Kausalanalyse und Machine Learning in R

Ein Leitfaden für reproduzierbare Forschung

Martin C. Arnold, Christoph Hanck

2022-11-01

Inhaltsverzeichnis

Einleitung

In den letzten Jahren hat sich die datengetriebene Forschung in vielen Fachgebieten, insbesondere in der Ökonometrie und den empirischen Wirtschaftswissenschaften, grundlegend verändert. Haupttreiber dieses Wandels ist die zunehmende Verfügbarkeit von *big data* – hochdimensionale Datenmengen die regelmäßig in Unternehmen und öffentlichen Institutionen anfallen und die Entwicklung sowie den Einsatz neuer statistischer Verfahren im Gebiet *machine learning* prominent gemacht haben. Mit diesen Verfahren können große Datenmengen schnell und effizient verarbeitet und analysiert werden, was ihre zunehmende Relevanz für die evidenzbasierte Entscheidungsfindung in Politik und Wirtschaft begründet.

Ein weiterer Treiber der empirischen wirtschaftswissenschaftlichen Forschung ist die[1]

- Data Science / Varian zitieren: der hat recht behalten
- Reproducibility
- Programmierung / Relevanz von Coding
- kanonischer Unterrichtskatalog in Ökonometrie aufbrechen
- Alles zusammenlegen als ein wichtiger Block für Wirtschaftswissenschaftler und solche die es werden wollen. Und es wird noch wichtiger werden.

(Angrist und Pischke 2010)

[1] Bekannst als *the credibility revolution in empirical economics* (Angrist und Pischke 2010).

1 Statistische Programmierung mit R

Dieses Kapitel ist *nicht* als umfassende Einführung in R gedacht, sondern behandelt Kernkonzepte und soll der Selbsteinschätzung dienen. Wenngleich die Inhalte deutlich über ein Hallo-Welt-Beispiel[1] hinausgehen, betrachten wir hier **absolutes Basiswissen**. Dieses ist Vorraussetzung für das Verständnis fortgeschrittener Konzepte in späteren Kapitel. Falls Sie bereits über solide Grundkenntnisse im Umgang mit R verfügen, können Sie problemlos zu Kapitel XYZ springen. Sollten Sie nicht oder nur teilweise mit den hier gezeigten Befehlen vertraut sein, empfiehlt sich eine Erarbeitung bzw. Wiederholung der Grundlagen. Nachstehede Ressourcen finden wir hilfreich:

- Feedbackgestütze interaktive Übungsaufgaben bei DataCamp, bspw. Einführung in R[2]
- Open-source-Literatur wie der Umfangreiche Leitfaden von Ellis und Mayer (2023).

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

[1] https://de.wikipedia.org/wiki/Hallo-Welt-Programm

[2] Ein Teil des hier angebotenen Katalogs (exlusive *Einführung in R*) ist kostenpflichtig.

1.0.0.1 Pinguine und Pipes

2 Regression Discontiniuty Designs

Regression Discontinuity Design (RDD) ist ein Ansatz für die Schätzung von Treatment-Effekten mit Regression, wenn durch einen experimentell oder natürlich gegebenen Umstand die Behandlung an einem Schwellenwert (C) einer *Laufvariable* (X) sprunghaft beeinflusst wird. Ein RDD-Schätzer brücksichtigt lediglich Beobachtungen mit Ausprägungen von X, die knapp ober- oder knapp unterhalb von C liegen. Die zentrale Idee hierbei ist, dass Individuen nahe bei C im Durchschnitt ähnliche Merkmale aufweisen. RDD isoliert Variation auf dem Pfad *Oberhalb* $C \rightarrow$ *Treatment* $\rightarrow Y$. Somit können Backdoor-Pfade über X oder weitere (möglichweise unbeobachtbare) Confounder (Z) vermieden werden, siehe Abbildung **??**.

