## Definition & Quelle

SI.POV.DDAY

Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population)

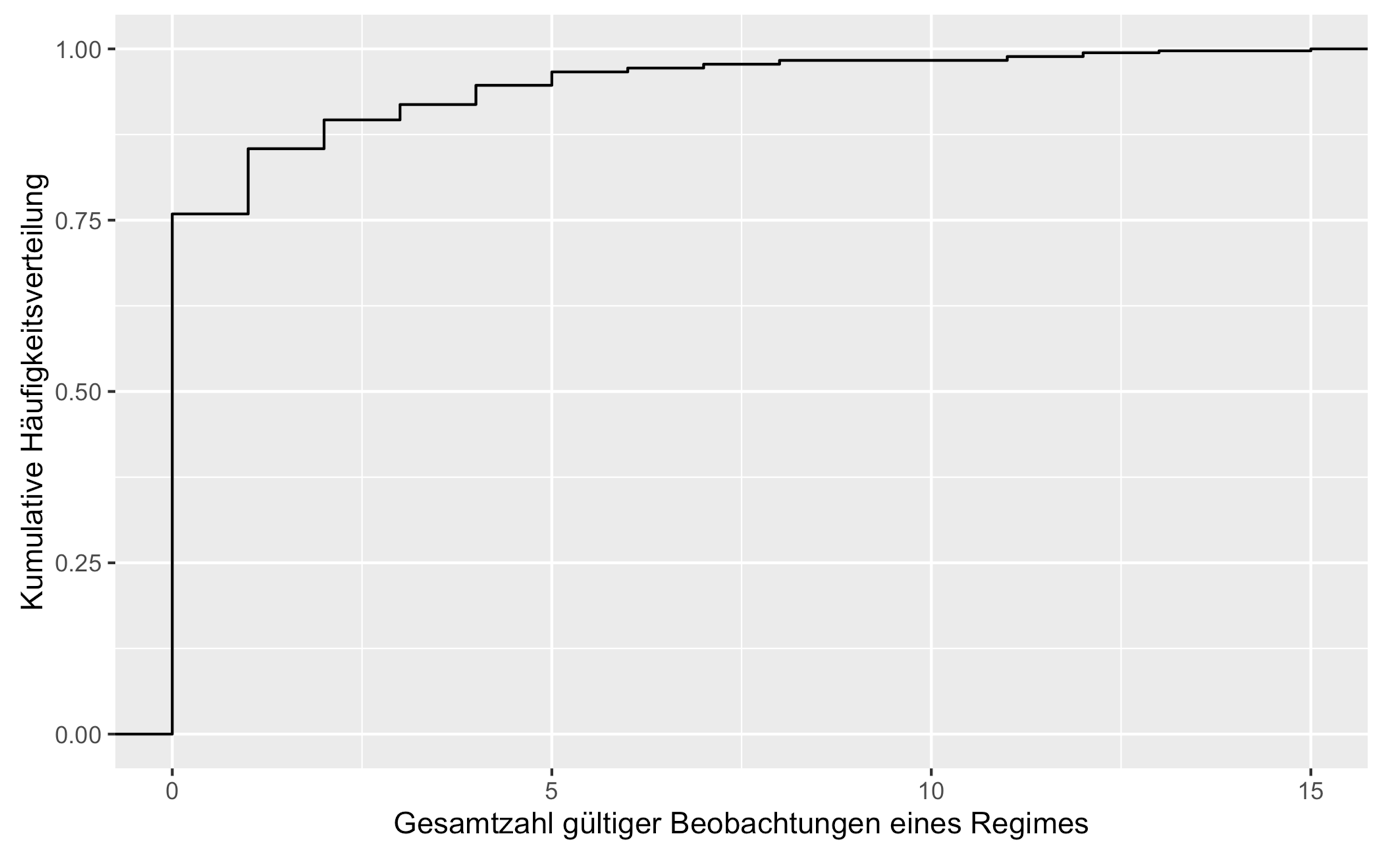
<http://data.worldbank.org/indicator/SI.POV.DDAY> (27.08.2016)

