BAYESIAN MODEL AVERAGING TO PREDICT INTEREST IN ELECTRIC VEHICLES IN GERMANY

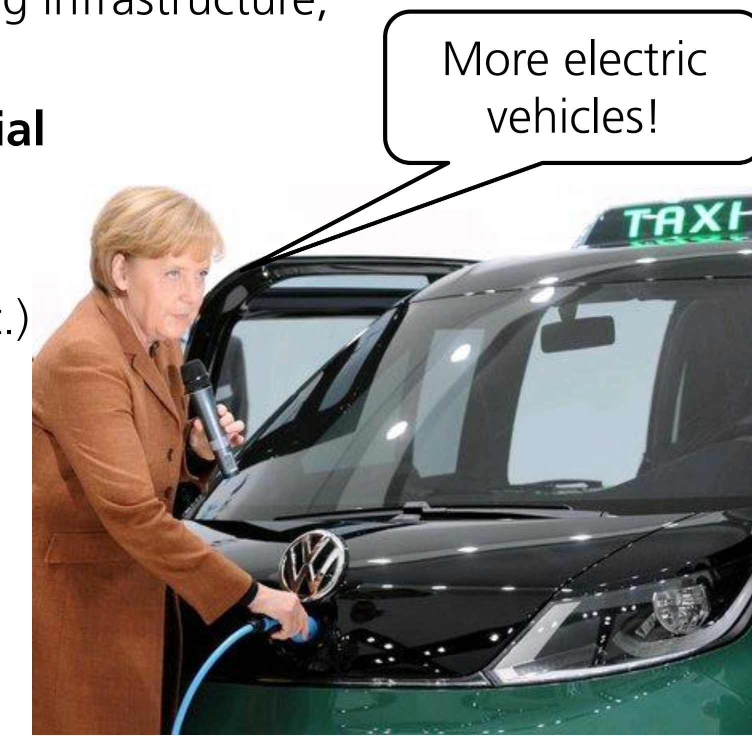
Patrick Plötz, Julius P. Wesche and Elisabeth Dütschke
Fraunhofer Institute for Systems and Innovation Research ISI, Germany

Fraunhofer
ISI

IAEE Conference, Vienna 2017

Identification of the buyer groups is essential for policies & industry

- **Policy makers are widely supporting electric vehicles (EVs)**
  - Policies that are applied differ widely, including CO_2-dependent taxes, subsidies, built-up of public charging infrastructure, pilot projects and many more
  - **Identification of user groups is essential for effective policies & car makers**
  - Many studies on first users (high-income, men, techies, pro-environment, high educ.)
  - **but little on next groups**
  - High uncertainty in influencing factors



- **Aim: Identify the potential early Majority of electric vehicles on a sound statistical basis**

Fraunhofer ISI

Data: representative survey of 1,000 Germans on EV purchase intention

- **Representative survey of 1,017 Germans**
- Variables on
  - **Interest in EV, purchase intention**
  - Symbolic value of cars (on 1 – 7 likert scale):
    - "Using a ICE / EV I can determine my route by myself " (*Autonomy*)
    - "An ICE / EV takes me everywhere" (*Freedom*)
    - Willingness to pay more (WTPM) in % for EV

Variable	**mean**	**SD**
Age	48.3	16.6
HH members	2.4	1.2
Technology affinity	4.4	0.9
Availability ICE	5.2	1.3
Availability EV	3.4	1.7
Independence ICE	5.3	1.2
Independence EV	3.9	1.6
Freedom ICE	5.4	1.0
Freedom EV	3.7	1.6
Autonomy ICE	5.5	1.0
Autonomy EV	4.0	1.6
WTPM	5.3	9.9

- Classification into **adopter groups** following Peters et al. (2011)
  - Participant owns an EV → *Innovator*
  - Participant wants to buy & is interested in EVs → *early adopter*
  - Participant does not want to buy but is interested in EVs → **early majority**

Results: Single logistic model

- Rogers' Adopter groups for EVs in survey:
  - 0.4% innovators own an EV; 1% with purchase intention → early adopter
  - **49% are interested → Majority group → want to understand them**
- Dependent variable: **Interest in EV?** (binary: yes/no)
- Results **of logistic regression model**:

Predictor	Estimate	Std. Error	p-value
(Intercept)	-5.16	0.63	< 2e-16 ***
Technology affinity	0.75	0.09	9.2e-16 ***
WTPM in %	0.06	0.01	2.5e-11 ***
Education	0.20	0.07	0.0020 **
Autonomy EV	0.35	0.05	1.4e-11 ***
Freedom ICE	-0.31	0.08	2.4e-4 ***
Drivers license	1.10	0.28	6.8e-5 ***

- **Problem: many other factors play role & model seems overly confident**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fraunhofer ISI

Method: Bayesian model averaging compensates for over-confidence of simple regression

- BMA to account for **model uncertainty** (significance and size of effect depends on the number of variables chosen for the regression)
- BMA provides a coherent way to integrate model uncertainty (Hoeting et al. 1999)
- Result: **posterior distribution for the coefficient** given data as the average of individual models each weighted by its posterior model probability:

$$\mathrm{pr}(\beta|D) = \sum_{k=1}^{K} \mathrm{pr}(\beta|M_k, D)\, \mathrm{pr}(M_k|D)$$

Probability of coeff's given data

Probability of coeff's given model+data

Probability of model given the data

- Difficult: posterior model probabilities due to integrated likelihood (high-dimensional integral), and the large number of possible models
- We use the BMA package & R statistical software with **2,000 models total**

Fraunhofer ISI

The top ten models have different sizes and lead to different coefficients

- Top ten models by posterior model probability (approx. related to BIC)
- As expected, coefficients vary and single model p-values are over-confident

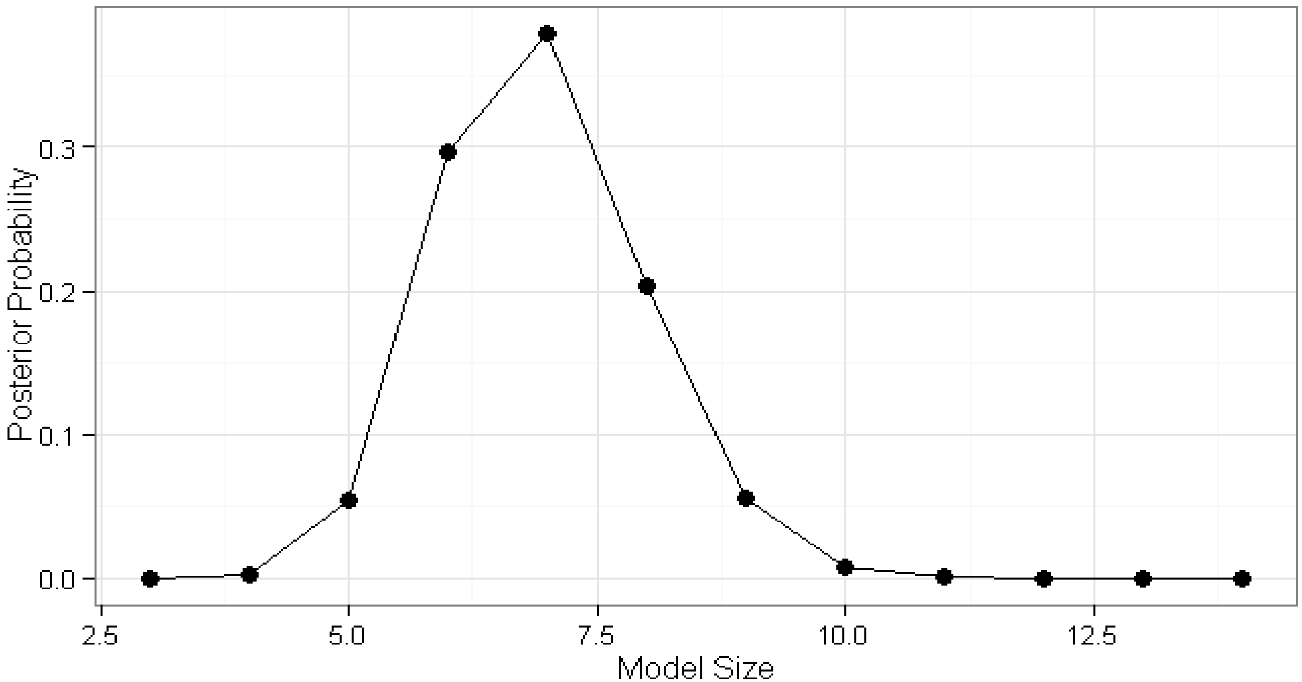
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}
Posterior Probability	13.1%	5.0%	3.8%	3.2%	3.0%	3.0%	2.5%	2.2%	2.1%	1.8%
Intercept	-5.16	-5.49	-5.39	-4.52	-4.89	-4.97	-5.18	-5.19	-5.20	-5.55
Technology affinity	0.75	0.75	0.75	0.77	0.76	0.76	0.74	0.72	0.75	0.73
WTPM	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Autonomy EV	0.35	0.35	0.35	0.35	0.34	0.34	0.36	0.36	0.30	0.35
Drivers license (yes)	1.10	1.26	1.19	1.23	1.32	1.39	0.99	1.08	1.08	1.14
Freedom ICE	-0.31	-0.28	-0.27	-0.31	-0.27	-0.28	-0.43	-0.30	-0.31	-0.42
Education	0.20	0.18	0.18	.	.	.	0.20	0.20	0.19	0.18
Public Transp. (yes)	.	0.36	.	.	.	0.41	.	.	.	0.41
Train available	.	.	0.40	.	0.48	.	.	.	.	.
Availability ICE	.	.	.	.	.	.	0.16	.	.	0.19
Sex (male)	.	.	.	.	.	.	.	0.28	.	.
Availability EV	.	.	.	.	.	.	.	.	0.10	.

