4 Konzentrations- und Armutsmessung

4.1 Vorbemerkungen

Konzentration: Ausmaß der Ballung von großen Anteilen an der gesamten Merkmalssumme auf wenige Einheiten.

- Fahrmeir, L. & Künstler, R. & Pigeot, I. & Tutz, G. (7. Auflage, 2010): Statistik: Der Weg zur Datenanalyse. Springer, Berlin, Heidelberg, New York.
- Riedl, B. (2011), Fraktionalisierung von Parteiensystemen aus statistischer Sicht. Bachelorarbeit, Institut für Statistik, LMU München.
- Toutenburg, H., Heumann, C (7. Auflage 2009): Deskriptive Statistik. Springer Verlag, Berlin, Heidelberg, New York.
- Wagschal, Uwe (1999): Statistik für Politikwissenschaftler. Oldenbourg Wissenschaftsverlag, Oldenbourg.
- Links zur Armuts- und Reichtumsberichterstattung der Bundesregierung: `http://www.armuts-und-reichtumsbericht.de/DE/Startseite/start.html`

Bem. 4.1. *Durchgängige Annahmen in Kapitel 4.2 und 4.3*

- Hier zunächst deskriptive Sicht; endliche Gesamtheiten
- X sei ein verhältnisskaliertes Merkmal (mit Urliste $x_1, \ldots, x_n$) und der Größe nach geordneten Ausprägungen $a_1 < a_2 < \ldots < a_k$.

  Ferner seien $f_1, \ldots, f_k$ und $h_1, \ldots, h_k$ die zugehörigen absoluten bzw. relativen Häufigkeiten; die empirische Verteilungsfunktion werde mit $F_X(\cdot)$ oder $F(\cdot)$ bezeichnet.
- Zudem sei $x_i \geq 0$, für alle $i = 1, \ldots, n$, und $\sum\limits_{i=1}^{n} x_i > 0$.

- Betrachtet werden die der Größe nach geordneten Daten:

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

(Achtung: Die Klammern im Index werden in der Literatur oft weggelassen (z.B in (Fahrmeir et al. (2010)); dann muss man vor dem Anwenden der dortigen Formeln die Daten ordnen. Allerdings wird auch hier angenommen, dass $a_1 < a_2 < \ldots < a_k$ gilt, dass also diese Ausprägungen bereits geordnet sind.)

Es gibt eine Unmenge an Konzentrations- und Ungleichheitsmaßen, z.B. Rinne (2003, Taschenbuch der Statistik. Frankfurt am Main) führt über 50 verschiedene Maße auf, hier natürlich nur Beschränkung auf ein paar wesentliche.

Im Folgenden wird öfter, um die verbale Beschreibung nicht zu komplex werden zu lassen, von „reich“ und „arm“ gesprochen, auch wenn andere Merkmale als Einkommen und Vermögen betrachtet werden. Reich steht dann einfach für mit relativ grossem Anteil an der Gesamtsumme (z.B. Umsatz, Stimmanteile).

4.2 Relative Konzentration

4.2.1 Die Lorenzkurve

Definition 4.2.

Gegeben sei die geordnete Urliste $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ eines verhältnisskalierten Merkmals X, das den Annahmen aus Bemerkung 4.1 genügt.
Die stückweise lineare Kurve durch die Punkte $(0, 0)$, (u_1, v_1), (u_2, v_2), …, (u_n, v_n) $= (1, 1)$, wobei für jedes $j = 1, \ldots, n$

$$u_j := \frac{j}{n}$$

und

$$v_j := \frac{\sum\limits_{i=1}^{j} x_{(i)}}{\sum\limits_{i=1}^{n} x_i} = \frac{\sum\limits_{i=1}^{j} x_{(i)}}{\sum\limits_{i=1}^{n} x_{(i)}}$$

heißt Lorenzkurve von X.

- u_j ist der Anteil von j Merkmalsträger an den n Merkmalsträgern: $u_j = F(x_{(j)})$
- v_j der anteilige Beitrag der j kleinsten Merkmalsträgern zur Gesamtsumme $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_{(i)}$ des Merkmals („kumulierte relative Merkmalssumme“)

Bem. 4.3.

a) Insbesondere bei größeren Datensätzen vereinfacht sich die Berechnung wesentlich, wenn man die relativen/absoluten Häufigkeiten $f_1, \ldots, f_k$ bzw. $h_1, \ldots, h_k$ der der Größe nach geordneten Merkmalsausprägungen $a_1 < a_2 <, \ldots, < a_k$ benutzt. Dann ist für $j = 1, \ldots, k$

$$u_j = \sum_{l=1}^{j} \frac{h_l}{n} = \sum_{l=1}^{j} f_l = F(a_j) \tag{4.1}$$

und

$$v_j = \frac{\sum\limits_{l=1}^{j} h_l \cdot a_l}{\sum\limits_{l=1}^{k} h_l \cdot a_l} = \frac{\sum\limits_{l=1}^{j} f_l \cdot a_l}{\sum\limits_{l=1}^{k} f_l \cdot a_l} \tag{4.2}$$