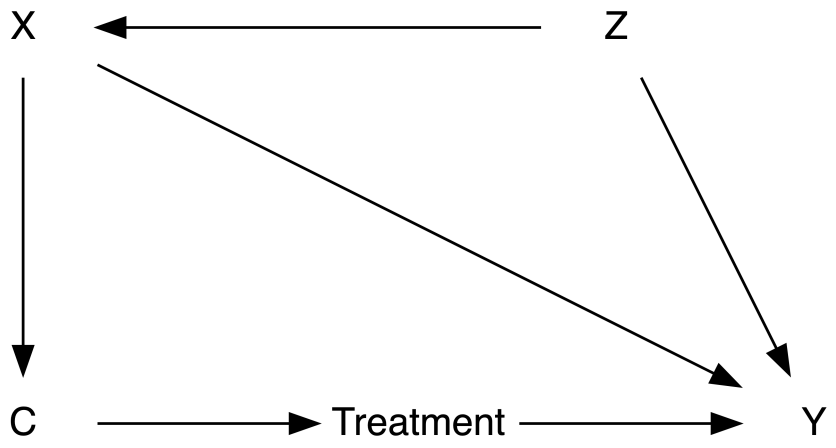

Abbildung 2.1: Causal Diagram für RDD

Der kausale Effekt wird dabei als durchschnittlicher Effekt der Diskontinuität auf die Outcome-Variable (Y) anhand von Beobachtungen *nahe bei* C ermittelt.

Hinsichtlich der Beeinflussung der Behandlung unterscheiden wir zwischen *Sharp* und *Fuzzy* Regression Discontinuity Designs (SRDD/FRDD). Bei einem SRDD ist die Zuordnung *deterministisch*, d.h. der Schwellenwert in der Laufvariable ist eine harte Grenze für die Gruppenzugehörigkeit: Die *Wahrscheinlichkeit* der Behandlung springt bei $X = C$ um $p = 100\%$.

Bei einem FRDD wird angenommen, dass die Zuordnung nicht perfekt durch den Schwellenwert C bestimmt ist: Die *Wahrscheinlichkeit* der Behandlung springt bei $X = C$ um $p < 100\%$.[1] Dies tritt auf, wenn das Überschreiten von C nicht die einzige Determinante einer Behandlung ist. Die Wahl zwischen SRDD und FRDD hängt von der Natur der Daten und der Forschungsfrage ab.

Der Geschätzte Treatment-Effekt ist ein s.g. *local average treatment effect* (LATE).

2.1 Sharp Regression Discontinuity Design

Modelle und funktionale Form

Die korrekte Spezifikation der funktionalen Form für RDD ist wichtig, um eine unverzerrte Schätzung des Effekts zu vermeiden. Die einfachste Form eines SRDD kann anhand der linearen Regression

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 (X_i - C) + u_i \tag{2.1}$$

geschätzt werden, wobei D_i Behandlung anzeigt: D_i ist eine Dummy-Variable für das Überschreiten des Schwellenwertes C, d.h.

$$D_i = \begin{cases} 0 & X_i < C \\ 1 & X_i \geq C. \end{cases}$$

[1] Im FRDD können also sowohl Treatment- als auch Kontroll-Beobachtungen auf beiden Seiten der Diskontinuität beobachtet werden – die Trennung der Gruppen ist "unscharf" (engl. *fuzzy*)

Hierbei ist zu beachten, dass $(X_i - C)$ die um den Schwellenwert zentrierte Laufvariable ist, sodass β_1 der Effekt der Behandlung bei $(X_i - C) \geq 0$ ist.

Modell (**??**) unterstellt, dass X links- und rechtsseitig von C denselben Effekt β_2 auf Y hat. Eine Alternative ist ein lineares Interaktionsmodell

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 (X_i - C) + \beta_3 (X_i - C) \times D_i + u_i. \quad (2.2)$$

In Modell (**??**) erfasst β_3 den Unterschied des Effekts von X auf Y für Beoabachtungen oberhalb von C gegenüber Beobachtungen unterhalb von C, sodass unterschiedliche lineare Effekte von X auf Y link- und rechtsseitig von C modelliert werden können.