- **Bayesian model averaging: weighted average of top 2,000 models!**

The top ten models have different sizes and lead to different coefficients

- Top ten models by posterior model probability (approx. related to BIC)
- As expected, coefficients vary and single model p-values are over-confident

	M_{10}
Posterior	1.8%
Intercept	-5.55
Technolo	0.73
WTPM	0.06
Autonom	0.35
Drivers lic	1.14
Freedom	-0.42
Educatior	0.18
Public Tra	0.41
Train avai	.
Availabili	0.19
Sex (mal	.
Availabili	.



- **Bayesian model averaging: weighted average of top 2,000 models!**

Fraunhofer ISI

Results: Factors influencing interest in EV including model selection uncertainty

- Combined effect from all models weighted by model probability
- Uncertainty much more transparent
- Probability that $\beta \neq 0$
- Mean and SD of coeff after all model runs

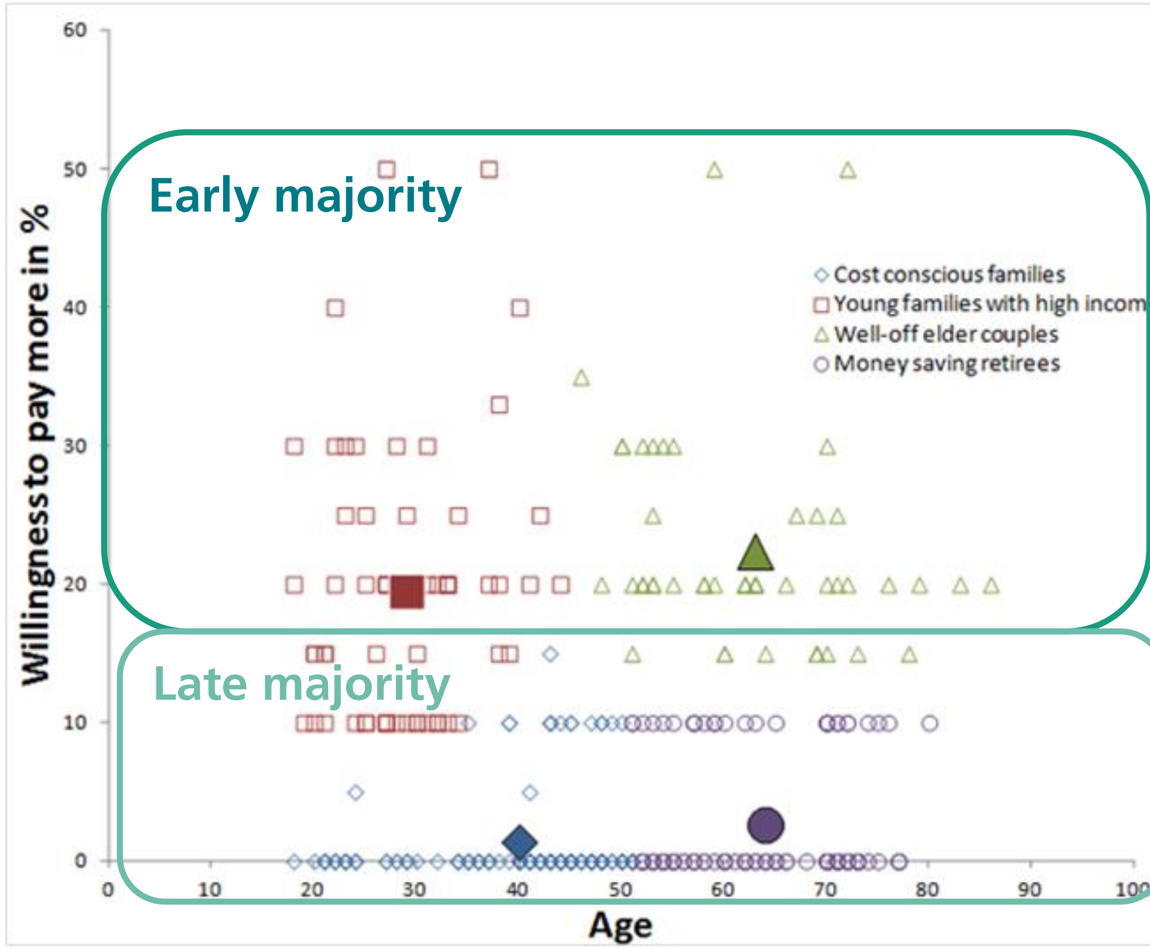
Predictor	Posterior Effect Probab. $P(\beta \neq 0)$	Posterior Mean	Posterior SD
Technology affinity	**100.0**	**0.742**	**0.096**
Willingness to pay more	**100.0**	**0.063**	**0.010**
Autonomy EV	**100.0**	**0.337**	**0.059**
Drivers license	**99.3**	**1.161**	**0.317**
Freedom ICE	**85.2**	**-0.270**	**0.149**
Education	64.7	0.123	0.106
Public Transport Avail.	30.1	0.118	0.203
Train Available	27.7	0.127	0.232
Availability ICE	18.3	0.033	0.079
sex	16.7	0.052	0.132
Availability EV	14.9	0.016	0.044
Autonomy ICE	7.9	-0.016	0.070
Working	6.4	0.015	0.069
Children	5.1	-0.012	0.066
Independence ICE	4.7	-0.006	0.038
Independence EV	3.4	0.002	0.018
HH members	3.0	0.002	0.016
Climate	2.2	0.003	0.034
