

# Bevölkerungsszenarien für die Stadt Zürich — Implementierung und Sensitivitätsanalyse mit

R-Meetup «adminR: R in Swiss Official Statistics»  
20. September 2018

Andreas Papritz  
Klemens Rosin



**Stadt Zürich**  
Statistik

# Zürich ist auf

Die Bevölkerung der Stadt wächst

Die Stadt Zürich wächst ungebrochen und sogar stärker als von den Statistikern erwartet. Setzt sich der Trend fort, wird die Bevölkerung 2020 ein Rekordniveau erreichen.

MICHAEL VON LEDEBUR

Es gab eine Zeit, da verlieh der Stadt Zürich gerne ein Rating. Nicht die Bonität wie meinst, und die Zuschreibung kein Kompliment. AAA für Ausländer und Arbeitskräfte der Stadt den Eindruck, der vor durchaus faktenbasierter neunziger Jahre Stadt jährlich umsonen. 1989 waren 440 180 Einwohner. Im Jahr 1962 sind inzwischen umgerückt. Statistisch noch 2,21 Jahre oder bis ehe Zürich den alten Wert erreicht. Es wäre das vorläufige einer Geschichte von Niedergang und Wiederaufstieg.

## Der höchste Wert seit 1947

Statistisch gesichert ist, dass die Stadt ein weiteres Mal kräftig gewachsen und rund 7628 Personen. Es ist seit 1947 (1916)

# Eine halbe Million bis 2035

Gemäss Prognosen wächst die Stadt bevölkerungsmässig am stärksten in Zürich-Nord. Und in der Altersgruppe der Teenager.

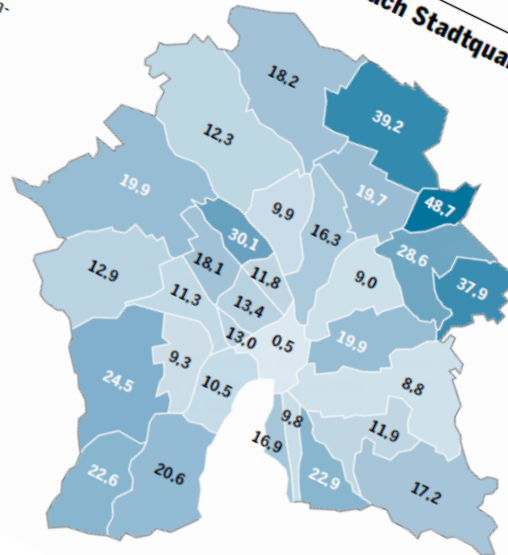
Yannik Wiget

Die Bevölkerung der Stadt Zürich wächst nach den neuesten Berechnungen des statistischen Amtes in den kommenden Jahren um fast 80 000 Personen. Im Jahr 2021 wird sie den historischen Höchststand von 440 180 Einwohnern überschreiten. Und es wird eine halbe Million Menschen leben.

Die Quartiere werden mit einem Anstieg gerechnet. Erwartete Wachstum der Einwohnerzahl bis ins Jahr 2035. Auch im Hirzenbach Quartier (30 Prozent) wird ein starkes Bevölkerungswachstum zu erwarten sein.

Veränderung in Prozent

## Bevölkerungswachstum nach Stadtquartier 2017-2035



etwa 17 Prozent



## Der Tannenbaum

Die Stadt wächst aber nicht nur in der Bevölkerung. Die Altersstruktur ändert sich auch. Die Altersgruppe der Teenager wird sich von 1916 bis 2035 verdreifachen.

Kreis 4

Kreis 5

# Inhalt

1. Zielgrößen und Modellierungsansatz
2. Workflow und Implementierung mit 
3. Resultate Szenarien 2018
4. Sensitivitätsanalysen mit 

# 1. Zielgrößen und Modellierungsansatz

# Modellierte Grössen, raum-zeitliche Auflösung

- Grösse der «Wirtschaftlichen Bevölkerung» in Stadt Zürich, unterschieden nach
  - Altersjahr
  - Geschlecht
  - vier Nationengruppen
    - Schweiz, Deutschland, Österreich, Liechtenstein
    - restliches Europa, restliche Welt
- Bevölkerungsbewegungen (Geburten, Todesfälle, Zuzüge, Wegzüge, ...) für dieselben Gruppen
- Räumliche Auflösung: 31 Quartiere
- Szenarienperiode: 2018 – 2035, Zeitschritt 1 Jahr

# Modell für Bevölkerungsszenarien

## Wohnungs- Modell

Wohnungsbestand →  
Kapazitätsgrenze für Bevölkerungsentwicklung

## Demografie- Modell

Bevölkerungsbestand im Jahr  $t$

= Bevölkerungsbestand im Jahr  $t-1$

+/- Einbürgerung

+/- Umzug innerhalb der Stadt

+ Geburten

- Sterbefälle

+ Zuzug

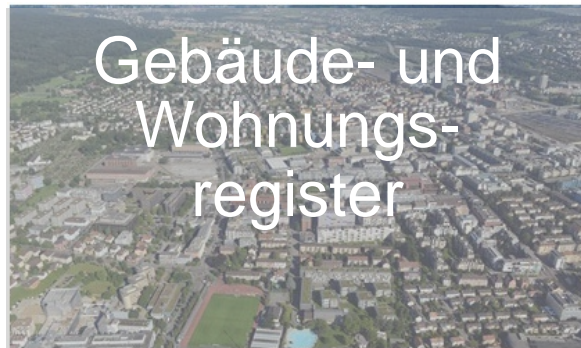
- Wegzug

} system-  
dynamische  
Prozesse

} dynamische  
Prozesse

## 2. Workflow und Implementierung mit R

# Workflow Wohnungsmodell



heute:  
aktueller Wohnungsbestand

kurzfristige Zukunft:  
bekannte Bauvorhaben

langfristige Zukunft:  
Flächenreserven



- Belegungsquote (# Personen pro Wohnung)
- Wohnfläche (m<sup>2</sup>) pro Person
- Anzahl Wohnungen
- Geschossflächen (m<sup>2</sup>)

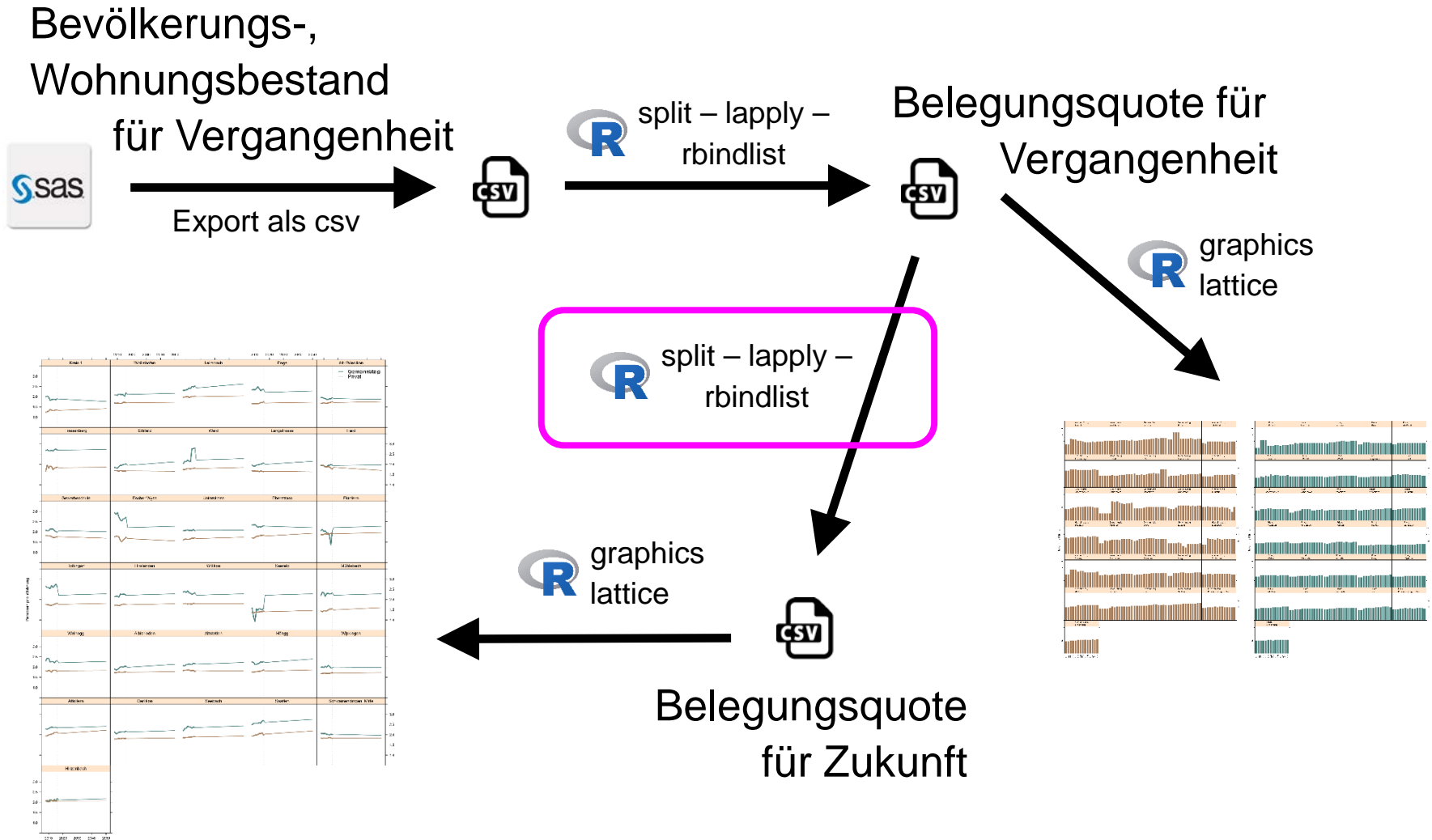
**maximale Bevölkerungsgrösse  
für jedes Jahr von 2018 – 2035**