Bandweite

Neben der funktionalen Form muss spezifiziert werden, welche Beobachtungen hinreichend nahe am Schwellenwert C liegen. Hierfür verwenden wir eine sogenannte Bandweite h, wobei

$$|(X_i - C)| \leq h \quad (2.3)$$

das Kriterium für eine Berücksichtigung von Beobachtung i bei der Schätzung ist. Unter Berücksichtigung einer Bandweite h wird der Regressionsansatz (**??**) als *local linear regression* mit Uniform-Kernelfunktion bezeichnet.[2] Der Uniform-Kernel ist neben dem Triangular-Kernel eine häufig in der Praxis genutzte lineare Kernelfunktion.[3] Der nachstehende Code plottet die Uniform- (grün) sowie die Triangular-Kernelfunktion (blau) wie in Abbildung **??** dargestellt.

```
library(ggplot2)
library(cowplot)
ggplot() +
    geom_function(
      fun = ~ ifelse(abs(.) <= 1, 1, 0), col = "green"
```

[2] *Local regression* ist ein nicht-parametrisches Verfahren. Hierbei kann die Beziehung zwischen Variablen flexibel modelliert werden.

[3] In der Praxis wird oftmals local linear regression mit linearen Kernelfunktionen zurückgegriffen und die Robustheit der Ergebnisse anhand flexiblerer Spezifikationen geprüft.

```
  ) +
geom_function(
  fun = ~ ifelse(abs(.) <= 1, 1-abs(.), 0), col = "blue"
  ) +
scale_x_continuous("x", limits = c(-1.5, 1.5),
                   labels = c("-h", 0, "h"),
                   breaks = c(-1, 0, 1)) +
scale_y_continuous("K(x)",
                   breaks = c(0, 1),
                   limits = c(0, 1.5)) +
theme_cowplot()
```

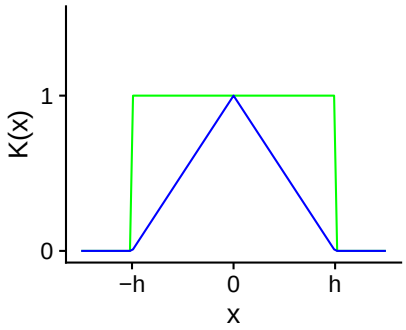

Abbildung 2.2: Uniform-Kernel auf [-h, h]

Die Wahl der Bandweite ist eine wichtige Komponente der RDD-Schätzung. Falls der wahre Zusammenhang nicht-linear ist, erlauben kleine Bandweiten eine Schätzung der Regressionsfunktion nahe des Schwellenwertes mit wenig Verzerrung. Allerdings kann diese Schätzung unpräzise sein, wenn nur wenige Beobachtungen (**??**) erfüllen. In der Praxis wird h daher anhand einer Schätzfunktion (G. Imbens und Kalyanaraman 2012) oder anhand von *Cross Validation* (bspw. G. W. Imbens und Lemieux 2008) bestimmt. Die später in diesem Kapitel betrachteten R-Pakete halten diese Methoden bereit.

2.1.1 Beispiel: Amtsinhaber-Vorteil (Lee 2008)

Lee (2008) untersucht den Einfluss des Amtsinhaber-Vorteils auf die Wahl von Mitgliedern des US-Repräsentantenhauses. Entfällt ein Stimmenanteil von mehr als 50% auf eine Partei, hat diese Partei den Vorsitz des Repräsentantenhauses gewonnen. Durch die Analyse der 6558 Wahlen im Zeitraum 1946-1998 mit RDD kommt die Studie zu dem Ergebnis, dass die amtsinhabende im Durchschnitt einen Vorteil von etwa 8%-10% bei der Wahl hat. Dieses Ergebnis kann als Vorteil kann durch verschiedene Faktoren erklärt werden, bspw. dass die amtierende Partei höhere finanzielle Ressourcen und von einer besseren Organisation als die Opposition profitiert. Anhand der Datensätze `house` und `house_binned` illustrieren wir nachfolgend die Schätzung von SRDD-Modellen für den Wahlerfolg der demokratischen Partei, wenn diese Amtsinhaber ist.

Wir verschaffen uns zunächst einen Überblick über den Datensatz `house`.

R lädt. Etwas Geduld bitte…

Der Datensatz `house` enthält Stimmenanteile der Demokraten bei der Wahl zum Zeitpunkt T (`StimmenT`) sowie um den Schwellenwert von 50% zentrierte Stimmenanteile bei der vorherigen Wahl zum Zeitpunkt $T-1$ (`StimmenTm1`).