b) Ist bei klassierten Daten mit den Klassen $[c_0, c_1), [c_1, c_2), \ldots, [c_{k-1}, c_k]$ die Merkmalsverteilung in den Klassen nicht bekannt – und will man als Funktionswert dennoch eine einzelne Zahl –, so nimmt man wie beim arithmetischen Mittel als Approximation an, dass alle Ausprägungen in dieser Klasse auf die Klassenmitte $m_\ell = \frac{c_{\ell-1}+c_\ell}{2}$ fallen. Damit erhält man mit f_ℓ und h_ℓ, $\ell = 1, \ldots, k$, als relative bzw. absolute Klassenhäufigkeiten und $a_\ell = m_\ell$, $\ell = 1, \ldots, k$, aus (4.1) und (4.2) direkt:

$$u_j = \sum_{l=1}^{j} \frac{h_l}{n} = \sum_{l=1}^{j} f_l = F(a_j) \tag{4.3}$$

$$v_j = \frac{\sum_{\ell=1}^{j} h_\ell \cdot m_\ell}{\sum_{\ell=1}^{k} h_\ell \cdot m_\ell} = \frac{\sum_{\ell=1}^{j} f_\ell \cdot m_\ell}{\sum_{\ell=1}^{k} f_\ell \cdot m_\ell} \quad . \tag{4.4}$$

Während bei Lorenzkurven unklassierter Daten nur die Punkte $(0, 0), (u_1, v_1), \ldots$ interpretierbar sind, interpretiert man bei klassierten Daten auch die linearen Zwischenstücke.

Bem. 4.4. [Zur Interpretation der Lorenzkurve]

4.2.2 Der Gini-Koeffizient

Definition 4.5. [Gini-Koeffizient]

Gegeben sei die geordnete Urliste $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ eines verhältnisskalierten Merkmals X, das den Annahmen aus Bemerkung 4.1 genügt.

$$G := \frac{2 \cdot \sum_{i=1}^{n} i \cdot x_{(i)}}{n \sum_{i=1}^{n} x_i} - \frac{n+1}{n} \tag{4.5}$$

heißt *Gini-Koeffizient* von X und

$$G_{\text{norm}} := \frac{n}{n-1} \cdot G \tag{4.6}$$

normierter Gini-Koeffizient (Lorenz-Münzner-Koeffizient) von X.

Bem. 4.6.

- Man kann durch die Herleitung über die Trapezformel zeigen (Toutenburg & Heumann (2009)), dass gilt:

$$\begin{aligned} G &= \frac{\text{Fläche zwischen Winkelhalbierender und Lorenzkurve}}{\text{Fläche zwischen Winkelhalbierender und Abszisse}} \\ &= 2 \cdot (\text{Gesamtfläche unter der Winkelhalbierenden – Fläche unter der Lorenzkurve}) \\ &= 2 \cdot \text{Fläche zwischen Winkelhalbierender und Lorenzkurve} \end{aligned}$$

In der Literatur gibt es verschiedene äquivalente Formen, den Gini-Koeffizient zu berechnen.

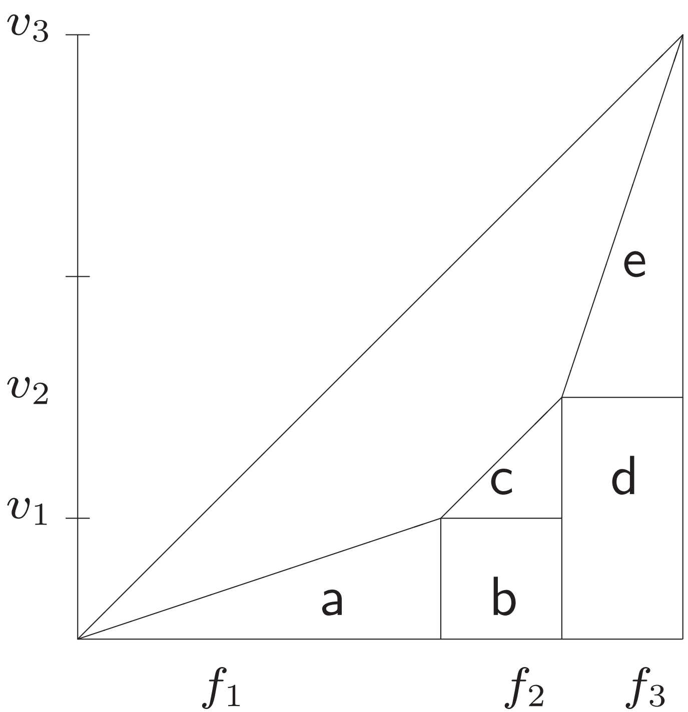
- Betrachtet man wie in Bemerkung 4.1 die geordneten Ausprägungen $a_1 < a_2 < \ldots < a_k$ mit den Häufigkeiten $h_1, h_2, \ldots, h_k$, so gilt:

$$G = \frac{\sum_{l=1}^{k}(u_{l-1} + u_l) f_l \cdot a_l}{\sum_{l=1}^{k} f_l \cdot a_l} - 1 = \frac{\sum_{l=1}^{k}(u_{l-1} + u_l) h_l \cdot a_l}{\sum_{l=1}^{k} h_l \cdot a_l} - 1 = 1 - \sum_{l=1}^{k} f_l (v_{l-1} + v_l) \tag{4.7}$$

- Es gilt bei minimaler Konzentration $G = 0$ und bei maximaler Konzentration $G = \dfrac{n-1}{n}$;

- damit ist also $G_{\text{norm}} = 0$ bei minimaler Konzentration und $G_{\text{norm}} = 1$ bei maximaler Konzentration. (Ist n sehr groß, so ist $\frac{n-1}{n} \approx 1$, also $G^* \approx G$.)