# Workflow Wohnungsmodell

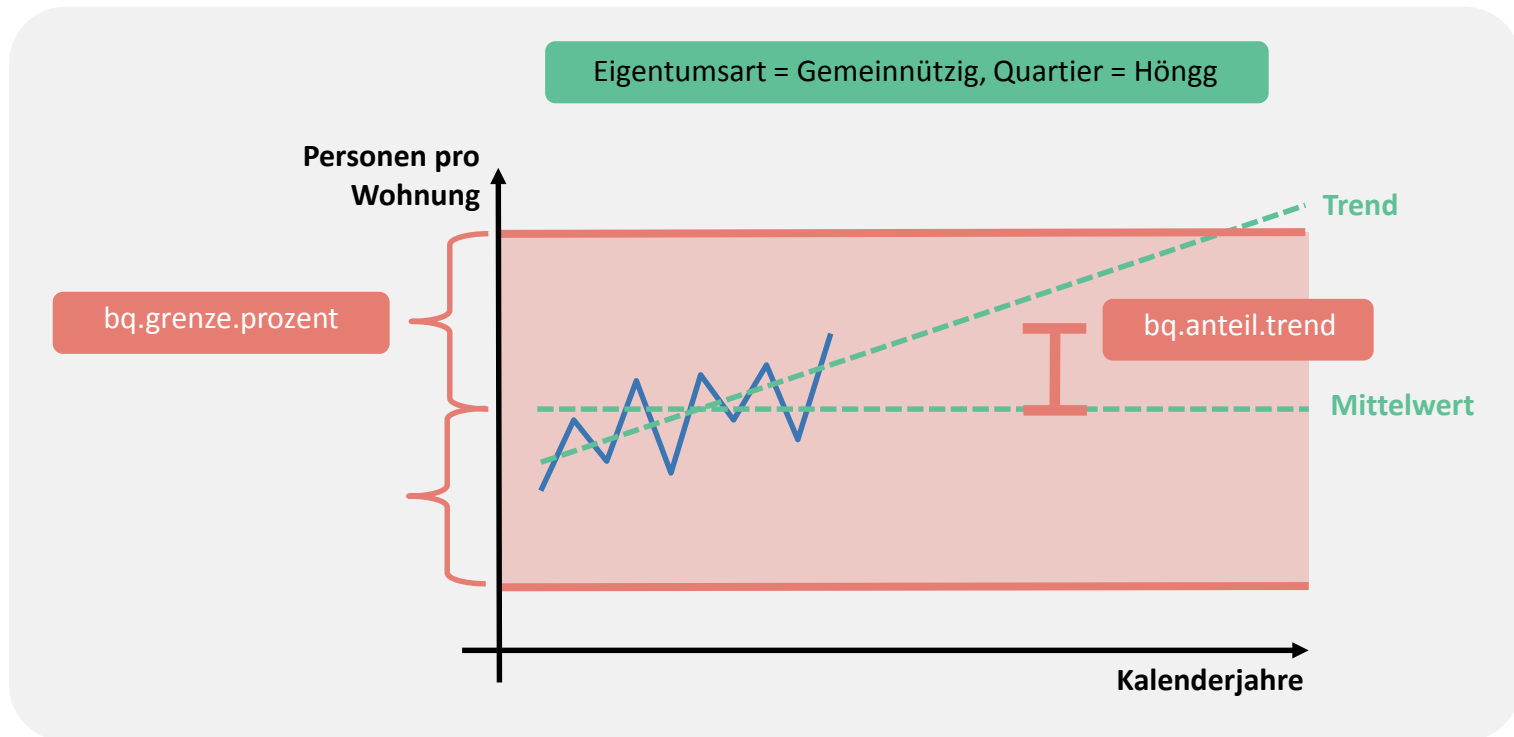


# Wohnungsmodell: Modellierung Belegungsquote



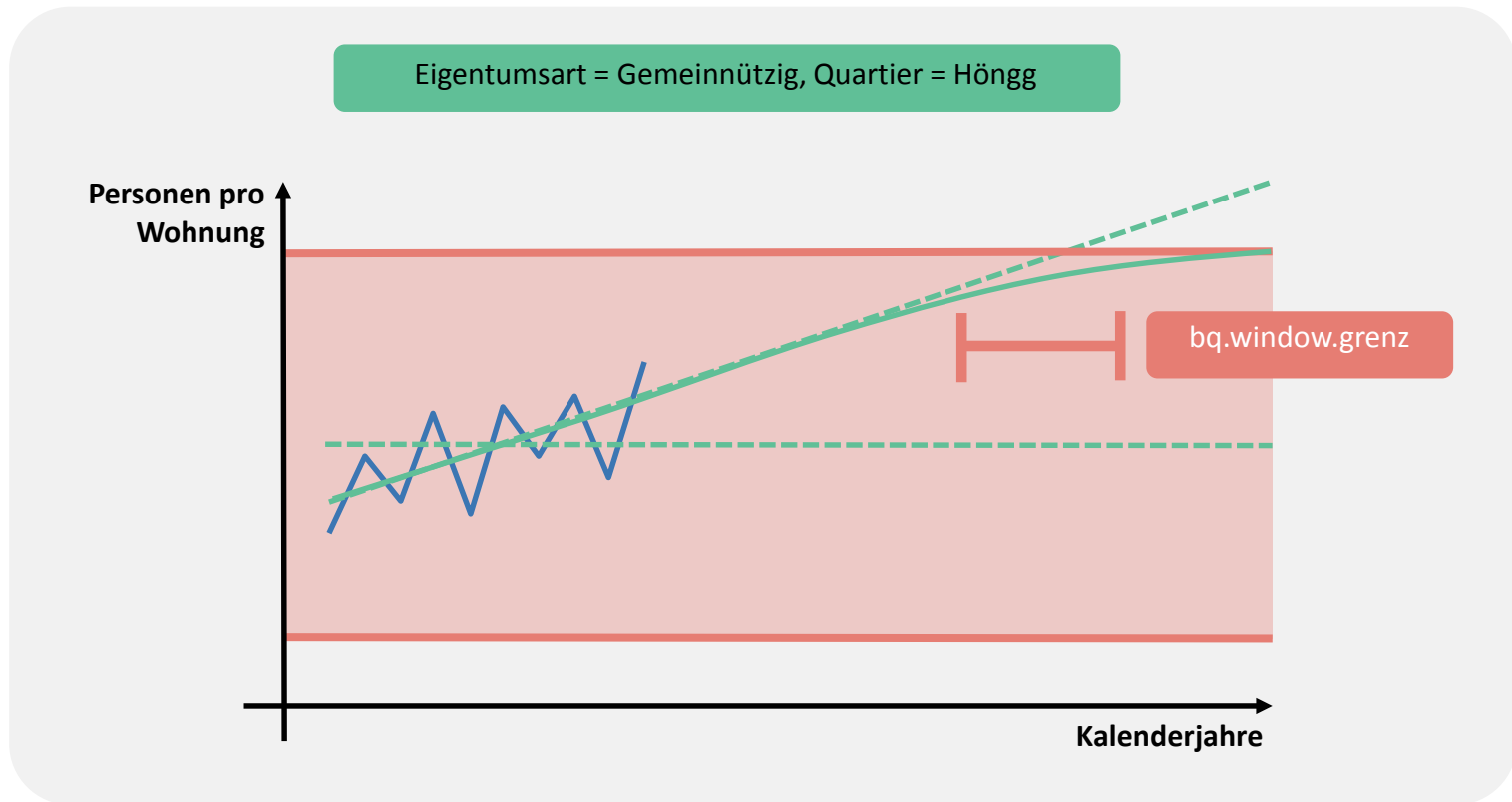
# Modellierung Zeittrend Belegungsquote

gewichtetes Mittel von Mittelwert und linearem Trend



# Modellierung Zeittrend Belegungsquote

## Knickpunkt vermeiden



# split, lapply, rbindlist

```
1  # Faktor zur Kodierung Kombinationen Eigentumsart x Quartier
2
3  sf <- with(data.in,
4    factor(
5      paste(eigart, quart),
6      levels = paste(
7        rep(lookup.eigart$eigart, each = n.quart),
8        lookup.quart$quart)))
9
10 # Input-Liste mit Dataframes fuer jede Kombination von Quartier und Altersjahr
11
12 data.in.list <- split(data.in, sf)
13
14 # Schaetzung der zeitlichen Veranderung der Belegungsquote ueber Basisjahre
15 # und Prognose der Belegungsquote fuer Szenarienjahre
16 # für jede Kombination von Quartier und Altersjahr
17
18 data.out.list <- lapply(data.in.list,
19   fun.prognose.rate.szen,
20   formula      = bq ~ jahr,
21   anteil.trend = bq.anteil.trend,
22   grenze.prozent = bq.grenze.prozent,
23   window.grenz = bq.window.szen,
24   lmfun        = lmrob)
25
26 # Output-Liste in Dataframe umwandeln
27
28 data.out <- as.data.frame(rbindlist(data.out.list))
```