`house_binned` ist eine aggregierte Version von `house` mit Mittelwerten von jeweils 50 gleichgroßen Intervallen oberhalb und unterhalb der Schwelle von 0. Dieser Datensatz eignet sich, um einen ersten Eindruck des funktionalen Zusammenhangs auf beiden Seiten zu gewinnen. Wir stellen die klassierten Daten mit `ggplot2` graphisch dar.

R lädt. Etwas Geduld bitte…

Die Grafik zeigt eindeutig einen Sprung an der Stelle $StimmenTm1 = 0$. Weiterhin erkennen wir, dass der Zusammenhang nahe 0 jeweils gut durch eine lineare Funktion approximiert werden kann. Eine Modell-Spezifikation mit gleicher Steigung auf beiden Seiten des Schwellenwertes scheint hingegen ungeeignet.

Als nächstes fügen wir dem Datensatz eine Dummyvariable `D` hinzu. Diese dient als Indikator für den Wahlgewinn in der letzten Wahl.

R lädt. Etwas Geduld bitte…

Als nächstes schätzen wir das Modell

$$\text{StimmenT}_i = \beta_0 + \beta_1 D_i + \beta_2(\text{StimmenTm1}_i - 50) + \beta_3(\text{StimmenTm1}_i - 50) \times D_i + u_i \tag{2.4}$$

bei einer Bandweite von $h = 1$. Aufgrund der Skalierung der Daten bedeutet dies die Verwendung des *gesamten* Datensatzes für die Schätzung.

R lädt. Etwas Geduld bitte…

Der geschätzte Koeffizient von D (`DTRUE`) beträgt etwa 0.12 und ist hochsignifikant. Übereinstimmend mit der (oben erstellten) Grafik erhalten wir also eine positive Schätzung des Treatment-Effekts. Diese Schätzung könnte jedoch invalide sein:

- Die (implizite) Wahl von $h = 1$ macht die Isolation des relevanten Frontdoor-Paths ($C = 0 \rightarrow$ Treatment $\rightarrow$ StimmenT) wenig plausibel.
- Weiterhin könnte die lineare funktionale Form unser Regression für den gesamten Datensatz inadäquat sein: Die lineare Approximation könnte lediglich "lokal" zu 0 gut sein und anderweitig in einer verzerrten Schätzung des Effekts resultieren.

```
round(rdd::IKbandwidth(X = house$StimmenTm1, Y = house$StimmenT), digits = 4)
```

```
[1] 0.2685
```

R lädt. Etwas Geduld bitte…

Die nachstehende interaktive Abbildung zeigt die klassierten Daten aus Lee (2008) auf dem Intervall $[-0.5, 0.5]$ gemeinsam mit einer nicht-parametrischen Schätzung des Zusammenhangs von `StimmenT` und `StimmenTm1` mit LOESS.[4] Über die Inputs kann eine gemeinsame Bandweite $l \in (0, 1]$ für den LOESS-Schätzer auf beiden Seiten des Schwellenwerts 0 und die Bandweite der Daten (h) um den Schwellenwert angepasst werden. Beachte:

- Der geschätzte Zusammenhang ist etwa linear für die voreingestellten Parameter ($l = 1$, $h = 0.28$).
- Für kleine l passt sich die Schätzung Besser an die Datenpunkte an. Zu kleine Werte führen jedoch zu *overfitting*. Insbesondere hat die geschätzte Funktion eine zu extreme Steigung nahe des Schwellenwerts $\rightarrow$ verzerrte Schätzung des Effekts!
- Kleine Werte h bedeuten eine geringe Datenbasis, sodass LOESS selbst bei großen l zu nicht-linearen Anpassungen der Regressionsfunktion tendiert. Für solche Parameterkombinationen erfolgt die Schätzung des Effekts mit hoher Varianz.

[4] LOESS ist eine nicht-parametrische Regressionsmethode, die lokale Gewichte verwendet, um nicht-lineare Zusammenhänge zwischen Variablen zu modellieren und geglättete Funktionen anzupassen.