Graphische Veranschaulichung



$$
\begin{aligned}
\underline{\text{Gini Koeffizient}} &= 2 \cdot \textit{Fläche zwischen Winkelhalbierender und Lorenzkurve} = \\
&= 2 \cdot (\frac{1}{2} - (a + b + c + d + e)) = \\
&= 1 - 2 \cdot (a + b + c + d + e) = \\
&= 1 - 2\left(\frac{f_1 \cdot v_1}{2} + f_2 \cdot v_1 + \frac{f_2 \cdot (v_2 - v_1)}{2} + f_3 \cdot v_2 + \frac{f_3 \cdot (v_3 - v_2)}{2}\right) = \\
&= 1 - 2\left(\frac{f_1 \cdot v_1}{2} + \frac{f_2 \cdot v_1}{2} + \frac{f_2 \cdot v_2}{2} + \frac{f_3 \cdot v_2}{2} + \frac{f_3 \cdot v_3}{2}\right) = \\
&= 1 - (f_1 v_1 + f_2 v_1 + f_2 v_2 + f_3 v_2 + f_3 v_3)
\end{aligned}
$$

4.2.3 Quantilsbezogene relative Konzentrationsmessung

Oft stehen die Daten nur in einer anderen Form zur Verfügung (vgl. etwa Beispiel 4.8): Gegeben sind dann bestimmte Quantilsbereiche (typischerweise Quartils-, Quintils- oder Dezilbereiche) und die Anteile des Merkmals, die auf den jeweiligen Bereich entfallen.

Wie kann man dann immer noch die Lorenzkurve berechnen?

Bem. 4.7. *(„Quantilsbezogene Darstellung“)*

Man betrachte für eine gegebene Lorenzkurve eines Merkmals X mit Urliste $x_1, \cdots, x_n$ mit großem n eine „Einteilung der Abszisse“

$$0 =: \alpha_0 < \alpha_1 < \ldots < \alpha_l < \ldots < \alpha_{q-1} < 1 =: \alpha_q$$

(so, dass $\alpha_\ell \cdot n$ ganzzahlig ist. Ferner seien die Ausprägungen von X echt größer als 0 und die Quantile eindeutig) seien $x_{\alpha_\ell}, \ell = 1, \ldots, q$ die zugehörigen Quantile.

Dann ergibt für $\ell = 1, \ldots, q$

$$z_\ell^* := v_{\alpha_\ell \cdot n} - v_{(\alpha_{\ell-1} \cdot n)+1}$$

denjenigen Anteil der Merkmalssumme, der auf Beobachtungen mit einer Merkmalsausprägung in $(x_{\alpha_{\ell-1}}, x_{\alpha_\ell}]$ entfällt.
Wählt man insbesondere die α_ℓ, $\ell = 0, \ldots, q$, äquidistant, also $\alpha_\ell = \ell \cdot \frac{1}{q}$, so erhält man sozusagen *quantilsbezogene* Daten; z_ℓ^*, $\ell = 1, \ldots, q$, sei dann als *ℓ-ter Quantilsanteil*

bezeichnet.

Umgekehrt werden Daten dieser Form $(\alpha_\ell, z_\ell^*)_{\ell=1,\dots,q}$ mit z_ℓ^* wie oben häufig verwendet, um Konzentrationsverhältnisse ohne Angabe der konkreten Merkmalsausprägungen zu charakterisieren. Die Kurve durch die Punkte (u_ℓ^*, v_ℓ^*) mit

$$u_\ell^* = \alpha_\ell \tag{4.8}$$

$$v_\ell^* = \sum_{r \leq \ell} z_r^* \tag{4.9}$$

für $\ell = 0, \dots, q$ werde als *induzierte Lorenzkurve* bezeichnet, und

$$G^* = 1 - \sum_{\ell=1}^{q} f_\ell^* (v_{\ell-1}^* + v_\ell^*) \tag{4.10}$$

mit

$$f_\ell^* := \alpha_\ell - \alpha_{\ell-1}, \quad \ell = 1, \dots, q \tag{4.11}$$

als *induzierter Gini-Koeffizient*.

Die induzierte Lorenzkurve ist wieder eine Lorenzkurve (zum Beispiel für das fiktive Merkmal X^* mit Ausprägungen $x_\ell^* = z_\ell^*/f_\ell^*$, $l = 1, \ldots, q$, und Häufigkeitsverteilung f_ℓ^*, $\ell = 1, \ldots, q$). Ihr Graph verläuft nicht unterhalb der Graphes der Lorenzkurve des Ausgangsmerkmals X; beide Graphen schneiden sich in den Punkten $(u_\ell^*, v_\ell^*)_{\ell=1,\ldots,q}$. Ist G der Ginikoeffizient von X, so gilt $G \geq G^*$.