## split, mclapply, rbindlist (parallelisiert)

```
1  # Faktor zur Kodierung Kombinationen Eigentumsart x Quartier
2
3  sf <- with(data.in,
4    factor(
5      paste(eigart, quart),
6      levels = paste(
7        rep(lookup.eigart$eigart, each = n.quart),
8        lookup.quart$quart)))
9
10 # Input-Liste mit Dataframes fuer jede Kombination von Quartier und Altersjahr
11
12 data.in.list <- split(data.in, sf)
13
14 # Schaetzung der zeitlichen Veranderung der Belegungsquote ueber Basisjahre
15 # und Prognose der Belegungsquote fuer Szenarienjahre
16 # für jede Kombination von Quartier und Altersjahr
17
18 data.out.list <- mclapply(data.in.list,
19   fun.prognose.rate.szen,
20   formula      = bq ~ jahr,
21   anteil.trend = bq.anteil.trend,
22   grenze.prozent = bq.grenze.prozent,
23   window.grenz = bq.window.szen,
24   lmfun        = lmrob, mc.cores = detectCores())
25
26 # Output-Liste in Dataframe umwandeln
27
28 data.out <- as.data.frame(rbindlist(data.out.list))
```

# Workflow Demografiemodell



Bestand und Bewegungen  
aktuell und in Vergangenheit



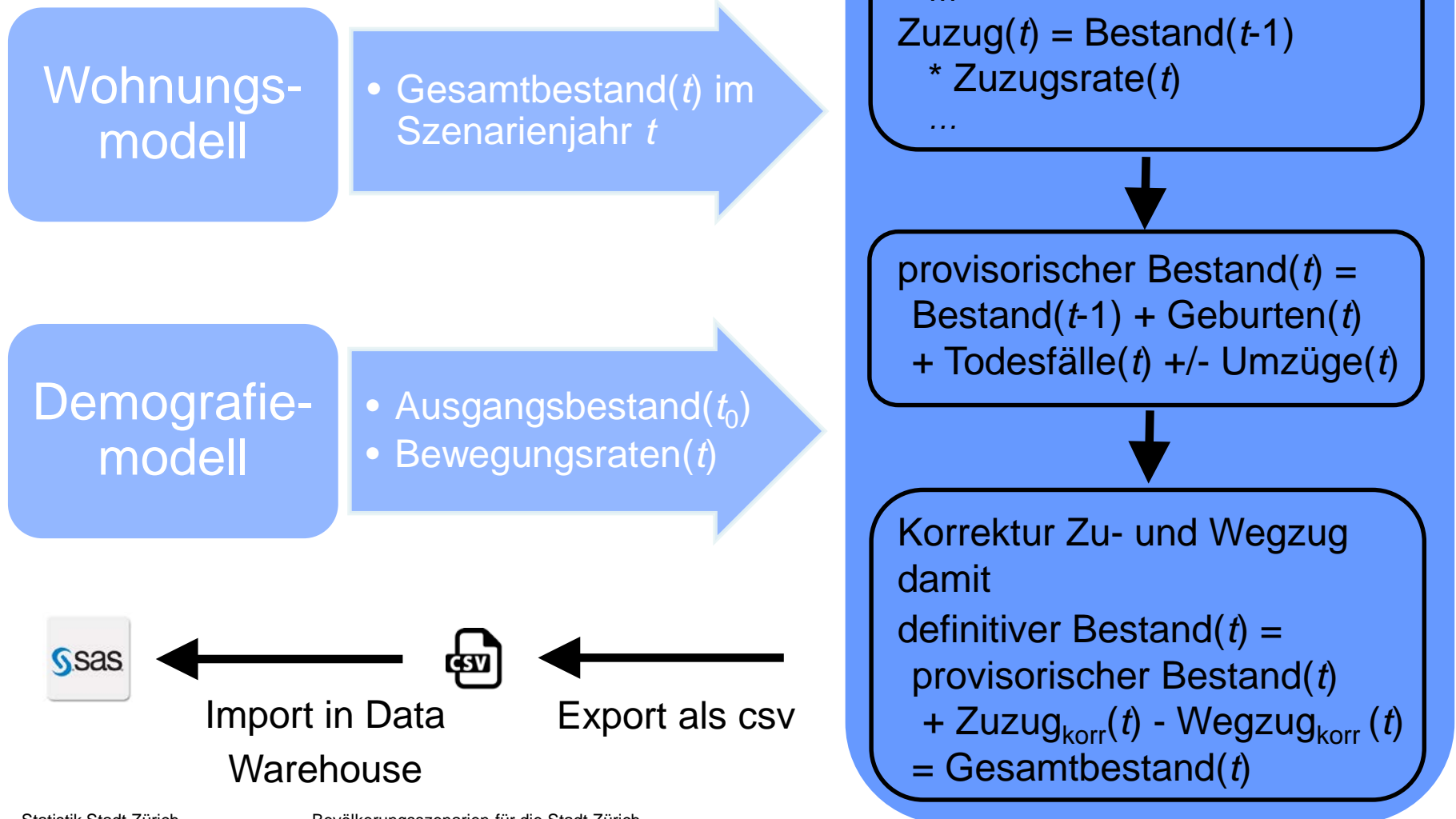
Sterbewahrscheinlichkeit  
aktuell und in Vergangenheit



- Bewegungsraten (Geburten, Todesfälle, Zu-, Wegzüge,..) für Vergangenheit und Zukunft
- zukünftiger Bestand und Bewegungen

**Bevölkerungsgrösse und -zusammensetzung  
für jedes Jahr von 2018 – 2035**

# Workflow Gesamtmodell





# Szenarien «Unten», «Mitte», «Oben»: Annahmen über künftige Nutzung des Wohnraums in Stadt

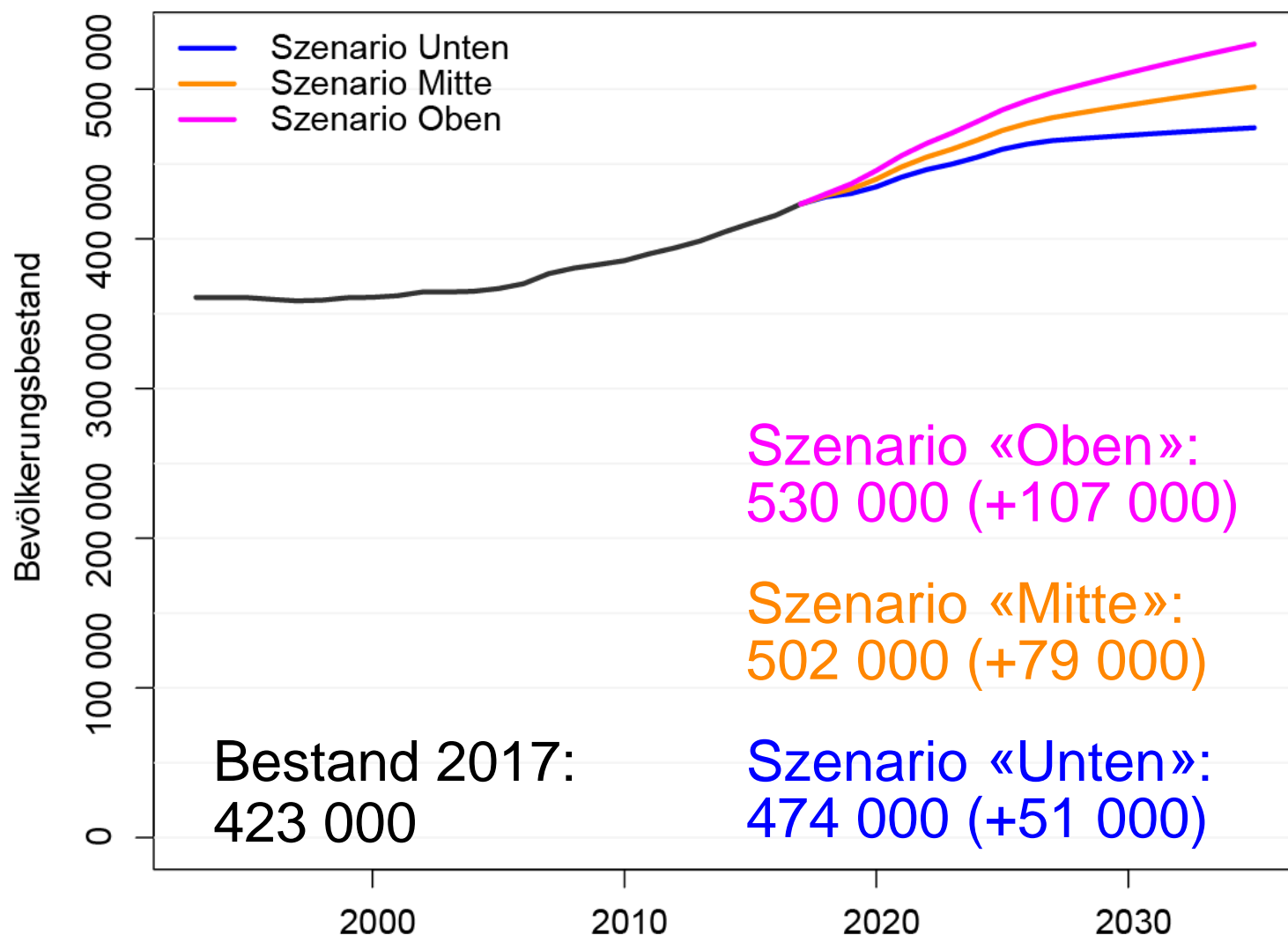
- Unterschiedliche Annahmen über künftigen
  - Ausbaugrad, Wohnanteil von Parzellen und Anteil Arealüberbauungen;
  - Anzahl Personen pro Wohnung und Wohnfläche pro Person
  - Annahmen kodiert als «Stellschrauben» des Modells