SRDD mit nicht-parametrischer Regression

```
//| echo: false
html`
<style>
circle {
  fill-opacity: .8;
  stroke: #000;
  stroke-opacity: 1;
}
.regression {
  fill: none;
  stroke: #000;
  stroke-width: 1.5px;
}
.axis line {
  stroke: #ddd;
}
.axis .baseline line {
  stroke: #555;
}
.axis .domain {
  display: none;
}
</style>
`

d3 = require("d3-array@3", "d3-axis@3", "d3-regression@1", "d3-scale@4", "d3-shape@3", "d3-s

margin = ({left: 55, right: 8, top: 13, bottom: 24});
base = Math.min(width, 500);
innerWidth = base - margin.left - margin.right;
innerHeight = base-100 - margin.top - margin.bottom;

viewof bandwidth = Inputs.range([.01, 1], {
  label: "Bandweite LOESS (l)",
  step: .01,
  value: 1
});
```

```
viewof bw_daten = Inputs.range([.05, .5], {
  label: "Bandweite Daten (h)",
  step: .01,
  value: .28
});

xScaleLoess = d3.scaleLinear()
   .domain([-.55, .55])
   .range([0, innerWidth]);

yScaleLoess = d3.scaleLinear()
  .domain([.2, .8])
  .range([innerHeight, 0]);

lineLoess = d3.line()
  .x(d => xScaleLoess(d[0]))
  .y(d => yScaleLoess(d[1]));

xAxisLoess = d3.axisBottom(xScaleLoess)
  .tickSize(innerHeight + 10)
  .tickValues([-.5, -.25, 0, .25, .5])
  .tickFormat(d => d);

yAxisLoess = d3.axisLeft(yScaleLoess)
  .tickSize(innerWidth + 10)
  .tickValues([.2, .35, .5, .65, .8])
  .tickFormat(d => d);

loessRegression = d3.regressionLoess()
  .x(d => d.StimmenTm1)
  .y(d => d.StimmenT)
  .bandwidth(bandwidth);
```

```
//| echo: false
//| fig-cap: "Nicht-parametrische Regression auf beiden Seiten des Cut-offs."

{
  const svg = d3.select(DOM.svg(innerWidth + margin.left + margin.right + 20, innerHeight +
```

```
const g = svg.append("g")
    .attr("transform", `translate(${margin.left}, ${margin.top})`);

g.append("g")
    .attr("class", "axis")
    .call(xAxisLoess);

g.append("g")
  .attr("class", "axis")
  .attr("transform", `translate(${innerWidth})`)
  .call(yAxisLoess);

// Add X axis label:
g.append("text")
  .attr("text-anchor", "end")
  .attr("font-size", 13)
  .attr("x", innerWidth)
  .attr("y", innerHeight + margin.top + 25)
  .text("Stimmenanteil Demokraten letzte Wahl");

// Y axis label:
g.append("text")
  .attr("text-anchor", "end")
  .attr("transform", "rotate(-90)")
  .attr("font-size", 13)
  .attr("y", -margin.left+10)
  .attr("x", -margin.top+10)
  .text("Stimmenanteil Demokraten");

// Distance at jump
g.append("text")
 .attr("x", 250) // x-Position des Textelements
 .attr("y", 200) // y-Position des Textelements
 .text("") // Textinhalt
 .attr("font-size", "14px") // Schriftgröße
 .attr("fill", "black"); // Textfarbe


g.selectAll("circle")
```

```
  .data(transpose(house_binned))
  .enter().append("circle")
  .attr("r", 2)
  .attr("cx", d => xScaleLoess(d.StimmenTm1))
  .attr("cy", d => yScaleLoess(d.StimmenT));

g.append("path")
    .attr("class", "regression")
    .datum(loessRegression(
      transpose(house)
        .filter(function(d){ return d.StimmenTm1 <= 0 & d.StimmenTm1 >= -bw_daten })
        )
      )
    .attr("d", lineLoess)
    .style("stroke", "red");

g.append("path")
    .attr("class", "regression")
    .datum(loessRegression(
      transpose(house)
        .filter(function(d){ return d.StimmenTm1 > 0 & d.StimmenTm1 <= bw_daten })
        )
      )
    .attr("d", lineLoess)
    .style("stroke", "red");

/* dashed line at cutoff */
g.append("line")
.attr("x1", xScaleLoess(0))
.attr("y1", 0)
.attr("x2", xScaleLoess(0))
.attr("y2", innerHeight)
.style("stroke", "black")
.style("stroke-dasharray", "1")
.style("stroke-width", "1");

/* dashed line data bw upper */
g.append("line")
.attr("x1", xScaleLoess(bw_daten))
```

```
  .attr("y1", 0)
  .attr("x2", xScaleLoess(bw_daten))
  .attr("y2", innerHeight)
  .style("stroke", "blue")
  .style("stroke-dasharray", "4")
  .style("stroke-width", "1");

  /* dashed line data bw lower */
  g.append("line")
  .attr("x1", xScaleLoess(-bw_daten))
  .attr("y1", 0)
  .attr("x2", xScaleLoess(-bw_daten))
  .attr("y2", innerHeight)
  .style("stroke", "blue")
  .style("stroke-dasharray", "4")
  .style("stroke-width", "1");

  return svg.node();
}
```