		1. Bereich	2. Bereich	...	q-ter Bereich
$\alpha_\ell = u_\ell^*$	kumul. Häufigkeiten	α_1	α_2	...	1
f_ℓ^*	Einzelhäufigkeiten	α_1	$\alpha_2 - \alpha_1$	...	$1 - \alpha_{q-1}$
z_ℓ^*	Einzelanteile	z_1^*	z_2^*	...	z_q^*
v_ℓ^*	Gesamtanteile	z_1^*	$z_1^* + z_2^*$	...	$z_1^* + z_2^* + \ldots + z_q^* = 1$

Bsp. 4.8. *[Verteilung des Privatvermögens in Deutschland (*4*. Armuts- und Reichtumsbericht, S. XII)]*

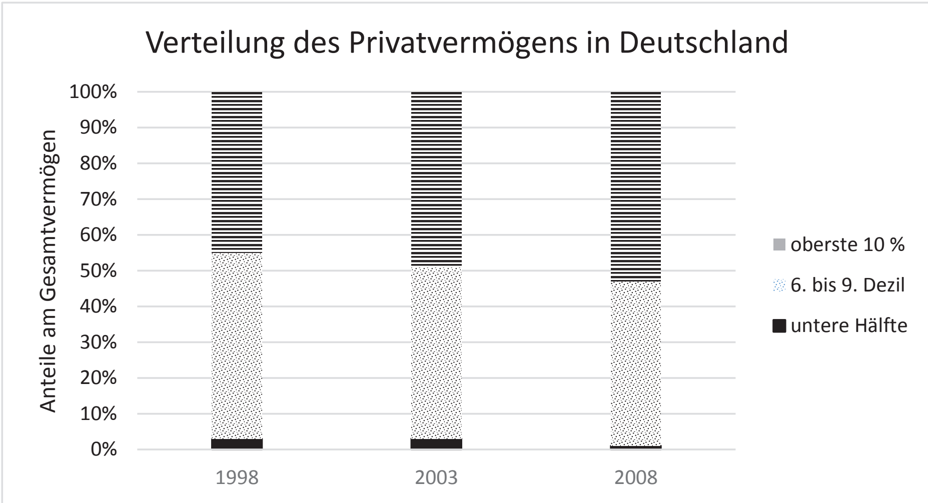


Abbildung 2: Verteilung des Privatvermögens in Deutschland. Quelle: Bundesministerium für Arbeit und Soziales (Hg.) (2013): Vierter Armuts- und Reichtumsbericht der Bundesregierung. Referat Information, Publikation, Redaktion, Bonn. (`www.bmas.de/DE/Service/Medien/Publikationen/a334-4-armuts-reichtumsbericht-2013.html`, aufgerufen am 02.12.16)

Anteile	in der unteren Hälfte	in den $6.$ bis $9.$ Dezil	im obersten Dezil
1998	4 %	51 %	45 %
2003	3 %	48 %	49 %
2008	1 %	46 %	53 %

Bestimmen Sie die Lorenzkurven und den Gini-Koeffizient der Verteilung. Illustrieren Sie die Berechnung des Gini-Koeffizienten auch graphisch.

4.2.4 Einige weitere quantilsbasierte Maße

Insbesondere basierend auf der „natürlichen, äquidistanten Quantilseinteilung“ $\alpha,\ 2\cdot\alpha,\ 3\cdot\alpha,\ldots$ wurden direkt weitere relative Konzentrationsmaße definiert, die natürlich entsprechend verallgemeinert werden können:

Robin-Hood-Index (maximaler Nivellierungssatz, Schutzkoeffizient, (z.B. Wagschal (1999, S.135ff))

- Wie viel müsste den Reichen weggenommen werden, um zu einer Konzentration von 0 zu kommen?

- Betrachte äquidistante Einteilung in q-Quantilsabschnitte:

$$\alpha_\ell = \ell \cdot \alpha, \quad \ell = 1, \ldots, q, \quad \text{mit } \alpha = \frac{1}{q}.$$

- Ermittle für jedes Quantil mit einem Anteil von höchstens α den Abstand seines Anteils zu α!

- Aufaddieren dieser Abstände liefert den Robin-Hood-Index. Dieser Anteil müsste verteilt werden, um zu einer gleichen Verteilung zu kommen!

Also mit der Notation

$$z_\ell^* \quad \text{Anteil im } \ell\text{-ten Quantil, } \ell = 1, \ldots, q$$

und mit

$$|a|_+ := \begin{cases} a & a > 0 \\ 0 & \text{sonst} \end{cases}$$

als Positivteil einer Zahl

$$RHI = \sum_{\ell=1}^{q} |\alpha - z_\ell^*|_+ \left(= \sum_{\ell=1}^{q} |z_\ell^* - \alpha|_+ = \frac{1}{2} \sum_{\ell=1}^{q} |z_\ell^* - \alpha| \right). \tag{4.12}$$

Es gilt ferner mit v_ℓ^* gemäß (4.9)

$$RHI = \max_\ell (\ell \cdot \alpha - v_\ell^*); \tag{4.13}$$

Der RHI ist also der „maximale senkrechte Abstand“ zwischen der Winkelhalbierenden und der Lorenzkurve durch $(u_\ell^*, v_\ell^*)_\ell$ und damit auch der maximale waagerechte Abstand, da ein rechtwinkeliges Dreieck mit Winkeln zu 45 Grad entsteht.