Wichtigste Parameter	«Unten»	«Mitte»	«Oben»
Ausbaugrad	75%	85%	90%
Wohnanteil (minimal vs. real vs. maximal)	-25%	0%	+25%
Arealüberbauungen (Anteil mit vs. ohne)	0%	50%	100%
Wohnflächenverbrauch (Anteil Trend)	0%	20%	50%
Belegungsquote (Anteil Trend)	0%	20%	50%

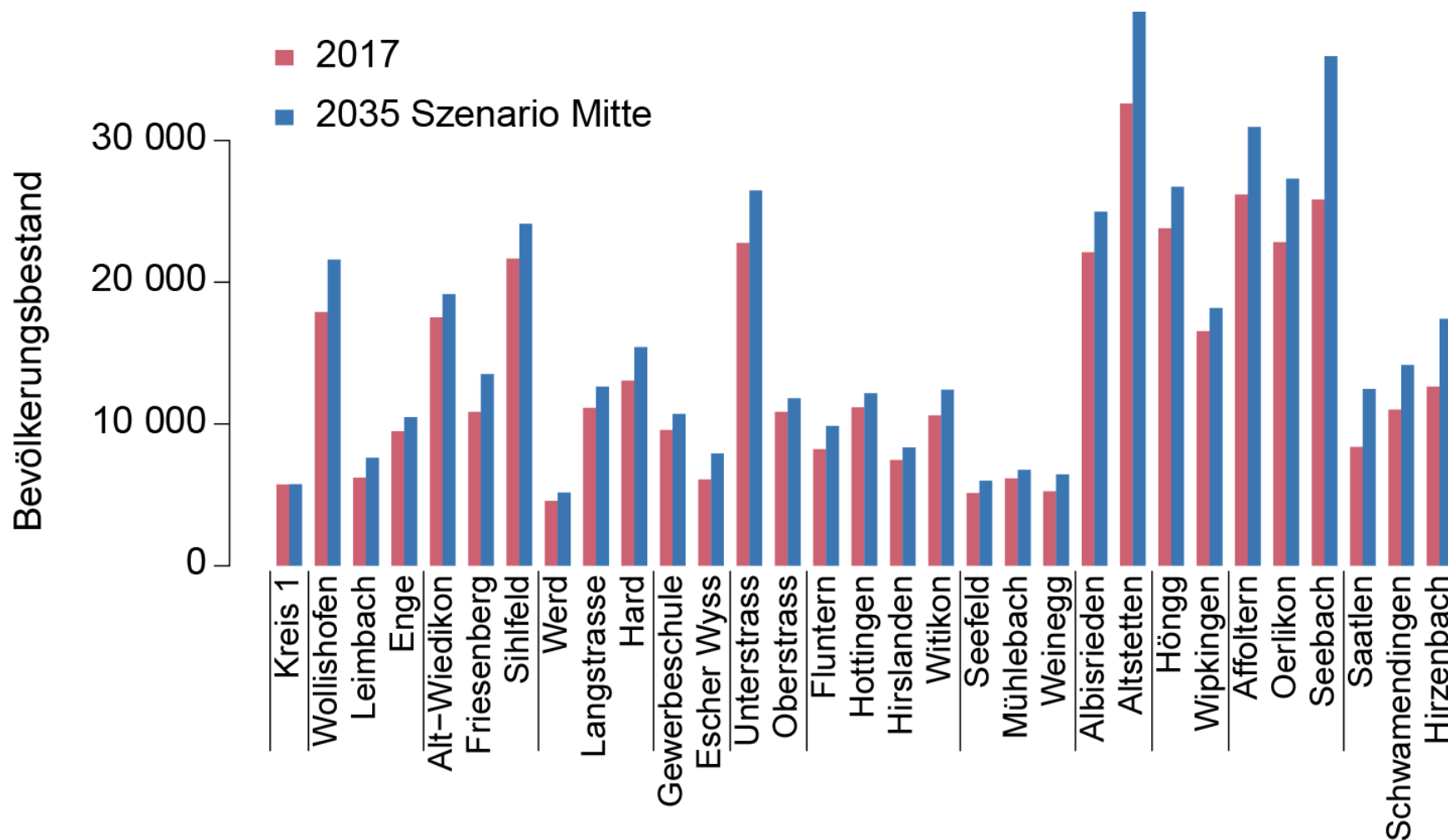
- Demographiemodell identisch in allen 3 Szenarien