2.2 Fuzzy Regression Discontinuity Design

2.3 Case Study: Protestantische Arbeitsethik

Die Studie *Beyond Work Ethic: Religion, Individual, and Political Preferences* (**BastenBetz2013?**) untersucht den Zusammenhang zwischen Religion, individuellen Merkmalen und politischen Präferenzen. Das Hauptaugenmerk ist die Rolle von Religiosität als Einflussfaktor auf politische Einstellungen. Die Hypothese ist, das Religiosität eines Individuums über den traditionellen Rahmen von Moralvorstellungen und sozialen Normen hinaus auch politische Präferenzen beeinflusst. Diese Theorie wird prominent von Max Weber (1930) vertreten, der argumentiert, dass die protestantische Arbeitsethik einen entscheidenden Einfluss auf die Entwicklung des Kapitalismus hatte. Laut Weber führte der protestantische Glaube an harte Arbeit,

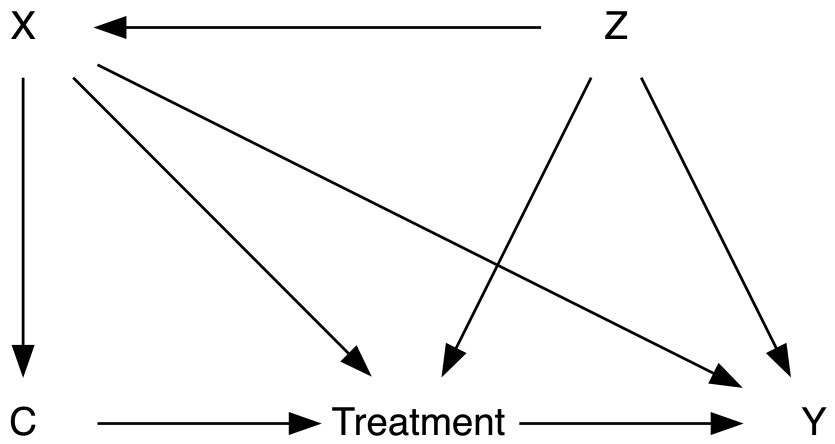

Abbildung 2.3: Causal Diagram für FRDD

ein sparsames Leben und ethisches Verhalten zur einer verbreiteten "Geisteshaltung" , die den wirtschaftlichen Erfolg förderte und den Aufstieg des Kapitalismus begünstigte.

(**BastenBetz2013?**) nutzen Daten aus dem World Values Survey,

Sie analysieren insbesondere die Zusammenhänge zwischen Religiosität, individuellen Merkmalen wie Geschlecht, Bildung und Einkommen sowie politischen Präferenzen wie links-rechts-Ausrichtung, Einstellungen zur Umverteilung und zur Einwanderung.