Bem. 4.9. *Quantilverhältnisse*

Ein weiteres anschauliches Maß sind Quantilsverhältnisse, etwa das

- z.B. $Dezilratio\ 90:10 = \dfrac{x_{0.9}}{x_{0.1}}$, falls $x_{0.1} > 0$

- beim Einkommensvergleich: also um welchen Faktor ist der untere Wert der 10% Reichsten größer als der obere Wert der 10% Ärmsten

- Minimale Konzentration: alles in einem Punkt $x_{0.1} = x_{0.9}$
  $\Rightarrow$ Dezilratio $= 1$,

- Umgekehrt Vorsicht: bei extremer Konzentration.
  Die Maßzahl könnte für Entwicklungsländer eventuell problematisch sein, ist aber etwa für Einkommensverhältnisse in OECD-Ländern ein sehr anschauliches Maß.

Bsp.

Einige generelle Anmerkungen

4.3 Absolute Konzentration

4.3.1 Vorbemerkungen: relative versus absolute Konzentration

Vgl. Einleitung: Manche Autor(inn)en verwenden für die „relative“ Konzentration den Begriff „Disparität“. Dann wird „absolute Konzentration“ meist schlicht als Konzentration bezeichnet.

4.3.2 Einige Maßzahlen der absoluten Konzentration

Definition 4.16.

Sei $0 \leq x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ die geordnete Urliste eines verhältnisskalierten Merkmals mit $\sum\limits_{i=1}^{n} x_i > 0$.

Mit

$$p_{(i)} := \frac{x_{(i)}}{\sum_{j=1}^{n} x_j}$$

heißt

$$CR_g := \sum_{i=n-g+1}^{n} p_{(i)}$$

Konzentrationsrate (vom Grade g).

Bsp. 4.17.

Zweitstimmenanteile polit. Parteien bei Bundestagswahlen 1949-2013

	1949	1953	1965	1972	1983	1994	2002	2005	2009	2013
CDU/CSU	31,0%	45,2%	47,6%	44,9%	48,8%	41,5%	38,5%	35,2%	33,8%	41,5%
SPD	29,2%	28,8%	39,3%	45,8%	38,2%	36,4%	38,5%	34,2%	23,0%	25,7%
FDP	11,9%	9,5%	9,5%	8,4%	7,0%	6,9%	7,4%	9,8%	14,6%	4,8%
Grüne	-	-	-	-	5,6%	7,3%	8,6%	8,1%	10,7%	8,4%
PDS/Die Linke	-	-	-	-	-	4,4%	4,0%	8,7%	11,9%	8,6%
Sonstige	27,9%	16,5%	3,6%	0,9%	0,4%	3,5%	3,0%	4,0%	7,0%	11,0%

Definition 4.18.

In der Situation von Def 4.16 heißt

$$H := \sum_{i=1}^{n} p_{(i)}^2$$

Herfindahl-Index. Die Größe $1 - H$ wird auch *Rae-Index* genannt.
In der Politikwissenschaft wird $\frac{1}{H}$ auch als *Anzahl der effektiven Parteien* bezeichnet.

Bsp. 4.21. [Herfindahl- und Rae-Index des deutschen Parteienwesens]

	1972	2005	2009	2013
CDU/CSU	44.9%	35,2%	33.8%	41.5%
SPD	45.8%	34.2%	23.0%	25.7%
FDP	8.4%	9.8%	14.6%	4.8%
Grüne	-	8.1%	10.7%	8.4%
Linke	-	8.7%	11.9%	8.6%
Sonstige (als eine Partei)	0.9%	4,0%	6.0%	11.0%