Abruf über R-Package WDI am 25.08.2016

## Datenverfügbarkeit

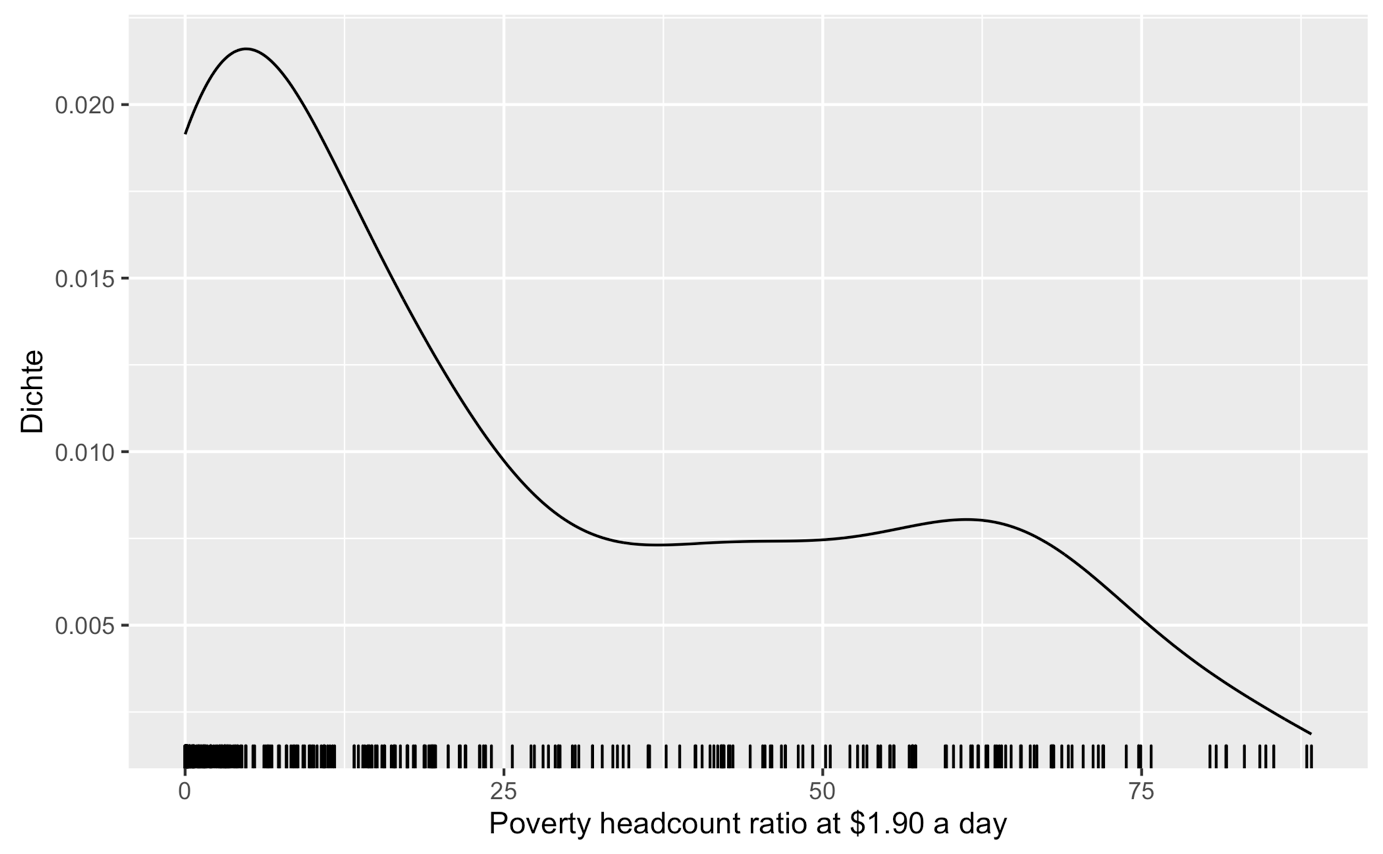
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Missing? | Electoral | Military | Monarchy | Oneparty | Ideocracy | Personalist | Total |
| Ja | 1388 | 574 | 656 | 466 | 672 | 423 | 4179 |
| Nein | 187 | 26 | 19 | 13 | 27 | 7 | 279 |
| Total | 1575 | 600 | 675 | 479 | 699 | 430 | 4458 |