# 3. Resultate Szenarien 2018

# Bevölkerungsentwicklung 2017 – 2035

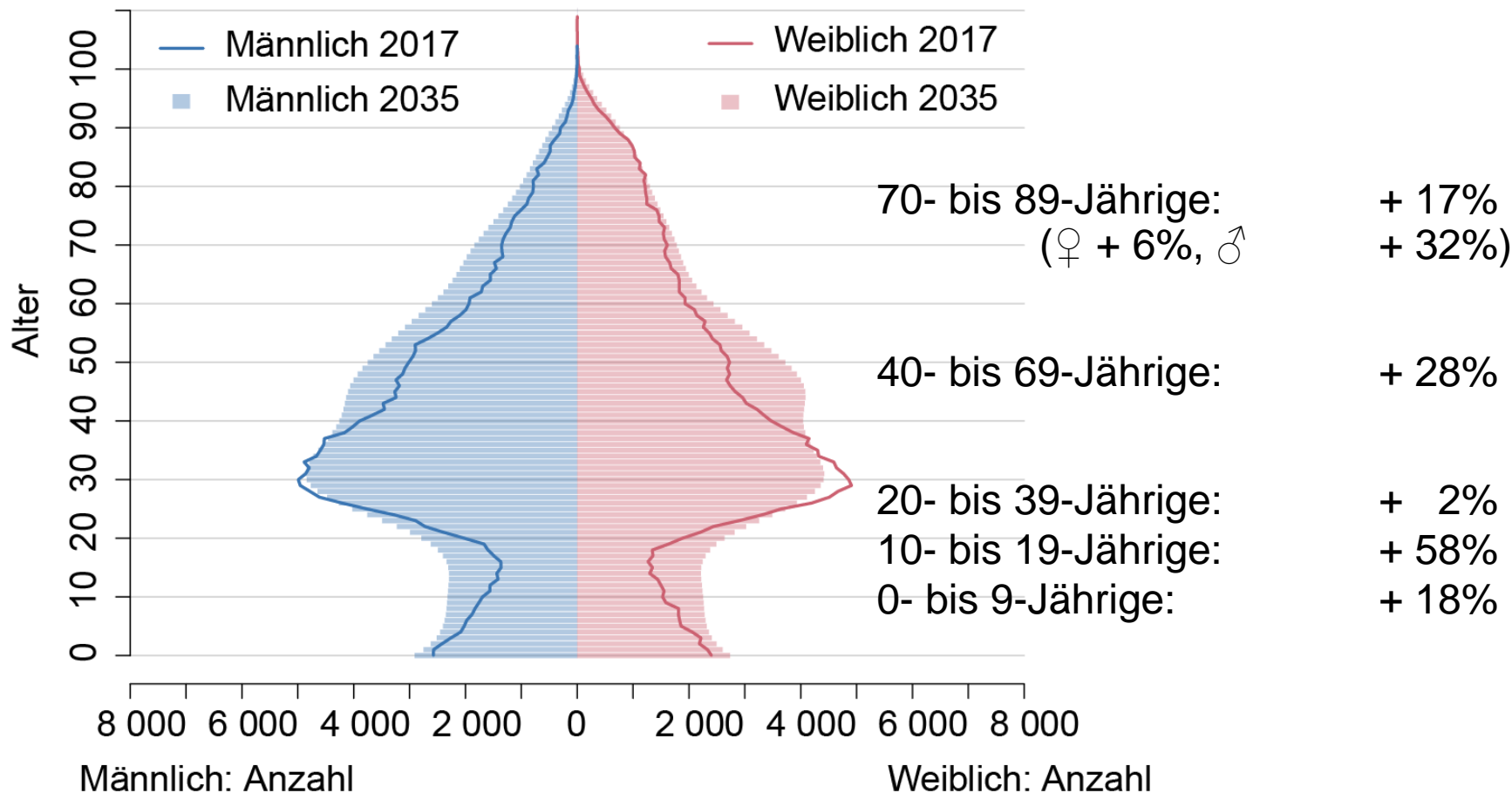


# Bevölkerung in den Quartieren 2017 und 2035



# Veränderung der Altersverteilung zwischen 2017 und 2035

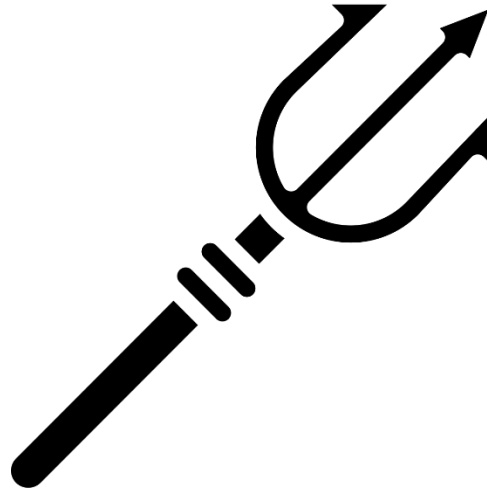
## Szenario Mitte



## 4. Sensitivitätsanalysen mit R

# Können wir das Modell verbessern?

- Nach der Veröffentlichung ist vor der Veröffentlichung
- Modell anpassen/verbessern/verfeinern
  - **Kunden**: Wer verwendet die Modellergebnisse? Akzeptanz?
  - **Validierung**: Stimmt die Szenarien in der Vergangenheit?
  - **Sensitivitäten**: Welche Parameter sind sensitiv?



# Kunden: Akzeptanz erarbeiten



## Akzeptanz bei Politik

- Jährliches Vorstellen der Bevölkerungsszenarien im **Stadtrat**
- Präsentationen bei **Geschäftsleitungen**



## Akzeptanz bei Fachpersonen

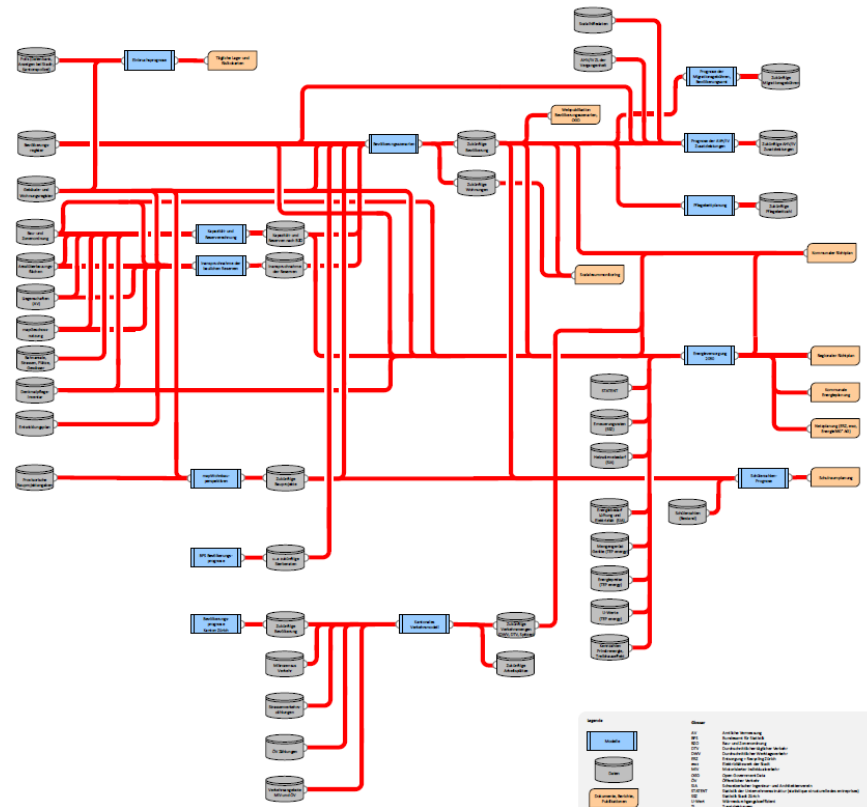
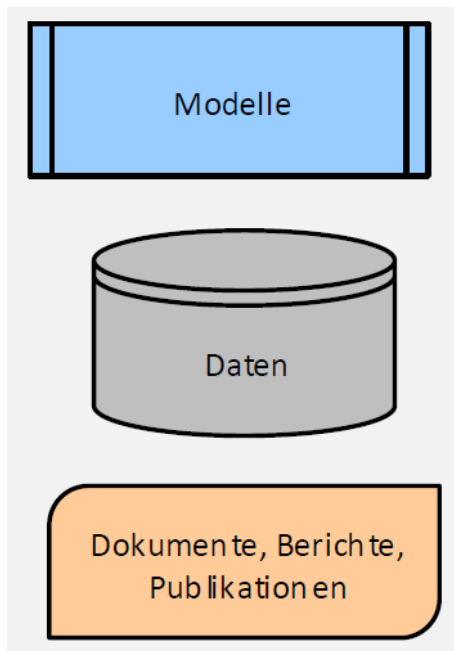
- **Schlüsselkunden** identifiziert (Nachfrage)
- Modellstruktur gemäss **Nachfrage programmiert**
- Modellstruktur grundsätzlich **gleich** halten
- **Dokumentation** öffentlich/verständlich