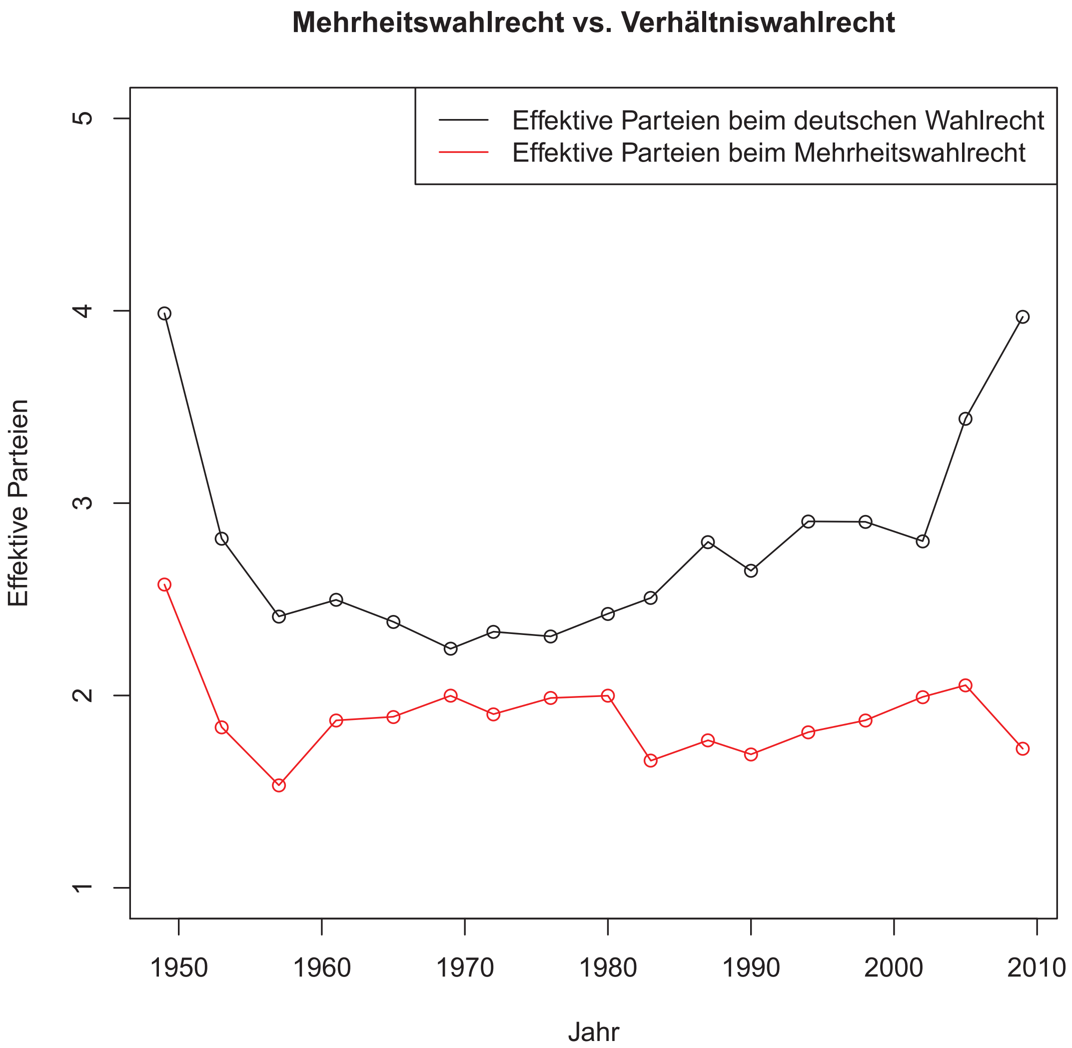


Abbildung 3: Effektive Parteien bei fiktiven Mehrheitswahlrecht über Erststimmen. Quelle: Bachelorarbeit B. Riedl, 2011, Institut für Statistik, LMU München.