* Über den gesamten Zeitraum von 1980 bis 2010 gibt es nur 279 Messwerte. Über 94 % aller Länderjahre liegen keine Informationen vor.
* Die Datenverfügbarkeit variiert mit dem Regimetyp. Über elektorale Regime liegen die meisten Informationen vor. Alle anderen Regimetypen kommen auf stets weniger als 30 Messwerte. Mit gerade einmal sieben Länderjahren können über personalistische Regime kaum Aussagen getroffen werden.

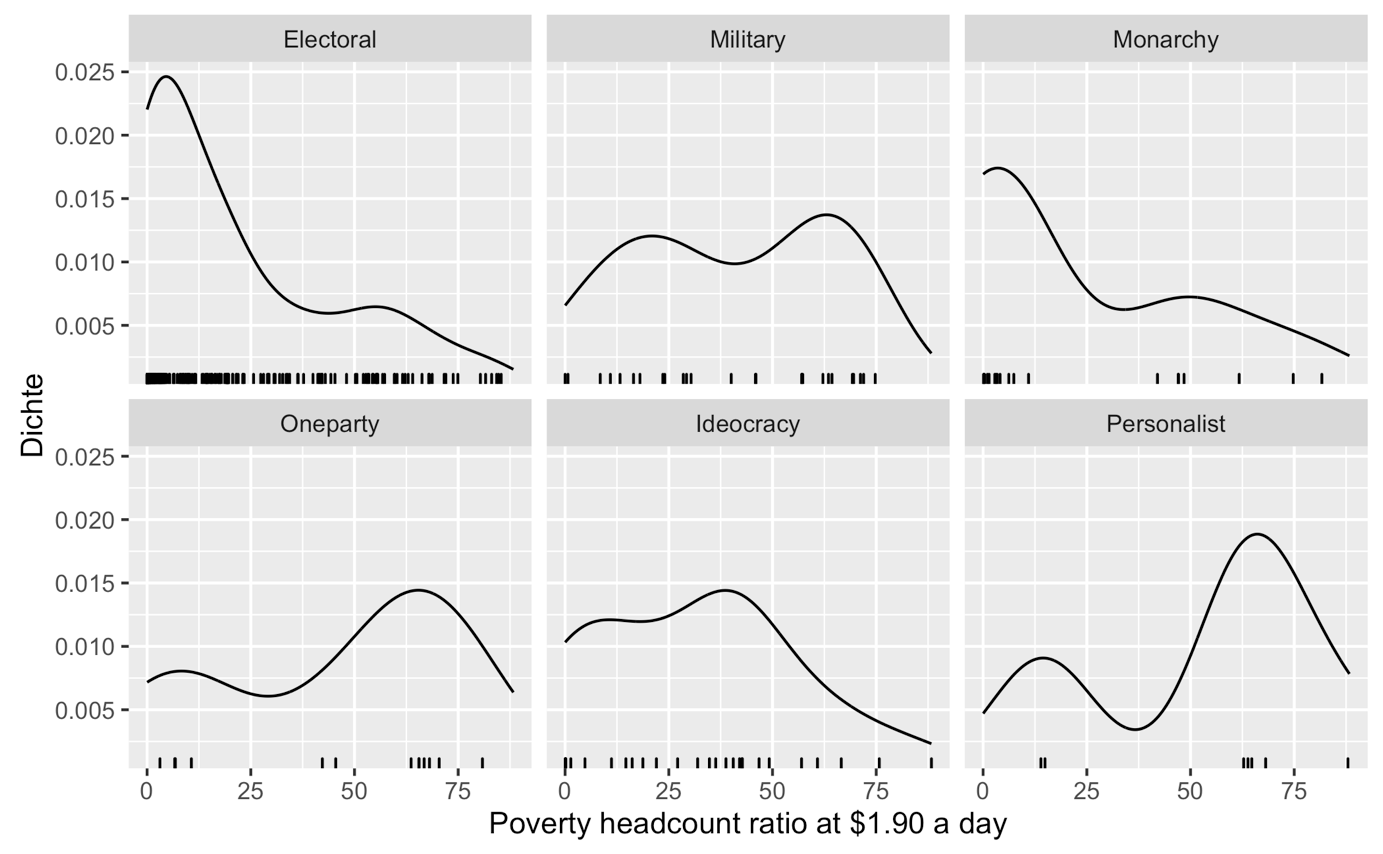


* Wie oft wurde Armut während der Lebenszeit eines Regimes erhoben? Die Grafik zeigt die Verteilungsfunktion der entsprechenden Auszählung.
* Für 75 % aller 360 autoritären Regime stehen überhaupt keine Informationen zur Verfügung!
* Für etwa 5 % aller Regime gibt es einen Messwert, die verbleibenden 20 % verteilen sich zwischen 2 und maximal 15 Messwerten.
* Es steht zu vermuten, dass kein einziges Regime vollständig erfasst wird.