Die Ergebnisse der Studie zeigen, dass Religiosität tatsächlich einen signifikanten Einfluss auf politische Präferenzen hat. Insbesondere stellen die Autoren fest, dass religiöse Menschen tendenziell konservativere Einstellungen haben und eher rechten politischen Parteien zuneigen. Dieser Effekt bleibt auch nach Kontrolle anderer Faktoren wie Bildung und Einkommen bestehen.

Darüber hinaus betonen die Autoren, dass der Zusammenhang zwischen Religion und politischen Präferenzen nicht allein

durch moralische Werte erklärt werden kann. Sie argumentieren, dass religiöse Institutionen auch eine soziale und politische Agenda verfolgen, die von den Gläubigen internalisiert wird. Diese Agenda kann beispielsweise Positionen zu Themen wie Abtreibung, gleichgeschlechtlicher Ehe, Einwanderung oder Umweltschutz umfassen.

Zusammenfassend zeigt das Paper "Beyond Work Ethic: Religion, Individual, and Political Preferences" von Christoph Basten und Frank Betz, dass Religiosität einen Einfluss auf politische Präferenzen hat, der über traditionelle Moralvorstellungen hinausgeht. Es deutet darauf hin, dass religiöse Menschen eher konservative politische Ansichten haben und dass religiöse Institutionen eine Rolle bei der Formulierung dieser Ansichten spielen können.

```
library(tidyverse)
library(haven)
library(vtable)
library(rdrobust)
```

```
BastenBetz <- read_dta('BastenBetz.dta')
```

```
# Table 1
T1 <- BastenBetz %>%
    filter(abs(borderdis) < 5.0283684) %>%
    transmute(
        group = ifelse(vaud == 1, "T", "C"),
        prot1980s = prot1980s * 100,
        reineink = reineink_pc_mean * 1000,
        noreligion1980s,
        altitude,
        pfl,
        pfr,
        pfi,
        gini = Ecoplan_gini
    ) %>%
    group_by(group) %>%
    summarise(
```

```
      across(everything(), list(Mean = mean, SD = sd, N = length))
      ) %>%
    pivot_longer(
      -group,
      names_to = c("variable", "statistic"),
      names_sep = "_"
    )
```

```
T1_wider <- T1 %>%
    pivot_wider(
        names_from = c("group", "statistic")
    )
```

```
T1_wider %>%
  gt(rowname_col = "Variable") %>%
  tab_spanner_delim(
    delim = "_",
  ) %>%
 tabopts
```

Tabelle 2.1: Datensatz `BastenBetz` – Zusammenfassende Statistiken

	C			T		
variable	Mean	SD	N	Mean	SD	N
prot1980s	9.428	5.695	49	83.245	11.411	84
reineink	43,692.71	3,369.175	49	47,253.272	5,342.36	84
noreligion1980s	1.729	1.499	49	2.95	2.726	84
altitude	642.592	120.23	49	639.607	113.564	84
pfl	48.239	4.774	49	39.508	5.723	84
pfr	43.049	2.634	49	39.19	5.025	84
pfi	52.642	2.94	49	47.086	3.368	84
gini	0.302	0.029	49	0.367	0.052	84

Imbens und Kalyanaraman (2012) zeigen

```r
bw_selection <- rdbwselect(
  y = BastenBetz$pfl,
  x = BastenBetz$borderdis,
  fuzzy = BastenBetz$prot1980s,
  bwselect = "mserd",
  kernel = "uniform")

summary(bw_selection)
```

```
Call: rdbwselect

Number of Obs.                  509
BW type                       mserd
Kernel                      Uniform
VCE method                       NN

Number of Obs.                  127          382
Order est. (p)                    1            1
Order bias  (q)                   2            2
Unique Obs.                      97          261

=============================================================
                  BW est. (h)    BW bias (b)
            Left of c Right of c  Left of c Right of c
=============================================================
     mserd     5.078      5.078     10.905     10.905
=============================================================
```

```r
OB <- bw_selection$bws[1]
```

```r
# Table 2: First stage results
# (1) (close to the) Imbens and Kalyanaraman (2012) optimal BW of 5.01 reported in
# the paper
FS1 <- lm(
  formula = prot1980s ~ vaud + borderdis + vaud * borderdis,
  data = BastenBetz %>% filter(abs(borderdis) < OB)
  )
```