Behavior, 16, 309-369.

Kuhlmann, S. 2011. 'Messung und Vergleich von Verwaltungsleistungen: Benchmarking-Regime in Westeuropa', Die Verwaltung: Zeitschrift für Verwaltung und Verwaltungswissenschaft, 2/2011.

Kuhlmann, S. und T. Jäkel 2013. 'Competing, Collaborating or Controlling? Benchmarking-Regimes in European Local Governments from a Comparative Perspective', Public Money and Management, 33, n/a-n/a.

Long, J.S. und J. Freese 2006. Regression Models for Categorical Dependent Variables Using Stata. 2 edn. College Station, Tex.: Stata Press.

Seckelmann, M. 2010. 'Durch Kooperation zum Wettbewerb? - Leistungsvergleiche nach Art. 91d GG', in B. Blanke, F. Nullmeier, C. Reichard und G. Wewer (eds), Handbuch zur Verwaltungsreform. Wiesbaden: VS-Verlag, pp. 571-581.

Siverbo, S. und T. Johansson 2006. 'Relative Performance Evaluation in Swedish Local Government', Financial Accountability & Management, 22, 271-290. doi: 10.1111/j.0267-4424.2006.00403.x.

Van Dooren, W., G. Bouckaert und J. Halligan 2010. Performance Management in the Public Sector. London: Routledge.

Appendix

Abbildung 12: brant-Teststatistik Modell Eins

```
. brant, detail

Estimated coefficients from j-1 binary regressions

                y>1         y>2
schlog00  -1.8944296  -4.1362548
  logbes   .40757244   .67555445
     spd   .65672321  -.76571824
   _cons   5.4685593   9.6001369

Brant Test of Parallel Regression Assumption

    Variable |      chi2   p>chi2    df
-------------+--------------------------
         All |      6.13    0.106     3
-------------+--------------------------
    schlog00 |      1.47    0.226     1
      logbes |      0.06    0.814     1
         spd |      3.89    0.049     1
-------------+--------------------------

A significant test statistic provides evidence that the parallel
regression assumption has been violated.

.
```

Abbildung 13: brant-Teststatistik Modell Drei

```
Estimated coefficients from j-1 binary regressions

                y>1         y>2
schlog00  -2.244382  -4.2496677
  logbes   .77591191   .83771983
     spd   .64516686  -.79872946
     alq   .19694676   .08158957
   _cons   3.1694485   8.4516421

Brant Test of Parallel Regression Assumption

    Variable |      chi2   p>chi2    df
-------------+--------------------------
         All |      7.02    0.135     4
-------------+--------------------------
    schlog00 |      0.85    0.356     1
      logbes |      0.00    0.959     1
         spd |      3.66    0.056     1
         alq |      1.11    0.293     1
-------------+--------------------------

A significant test statistic provides evidence that the parallel
regression assumption has been violated.
```

Abbildung 14: brant-Teststatistik Modell Fünf

```
Estimated coefficients from j-1 binary regressions

                y>1         y>2
schlog00  -2.7625389  -4.0354767
  logbes  -.35344401   .81603191
     soz   .05492598  -.01318613
   _cons   8.7933983   9.0273982

Brant Test of Parallel Regression Assumption

    Variable |      chi2   p>chi2    df
-------------+-----------------------------
         All |      6.83    0.078     3
-------------+-----------------------------
    schlog00 |      0.36    0.548     1
      logbes |      0.89    0.345     1
         soz |      4.85    0.028     1
-------------+-----------------------------

A significant test statistic provides evidence that the parallel
regression assumption has been violated.
```

Abbildung 15: brant-Teststatistik Modell Sechs

```
Estimated coefficients from j-1 binary regressions

                y>1         y>2
schlog00  -1.9551511  -4.4411184
  logbes   .55917336    .7813037
     alq   .20341033   .07680297
   _cons   3.2716113   8.9262541

Brant Test of Parallel Regression Assumption

    Variable |      chi2   p>chi2    df
-------------+-----------------------------
         All |      3.22    0.359     3
-------------+-----------------------------
    schlog00 |      1.37    0.241     1
      logbes |      0.03    0.857     1
         alq |      1.26    0.262     1
-------------+-----------------------------

A significant test statistic provides evidence that the parallel
regression assumption has been violated.
```