

# Messung öffentlicher Aufmerksamkeit mit Wikipedia-Zugriffsstatistiken

Simon Munzert  
Universität Konstanz

Workshop an der ZU Friedrichshafen

03. November 2015

Motivation



# Existierende Messungen öffentlicher Aufmerksamkeit

- MIP-Befragungen (RePass 1971, Miller et al. 1976)
- Medien-Aufmerksamkeit (Baumgartner/Jones 1993, Epstein/Segal 2000)
- Internet-Suchdaten; Google Trends (Ripberger 2011, Mellon 2014)

# Existierende Messungen öffentlicher Aufmerksamkeit

- MIP-Befragungen (RePass 1971, Miller et al. 1976)
- Medien-Aufmerksamkeit (Baumgartner/Jones 1993, Epstein/Segal 2000)
- Internet-Suchdaten; Google Trends (Ripberger 2011, Mellon 2014)

	Umfragedaten (MIP)	Mediendaten	Internetsuchdaten
pros	Individualmessung, lange Zeitreihen, direkte Messung	geringe Kosten, Häufigkeit, Flexibilität, lange Zeitreihen	geringe Kosten, Häufigkeit, Flexibilität, breite Themenabdeckung
cons	Kosten, Spärlichkeit, wenige Issues, Validitätsprobleme	Messung auf Aggregatebene, Brückennahmen	Messung auf Aggregatebene, mangelnde Transparenz, begrenzte Zeitreihe

# Ein alternativer Ansatz

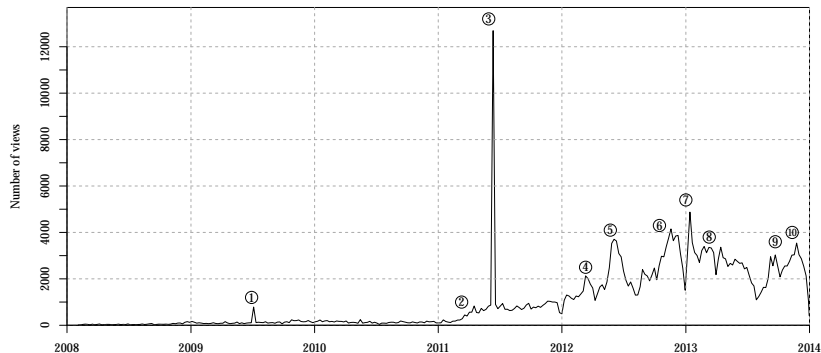
## Strategie

- Auswahl relevanter *Wikipedia*-Artikel, die mit einem bestimmten Issue verknüpft sind
- Beobachtung der täglichen Zugriffe
- Ableitung eines Aufmerksamkeitsmaßes

## ‘Theorie’

- *Wikipedia* ist DIE Informationsquelle im Web
- Nutzer interessieren sich für ein Thema (=Aufmerksamkeit) → schlagen es auf *Wikipedia* nach
- Vorteile: tägliche Messung, auf viele Issues anwendbar, vergleichende Analysen möglich, Messung von Aufmerksamkeit, nicht Meinung

# Beispiel



Wikipedia-Zugriffszahlen für den 'Energiewende'-Artikel.

# Validierungsstrategie

- **Augenschein- und Inhaltsvalidität:**  
Auswahl und Untersuchung geeigneter Artikel
- **Kriteriumsvalidität:**  
Vergleich mit Surveydaten
- **Prädiktive Validität:**  
Verwendung externer Kriterien (Ereignisdaten, politische Aktivitäten etc.)

# Daten

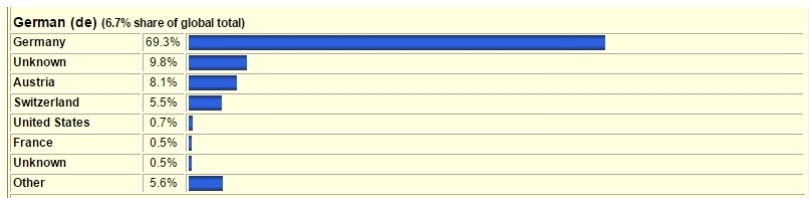
- MIP-Daten: Daten aus dem Politbarometer, seit 2006 alle zwei Wochen erhoben, vorgegebene Kategorien
- *Wikipedia*-Daten: tägliche rohe Zählungen, seit Februar 2008 verfügbar auf <http://stats.grok.se>
- 12 populäre Issues: Arbeitslosigkeit, Steuern, Renten, Inflation, Gesundheitsversorgung, Ausländer, Finanzkrise, Europa, Umwelt, Energiewende, Wirtschaft, Kinderbetreuung