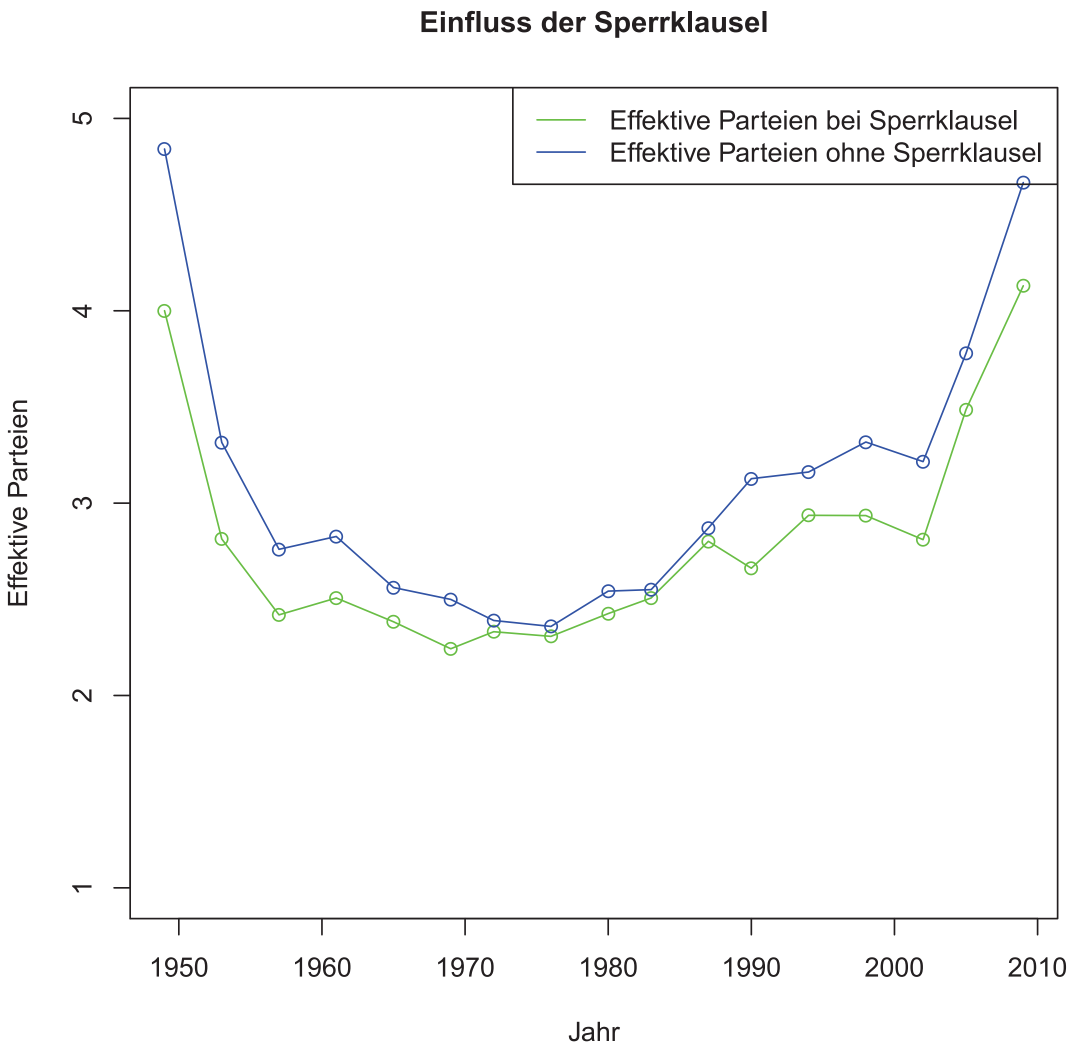


Abbildung 4: Einfluss der Sperrklausel auf die Anzahl der effektiven Parteien bei Wahl zum deutschen Bundestag. Quelle: Bachelorarbeit B. Riedl, 2011, Institut für Statistik, LMU München.

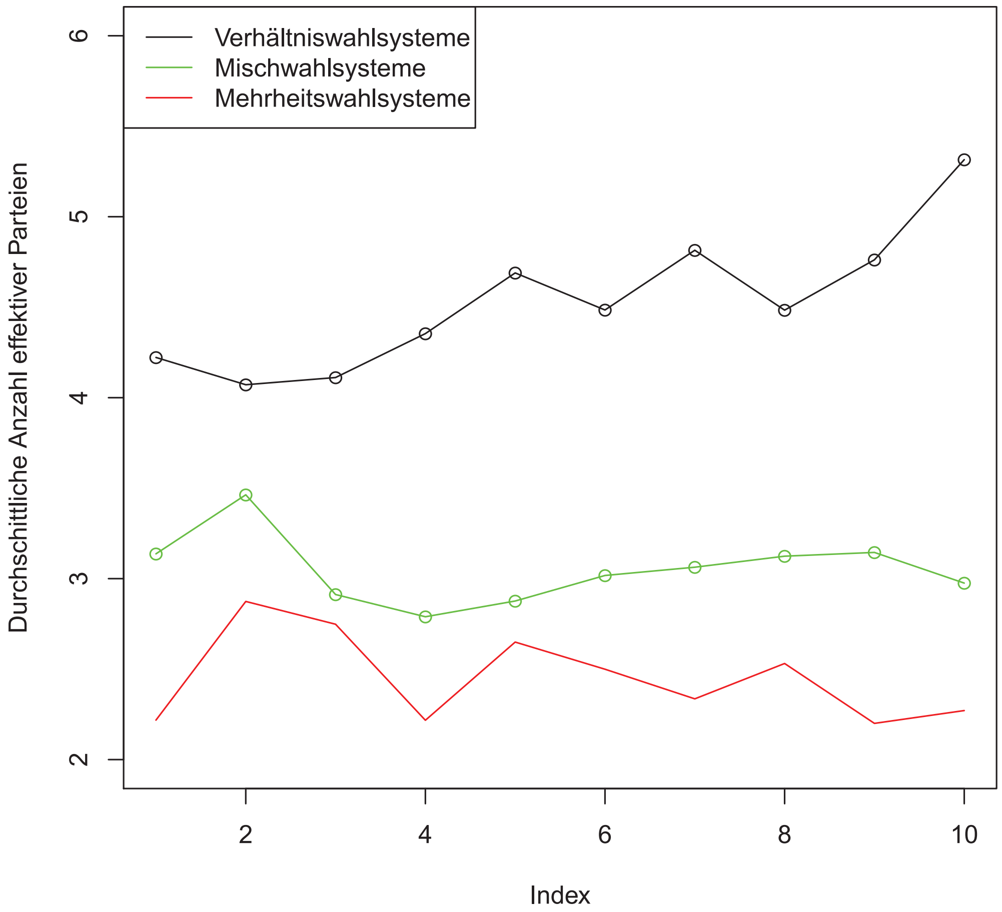


Abbildung 5: Entwicklung der Anzahl effektiver Parteien. Quelle: Bachelorarbeit B. Riedl, 2011, Institut für Statistik, LMU München; Laufindex: letzte 10 Wahlen