## Einfache uni- und bivariate Exploration

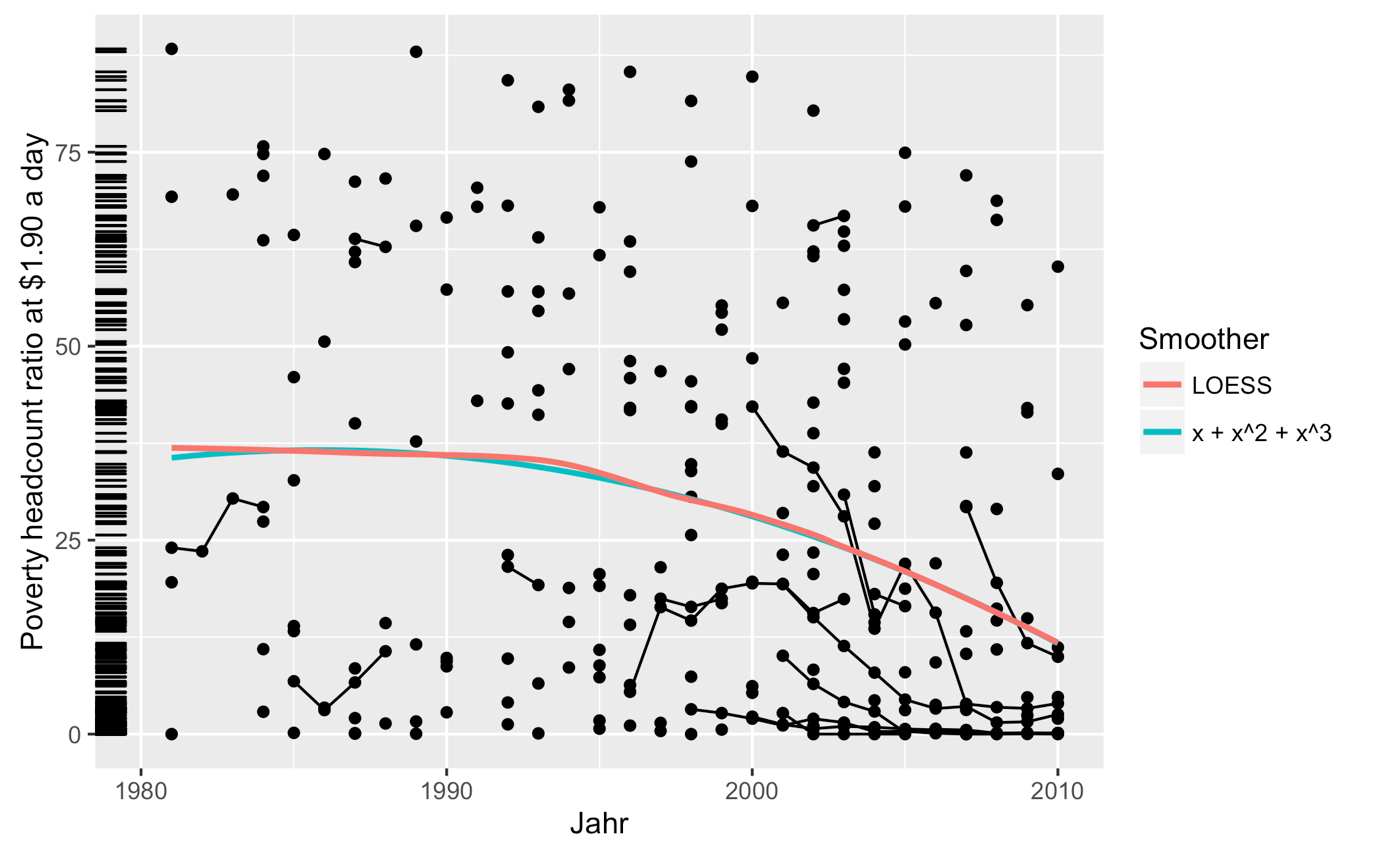


* Die Spannweite der Daten ist beträchtlich, mit 0 resp. 88% wird beinahe der gesamte theoretisch mögliche Wertebereich ausgeschöpft.
* Die meisten autoritären Regime erreichen Armutsraten von unter 10%. Allerdings deutet sich bei etwa 70% ein zweiter Gipfel an. Außerdem ist die Verteilung sehr schief. Bei Werten über 75% handelt es sich wahrscheinlich um Ausreißer.



* Armut variiert in allen Regimetypen über den gesamten Wertebereich.
* Keine Verteilung ist einfach, d.h. eingipflig und symmetrisch. Die überwiegend geringen Fallzahlen machen sich deutlich bemerkbar.
* In der Tendenz weisen elektorale Regime und Monarchien die geringsten Armutsraten auf. Aufgrund der mehrgipfligen Verteilungen sollten Aussagen über die anderen Regimetypen mit großer Vorsicht getroffen werden.

## Trendbestimmung



* Die Grafik zeigt ein Streudiagram der Indikatoren für Armut und Erhebungszeitpunkt. Abgebildet sind Länderjahre, Beobachtungen desselben Regimes werden durch eine schwarze Linie verbunden. Der Rugplot zeigt die univariate Verteilung der AV.
* Die Datenverfügbarkeit streut über den gesamten Erhebungszeitraum. Einzelne Messpunkte füllen die gesamte Skala der Armutsmessung aus.
* Beobachtungen, die über einen längeren Zeitraum aufeinander folgen, gibt es vermehrt erst nach 1990. Dabei handelt es sich zumeist um autoritäre Regime, die auf ein sehr geringes Armutsniveau verweisen können. Das weist auf einen Selektionsbias hin.
* Über den gesamten Zeitraum nimmt Armut in der Tendenz ab (Smoother). Die Parametrisierung über ein kubisches Polynom fängt diese Tendenz sehr gut ein. Komplexe, nicht-parametrische Annäherungen (LOESS) sind nicht notwendig. Aussagen zu Trends innerhalb einzelner Regime sind nicht sinnvoll, da meist nur eine einzige Beobachtung vorliegt.