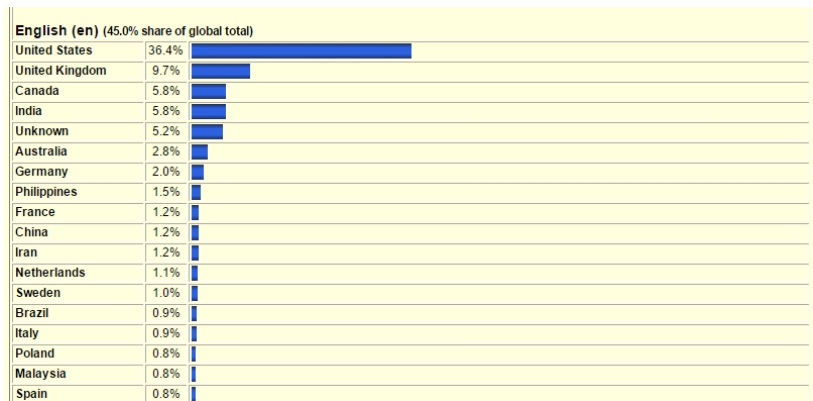
# Artikel (Auswahl)

Issue	Article name	# views	avg. views	max. views
unemployment	Arbeitslosigkeit	1,372,336	641.0	2,434
unemployment	Arbeitslosenstatistik	1,000,568	467.3	2,515
unemployment	Agenda_2,010	728,966	340.5	8,046
unemployment	Arbeitslosengeld	638,748	298.3	990
unemployment	Hartz_IV	459,114	214.4	1,598
taxes	Einkommensteuer_(Deutschland)	3,073,272	1,435.4	9,417
taxes	Umsatzsteuer_(Deutschland)	1,583,643	739.7	2,369
taxes	Steuerrecht_(Deutschland)	358,191	167.3	524
pensions	Gesetzliche_Rentenversicherung_(Deutschland)	2,755,671	1,287.1	3,826
pensions	Pension_(Altersversorgung)	853,575	398.7	1,987
pensions	Altersvorsorge	497,913	232.6	553
pensions	Rentenversicherung_(Erlebensversicherung)	222,401	103.9	250
inflation	Inflation	3,990,138	1,863.7	9,510
inflation	Deflation	1,519,495	709.7	9,100
inflation	Leitzins	926,044	432.5	12,352
inflation	Preisindex	739,752	345.5	1,087
healthcare	Private_Krankenversicherung	1,681,397	785.3	13,765
healthcare	Krankenversicherung	1,507,589	704.2	2,201
healthcare	Gesundheitssystem_Deutschlands	237,830	111.1	286
foreigners_integration	Einwanderung	871,058	406.8	1,597
foreigners_integration	Ausländer	464,084	216.8	1,457
foreigners_integration	Flüchtling	261,845	122.3	1,158
financial_crisis	Finanzkrise_ab_2,007	1,282,737	599.1	4,287
financial_crisis	Finanzkrise	549,077	256.5	5,567
financial_crisis	Bad Bank	280,080	130.8	6,029
financial_crisis	Bankenkrise	172,782	80.7	1,860