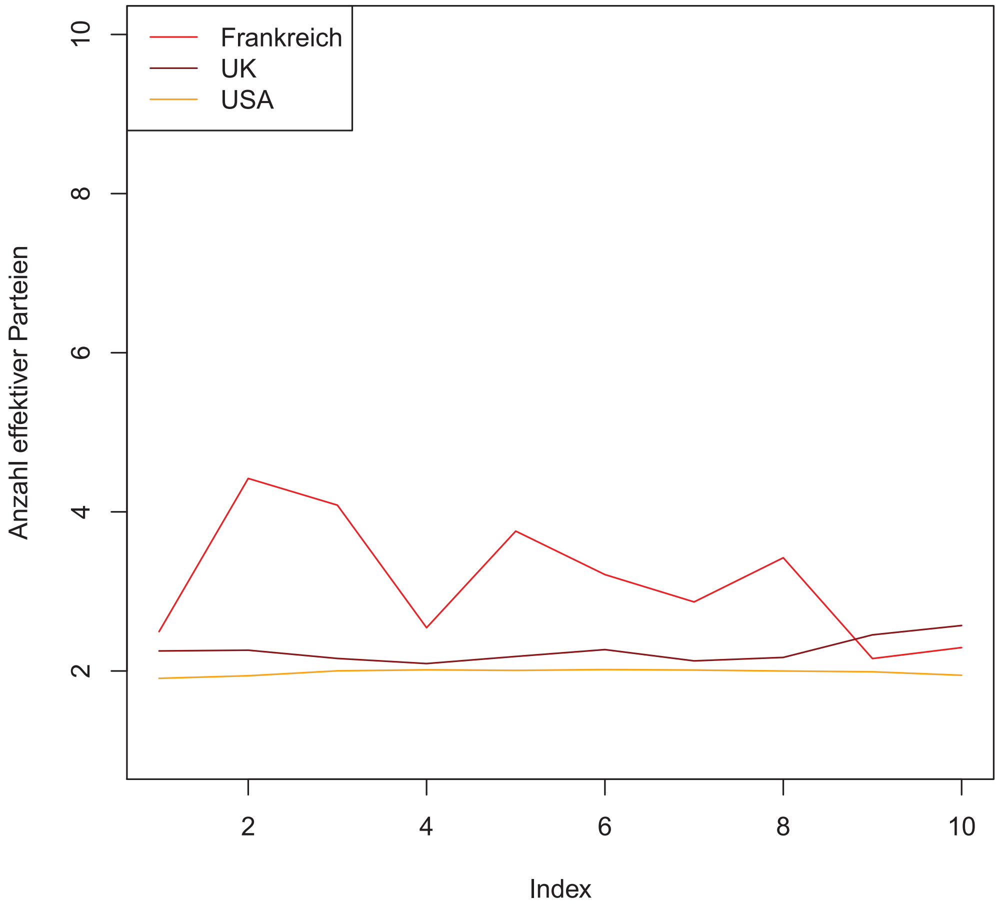


Abbildung 6: Entwicklung der Anzahl effektiver Parteien (Mehrheitswahlsysteme). Quelle: Bachelorarbeit B. Riedl, 2011, Institut für Statistik, LMU München; Laufindex: letzte 10 Wahlen

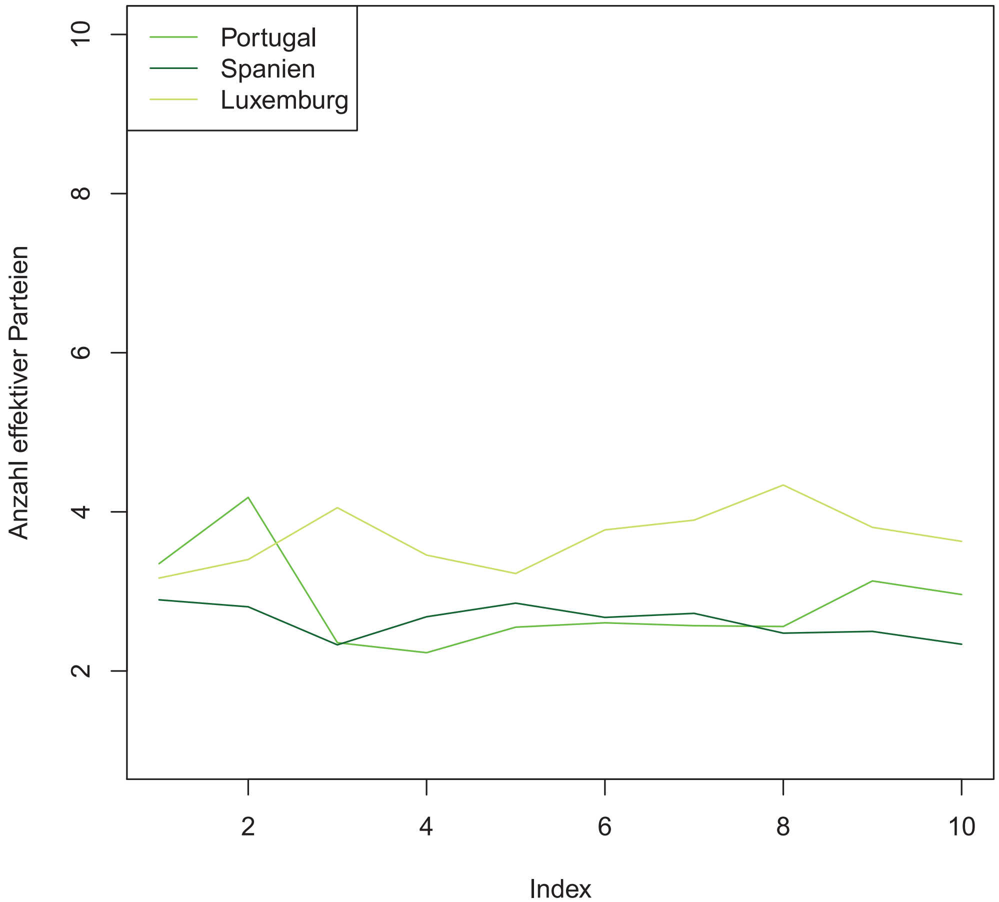


Abbildung 7: Entwicklung der Anzahl effektiver Parteien (Mischwahlsysteme). Quelle: Bachelorarbeit B. Riedl, 2011, Institut für Statistik, LMU München; Laufindex: letzte 10 Wahlen