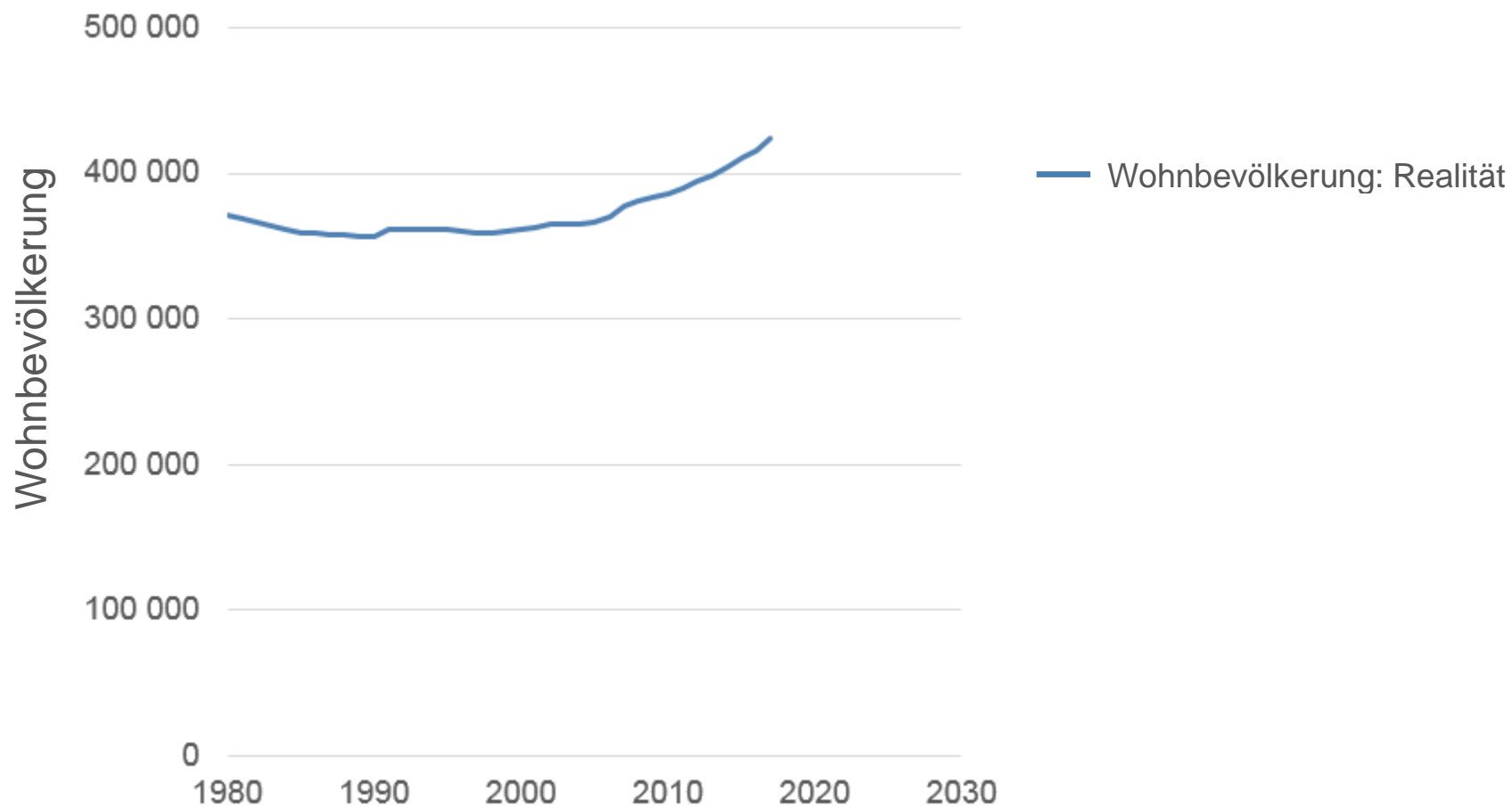
# Kunden: Prognoselandschaft



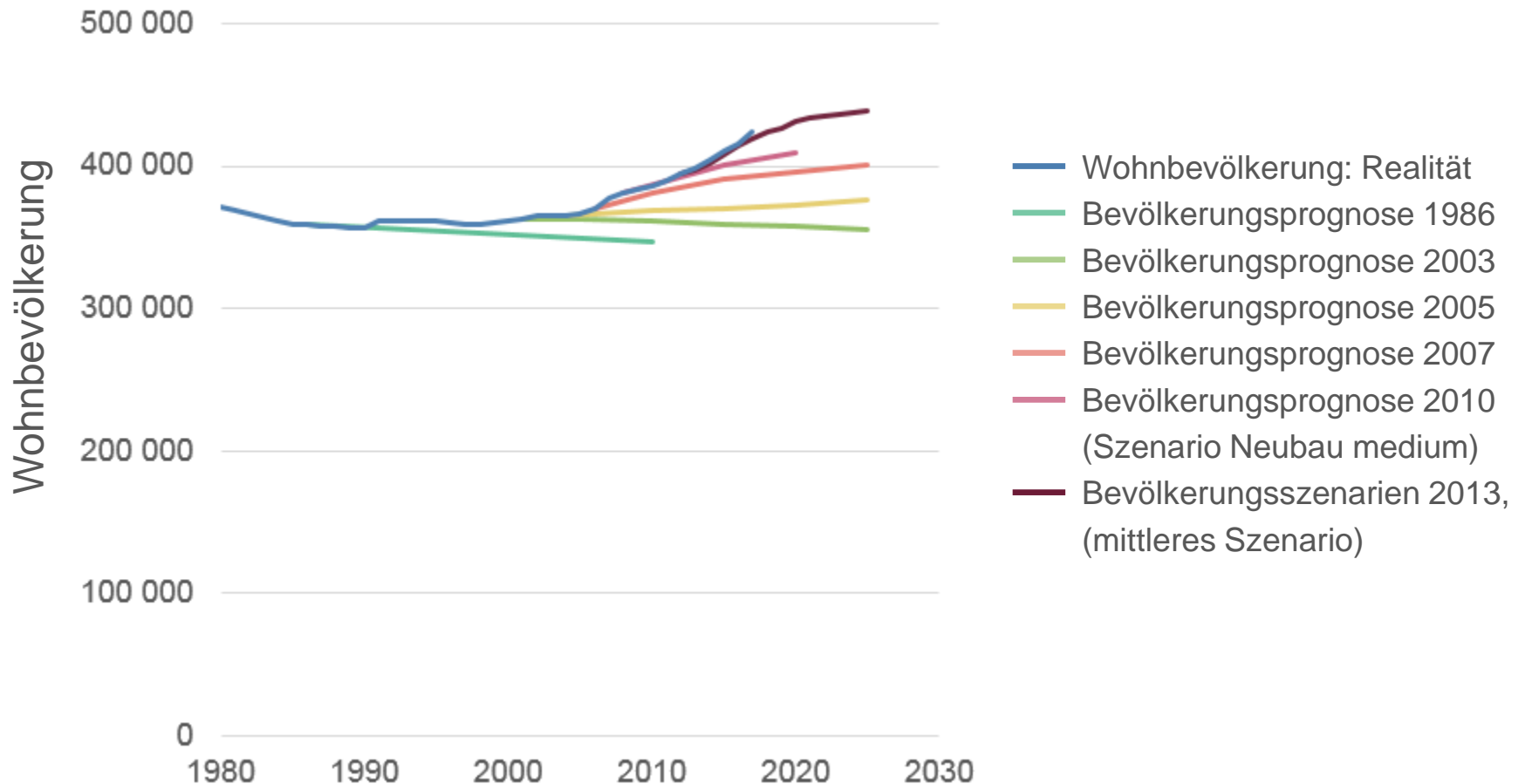
- Wo werden die Modellergebnisse verwendet?
- Abhängigkeiten/Zirkelschlüsse?



# Validierung



# Validierung



→ Wo stimmten die Szenarien nicht? Warum?

- Parameter und Werte sind veröffentlicht ([Link](#))
- Regionale Sensitivitätsanalyse ([sensitivity package](#)):  
pro Parameter ein Sensitivitätswert

The diagram consists of two parts. The top part shows a horizontal yellow beam balanced on a central orange triangular fulcrum. The left end of the beam is labeled 'Parameter' and the right end is labeled 'Output'. The bottom part shows the same beam tilted upwards on the right side. A green arrow points downwards at the 'Parameter' end, indicating an input force. The 'Output' end is higher than the 'Parameter' end.

# Sensitivity package

- Zielfunktion wählen (z.B. Bevölkerung im Jahr 2030)
- Funktionsinput: Parameter-Matrix  
(verschiedene Parameter-Kombinationen)
- Funktionsoutput: Werte für Zielfunktion  
(1 Wert pro Parameter-Kombination)

```
#-----  
#regional sensitivity analysis: Sobol  
#-----  
  
#model input: matrix with n parameter combinations  
#model output: n values for an objective function  
  
model.reg <- function(X)  
{  
  ninx <- nrow(X)  
  outreg <- rep(NA, ninx)  
  for (w in 1:ninx) {outreg[w] <- do.call(model.nse, list(as.numeric(X[w,])))}  
  outreg  
}
```

# Sensitivity package

- Parameter-Matrizen erstellen
- Verschiedene Methoden (siehe Saltelli et al., 2008)
- Output: Index 1. Ordnung, gesamter Index

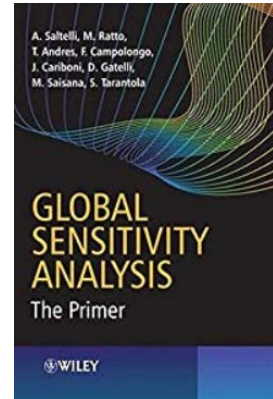
```
#two matrices with (different) random values

np <- nrow(para) #amount of parameters
nk <- 500 #amount of parameter combinations

pa <- data.frame(matrix(NA, nrow = nk, ncol = np))
names(pa) <- para[,1]
a.pa <- b.pa <- pa
for (i in 1:np){a.pa[,i] <- runif(nk, para$lower[i], para$upper[i])}
for (i in 1:np){b.pa[,i] <- runif(nk, para$lower[i], para$upper[i])}
head(a.pa)
head(b.pa)
#model.reg(a.pa[1:5,])

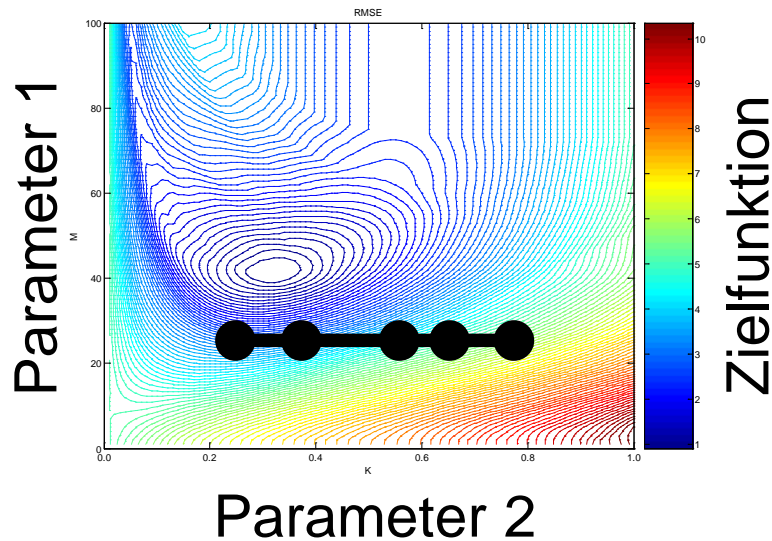
#sobol sensitivity indices
t0 <- Sys.time()
sobol <- sobol2002(model = model.reg, x1 = a.pa, x2 = b.pa, nboot = 10000)
Sys.time() - t0

#1st order index
sobol$s
```



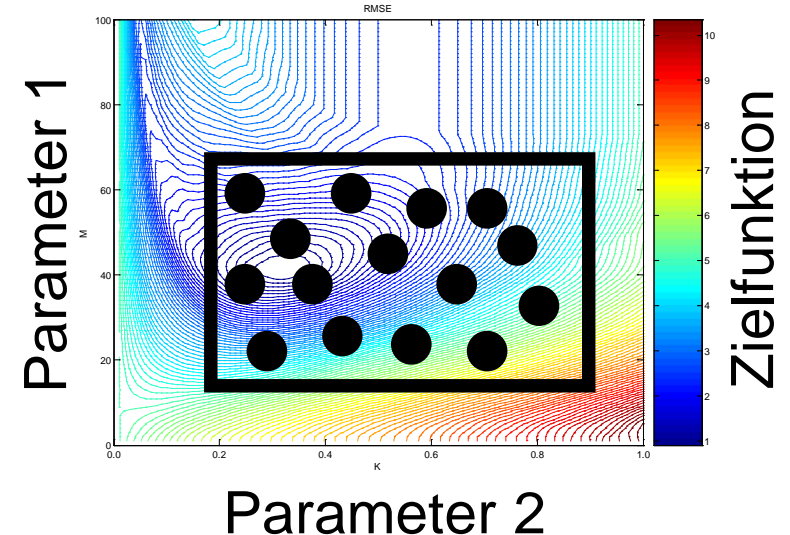
# Exkurs: Regionale Sensitivitätsanalyse