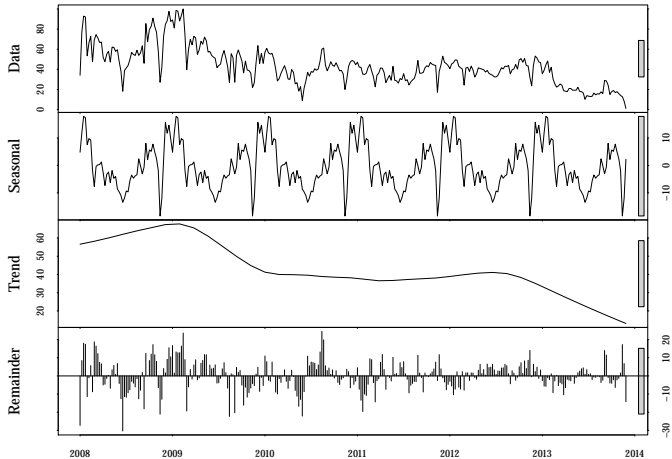
# Nutzerherkunft, deutsche Wikipedia



# Nutzerherkunft, englische Wikipedia



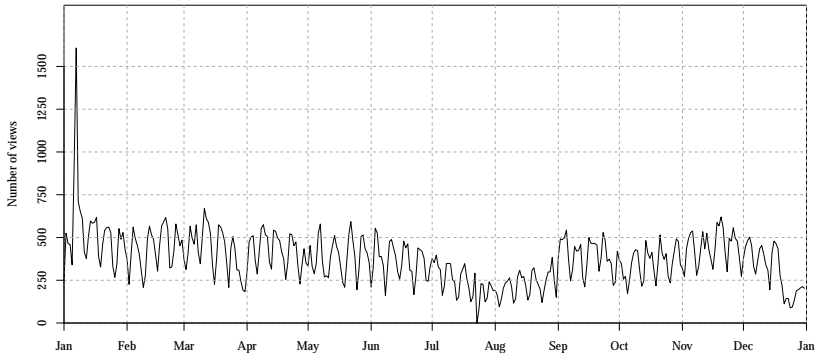
# Saisonalität



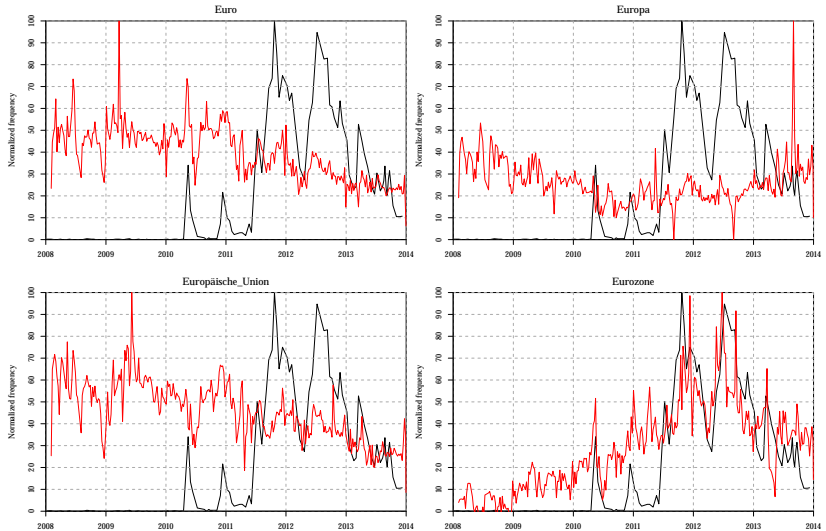
Seasonal trend decomposition by LOESS, *Wikipedia*-Artikel zu Arbeitslosengeld.

# Wöchentliche Saisonalität

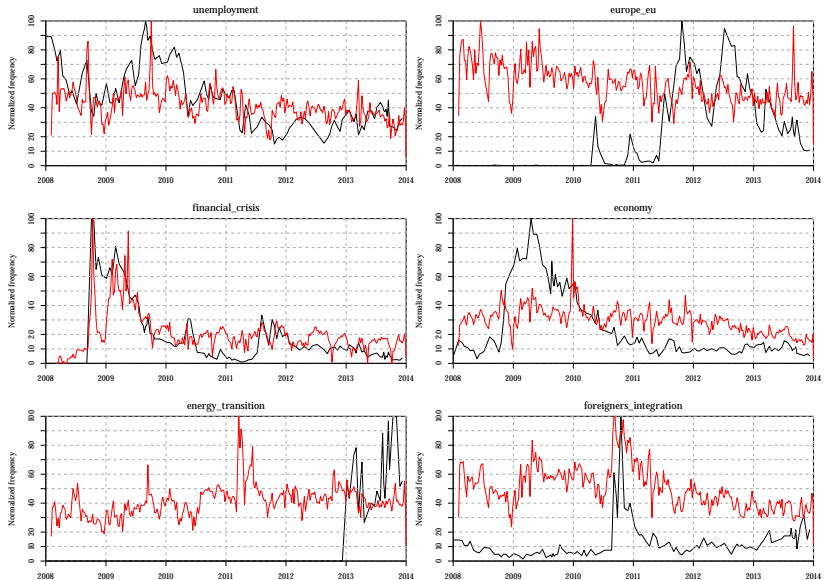
Wikipedia page view counts (red curve) time series for the 'Energiewende' article in 2013. The time series exhibits considerable seasonality because page view statistics use to drop at weekends.



# Kriteriumsvalidität



# Kriteriumsvalidität



# Kriteriumsvalidität

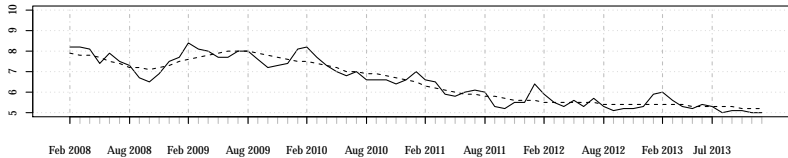
Assessment of criterion validity: OLS regressions of MIP on *Wikipedia* time series, aggregated by issue

Issue	Coefficient	N-W s.e.	R2
Unemployment	1.13	0.21	0.28
Europe/EU	-0.85	0.19	0.16
Financial crisis	1.26	0.14	0.61
Economy	1.27	0.27	0.21
Energy Transition	0.06	0.17	0.00
Foreigners	0.33	0.17	0.16
Child care	0.71	0.13	0.32
Inflation	0.31	0.22	0.08
Pensions	0.14	0.22	0.01
Environment	0.02	0.10	0.00
Taxes	0.24	0.12	0.04
Health care	0.60	0.13	0.34

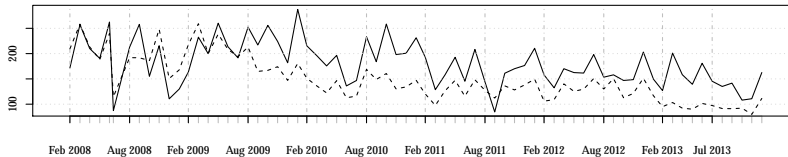


# Vorhersagevalidität

Monthly unemployment rate, raw (solid) and seasonally adjusted (dashed)



Wikipedia traffic volume of unemployment articles, aggregated count (solid) and article on unemployment benefits (dashed)



MIP Politbarometer data, unemployment issue