## Modellierungsansatz

* Zur Auswertung der Daten setze ich ein Mehrebenenmodell für longitudinale Daten ein. Mehrebenenmodelle zeichnen sich durch ihre Sensibilität für hierarchische Datenstrukturen aus. Im vorliegenden Fall handelt es sich um einen Längsschnitt, der autoritäre Regime über Zeit erfasst. Es ist daher zu erwarten, dass die jeweiligen Messungen für ein Regime miteinander korrelieren.[[1]](#footnote-1)
* Im Interesse einer effizienten Modellbildung suche ich zunächst nach einem Basismodell. Dieses soll Heterogenität zwischen autoritären Regimen quantifizieren und zeitabhängige Dynamiken der abhängigen Variablen beschreiben.
  + Zunächst unterstelle ich jedem Regime einen eigenen Achsenabschnitt (random/varying intercept). Er entspricht der Annahme, nicht alle autoritären Regime hätten denselben Erwartungswert auf der abhängigen Variablen. Das kommt der Behauptung gleich, autoritäre Regime unterschieden sich in ihrer sozialen Entwicklung.
  + Trends bestimme ich in zwei Varianten. 1. Beobachtungsjahr, 2. Regimedauer. Die technischen Details sind müßig (Zentrierung von 1. auf 1990, Berechnung orthogonaler Polynome). Letztlich erwies sich nur die erste Variante der Trendbestimmung als sinnvoll. Ich verwende ein kubisches Polynom und halte den Effekt für alle Regime konstant (fixed slope).
* Da nur 279 Beobachtungen für die Schätzung zur Verfügung stehen, weiche ich auf die bayesianische Methoden aus. Sie sind mittlerweile technisch wenig aufwändig und beugen der Kritik vor, wir hätten nicht genügend statistische Masse für üblichere Berechnungsmethoden. Darüber hinaus verwende ich wenig informative Priors für die Koeffizienten (je eine Normalverteilung mit Mittelwert 0 und Standardabweichung 10). Die Ergebnisse kommen deshalb der konventionellen Lösungsehr nahe, bei Verwendung einer uniformen Verteilung wären sie identisch.
* Bei der abhängigen Variablen handelt es sich um eine Prozentangabe. Ich muss sie logarithmieren, um unter Verwendung eines linearen Regressionsmodells sinnvolle Ergebnisse zu erhalten. Bei der letztlichen Darstellung substantieller Effekt kann man die Vorhersagen zurück transformieren.

Die Ergebnisse in Kurzform:

Mean StdDev lower 0.89 upper 0.89

Interc. -1.81 0.24 -2.21 -1.43

t -4.56 8.90 -18.87 9.46

t^2 -12.38 6.99 -24.07 -1.81

t^3 -26.90 4.80 -34.37 -19.17

Residualvarianz 0.62 0.03 0.56 0.67

Interceptvarianz 1.82 0.15 1.57 2.06

N = 279, Regime: 86

1. Die Unterschiede zwischen Regimen sind deutlich größer als die Unterschiede innerhalb von Regimen über Zeit (Interceptvarianz > Residualvarianz).
   1. Die Aufnahme von Regimetypen in das Modell ist sinnvoll, denn sie könnten Querschnittsvarianz, d.i. Unterschiede zwischen Regimen erklären.
   2. Die Transformation der abhängigen Variablen muss geändert werden, denn e^(-1.81+1.82) > 1. Das sollte gerade nicht passieren, wenn die abhängige Variable auf das Intervall von 0 bis 1 beschränkt ist.
2. Es gibt einen komplexen Trend auf der abhängigen Variablen, der sich bereits in einer Grafik weiter oben abzeichnete. Es ließe sich prüfen, ob dieser Trend über Regime variiert. Allerdings halte ich das in Anbetracht der kleinen Datenmenge mit häufig nur einer Beobachtung pro Regime nicht für sinnvoll.

1. Streng genommen gibt es eine weitere Hierarchieebene: Länder. Aufgrund der zahlreichen fehlenden Daten fallen Regime und Länder allerdings meistens zusammen. Nach ersten Analysen gelangte ich zu der Einschätzung, dass nur eine Hierarchieebene, Länder oder Regime, modelliert werden kann. Ich habe mich für Regime entschieden, weil: 1. Sie die Aussageebene des Artikels definieren; 2. Diese Modellierung den Umgang mit Regimetypen erleichtert. [↑](#footnote-ref-1)