## Lokale Sensitivitätsanalyse (LSA)



«es wird jeweils nur ein Parameter geändert»

## Regionale Sensitivitätsanalyse (RSA)



«es werden gleichzeitig mehrere Parameter geändert» → Parameter-Region

# Exkurs: Varianz-basierte Sensitivitätsindices

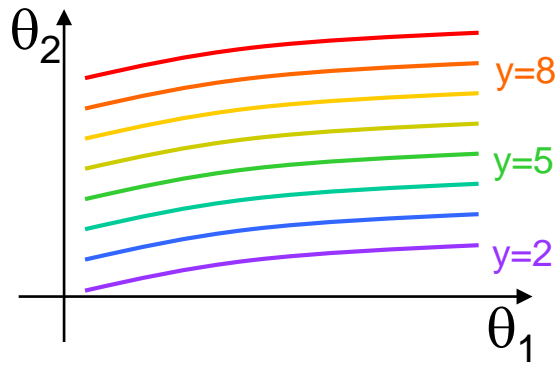
- Varianz-basierte Sensitivitätsindices  
(nach Saltelli et al., 2008<sub>1</sub>)
- Index 1. Ordnung:
  - $S_i$ : Sensitivitätsindex 1. Ordnung des i-ten Parameters
  - $\theta_i$ : Parameter i
  - $y$ : Zielfunktion (z.B. Wohnbevölkerung im Jahr 2030)

$$S_i = \frac{V(E[y|\theta_i = c])}{V(y)}$$



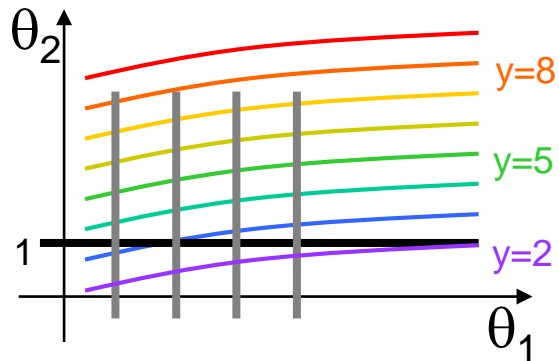
# Exkurs: Varianz-basierte Sensitivitätsindices

Modell A ( $\theta_2$  ist sensitiv)



# Exkurs: Varianz-basierte Sensitivitätsindices

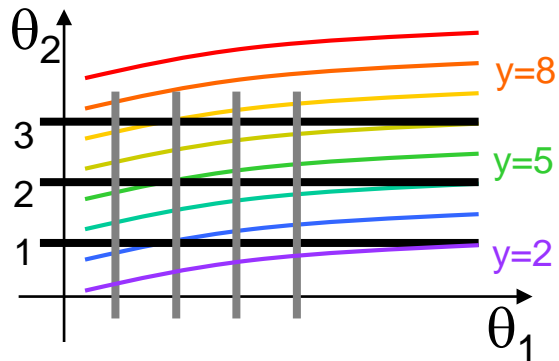
Modell A ( $\theta_2$  ist sensitiv)



$$E[y|\theta_2 = 1] = \frac{1}{4}(3.2 + 2.9 + 2.7 + 2.4) = 2.8$$

# Exkurs: Varianz-basierte Sensitivitätsindices

Modell A ( $\theta_2$  ist sensitiv)



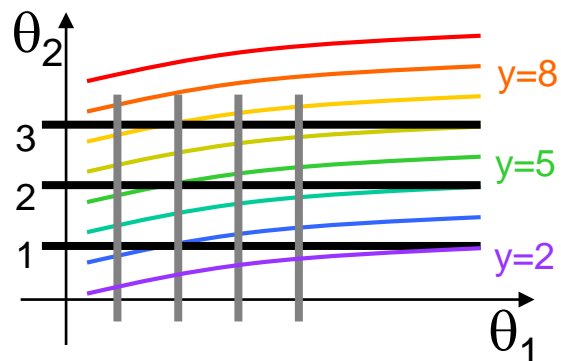
$$E[y|\theta_2 = 1] = \frac{1}{4}(3.2 + 2.9 + 2.7 + 2.4) = 2.8$$

$$E[y|\theta_2 = 2] = \frac{1}{4}(5.2 + 5.0 + 4.8 + 4.5) \approx 4.9$$

$$E[y|\theta_2 = 3] = \frac{1}{4}(7.4 + 7.0 + 6.6 + 6.4) \approx 6.9$$

# Exkurs: Varianz-basierte Sensitivitätsindices

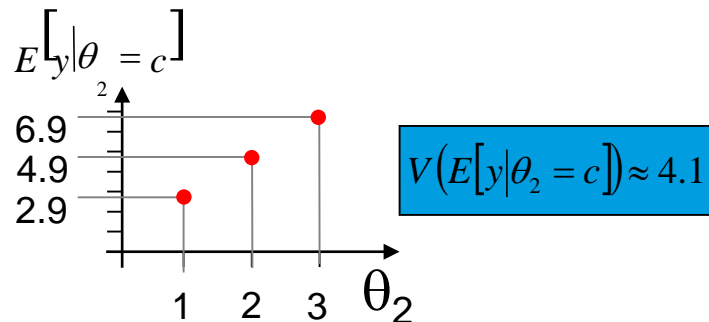
Modell A ( $\theta_2$  ist sensitiv)



$$E[y|\theta_2 = 1] = \frac{1}{4}(3.2 + 2.9 + 2.7 + 2.4) = 2.8$$

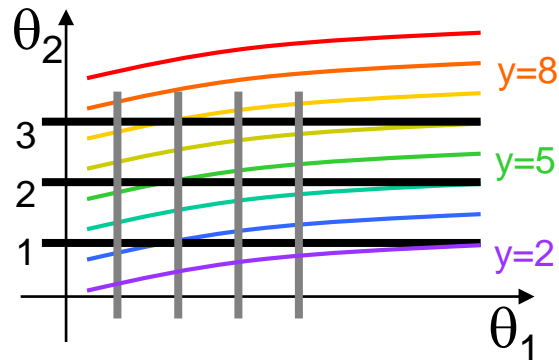
$$E[y|\theta_2 = 2] = \frac{1}{4}(5.2 + 5.0 + 4.8 + 4.5) \approx 4.9$$

$$E[y|\theta_2 = 3] = \frac{1}{4}(7.4 + 7.0 + 6.6 + 6.4) \approx 6.9$$



# Exkurs: Varianz-basierte Sensitivitätsindices

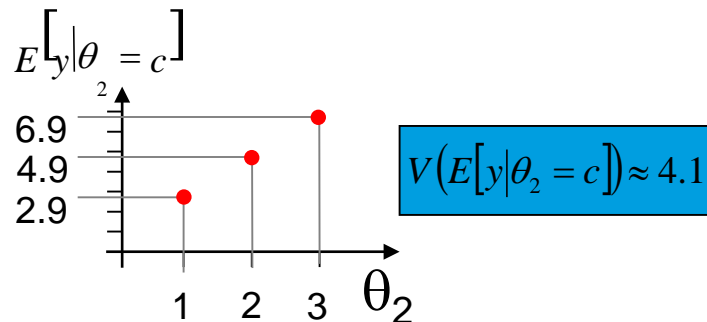
## Modell A ( $\theta_2$ ist sensitiv)



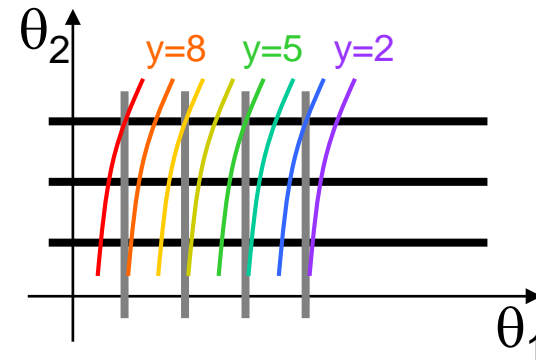
$$E[y|\theta_2 = 1] = \frac{1}{4}(3.2 + 2.9 + 2.7 + 2.4) = 2.8$$

$$E[y|\theta_2 = 2] = \frac{1}{4}(5.2 + 5.0 + 4.8 + 4.5) \approx 4.9$$

$$E[y|\theta_2 = 3] = \frac{1}{4}(7.4 + 7.0 + 6.6 + 6.4) \approx 6.9$$



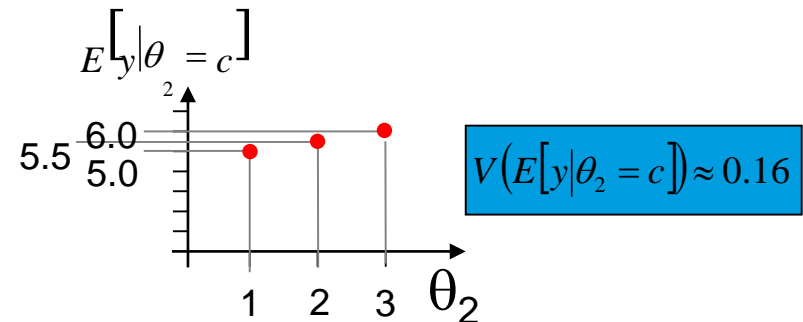
## Modell B ( $\theta_2$ ist nicht sensitiv)



$$E[y|\theta_2 = 1] = \frac{1}{4}(8.2 + 6.2 + 4.2 + 2.2) = 5.2$$

$$E[y|\theta_2 = 2] = \frac{1}{4}(8.5 + 6.5 + 4.5 + 2.5) = 5.5$$

$$E[y|\theta_2 = 3] = \frac{1}{4}(9.0 + 7.0 + 5.0 + 3.0) = 6.0$$



# Exkurs: Varianz-basierte Sensitivitätsindices

- Varianz-basierte Sensitivitätsindices  
(nach Saltelli et al., 2008<sub>1</sub>)
- Index 1. Ordnung:
  - $S_i$ : Sensitivitätsindex 1. Ordnung des i-ten Parameters
  - $\theta_i$ : Parameter i
  - $y$ : Zielfunktion (z.B. Wohnbevölkerung im Jahr 2030)

$$S_i = \frac{V(E[y|\theta_i = c])}{V(y)}$$

# Exkurs: Varianz-basierte Sensitivitätsindices

- Analog dazu: Beispiel für Index **2. Ordnung**  
(Sensitivitätsindex 2. Ordnung für Parameter 2 und 3)

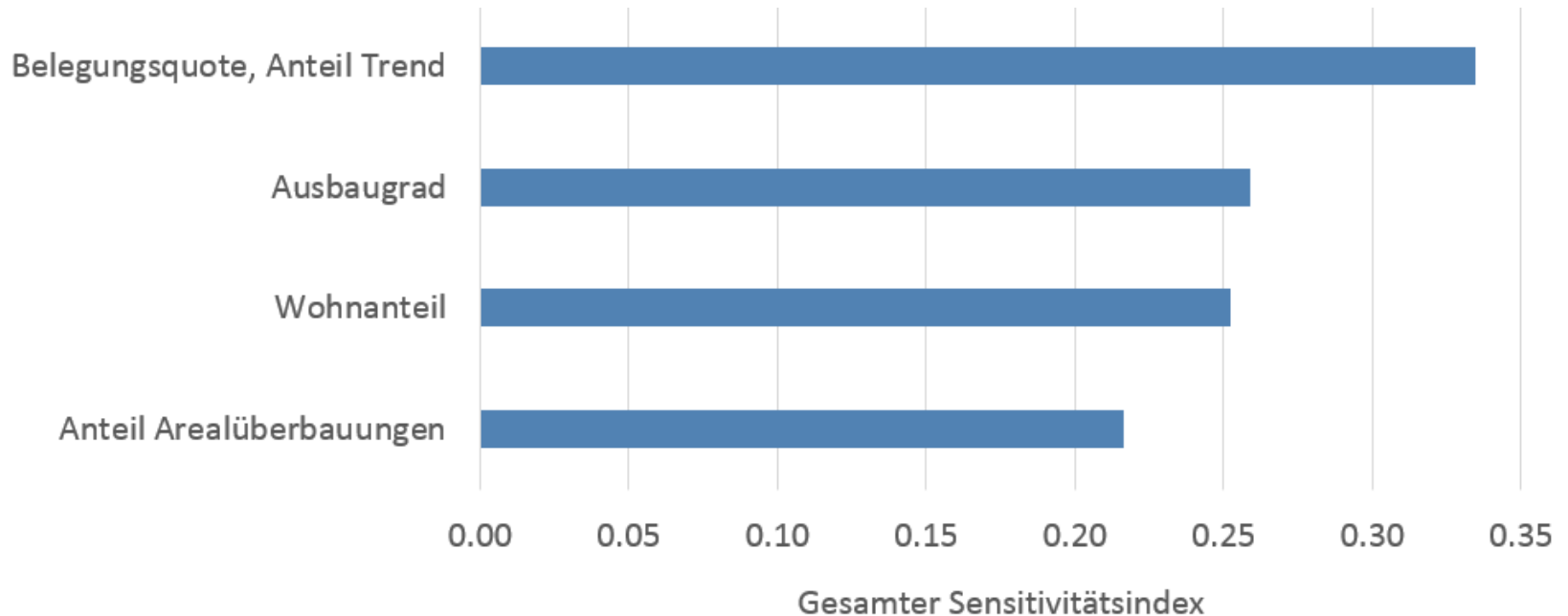
$$S_{23} = \frac{V(E[y | (\theta_2 = c \cap \theta_3 = c)])}{V(y)}$$

- Gesamter Sensitivitätsindex: Indices aller Ordnungen in denen der Parameter vorkommt

$$T_2 = S_2 + S_{12} + S_{23} + S_{24} + S_{234}$$

# Sensitivitätsanalyse: Ergebnisse

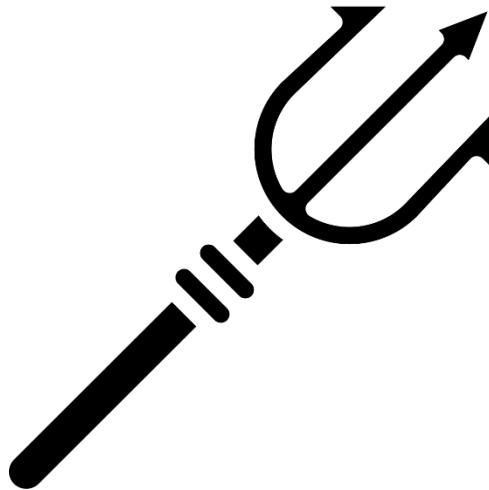
- Regionale Sensitivitätsanalyse für Zielfunktion Bevölkerungsbestand im Jahr 2030
- Sensitivste Parameter





# Erkenntnisse für Modellverbesserung

- Belegungsquote und Ausbaugrad sind sensitiv (**Sensitivitätsanalyse**)
- Nutzen bei Modellüberprüfung (**Validierung**)
- Vergleich mit Nachfrage (**Kundenperspektive**)



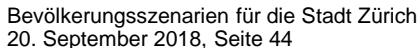
→ Detaillierte Datenanalysen, Priorisierung

# Besten Dank!

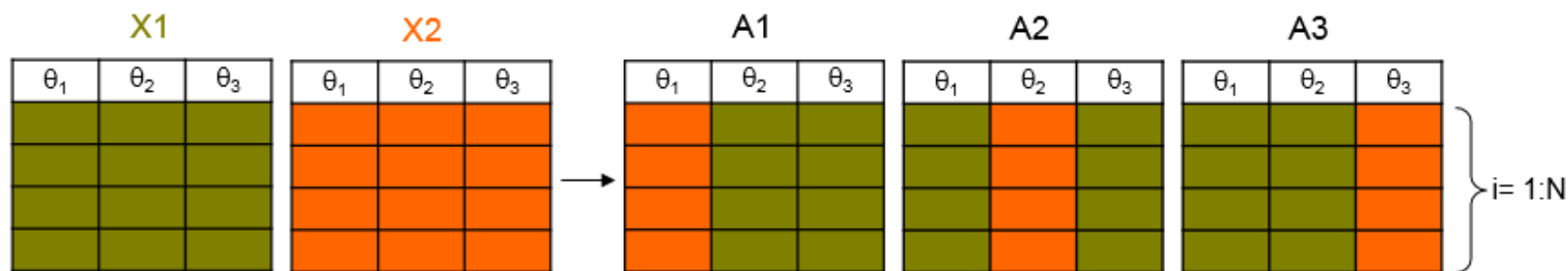


# Bevölkerung: Definitionen

	Wirtschaftlich	Zivilrechtlich	Ständig	
Schweizer/-innen				
Jahresaufenthalter/-innen, Niedergelassene				
Personen mit EDA-Ausweis				
Kurzaufenthalter/-innen				Aufenthalt $\geq$ 12 Monate
Kurzaufenthalter/-innen				Aufenthalt $<$ 12 Monate
Wochenaufenthalter/-innen				in der Stadt Zürich
Wochenaufenthalter/-innen				ausserhalb der Stadt Zürich
Personen mit Amtsadresse				
Asylsuchende, vorl. Aufgenommene				Aufenthalt $\geq$ 12 Monate
Asylsuchende, vorl. Aufgenommene				Aufenthalt $<$ 12 Monate



- three parameters; assumption:  $\theta_2$  sensitive,  $\theta_1$  und  $\theta_3$  not sensitive
- two parameter matrices X1 and X2



- result of the model run (objective function)

$RX_1$	$RX_2$	$RA_1$	$RA_2$	$RA_3$
0.71	0.16	0.76	0.15	0.76
0.83	0.35	0.89	0.39	0.84
0.12	0.94	0.14	0.92	0.10
0.25	1.53	0.28	1.59	0.21

- Index = covariance

- $Cov(RX_2, RA_1)$  = small  $\rightarrow$  low index  $\rightarrow$  not a sensitive parameter
- $Cov(RX_2, RA_2)$  = large  $\rightarrow$  high index  $\rightarrow$  sensitive parameter
- $Cov(RX_2, RA_3)$  = small  $\rightarrow$  low index  $\rightarrow$  not a sensitive parameter