Age	1.7	0.000	0.001
Freedom EV	1.7	0.000	0.011

compare these to p-values!

Fraunhofer ISI

The next big group of adopters can be divided into subgroups

- Use k-means clustering to find subgroups
- Early and late majority can be divided by willingness to pay more
- Both have younger and older group
- Socio-demographics differ too



Discussion and conclusions

Discussion and further research

- Some factors are rather pre-conditions and discrimative
- Interest, purchase intention and willingness to pay more currently directly related

Conclusions

- **Interest in new technology** and **symbolic value of car** (no or low association of freedom with conventional vehicles) are **most important distinguishing factors** for next big group of EV users
- **Bayesian model averaging** is coherent method against model uncertainty

Fraunhofer ISI

Thank you for your attention

Contact:

Patrick Plötz

Competence Center Energy Technologies and Energy Systems

Fraunhofer Institute for Systems and Innovation Research ISI

patrick.ploetz@isi.fraunhofer.de

Sources:

Peters, A. and Dütschke, E. (2014): How do consumers perceive electric vehicles? A comparison of German consumer groups, Journal of Environmental Policy & Planning **16**, 359-377,

Peters, A., Agosti, R., Popp, M., Ryf, B. (2011): Electric mobility – a survey of different consumer groups in Germany with regard to adoption. Proceedings to ECEEE Summer Study, June 2011, Belambra Presqu'île de Giens, France.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. Statistical science, 382-401.

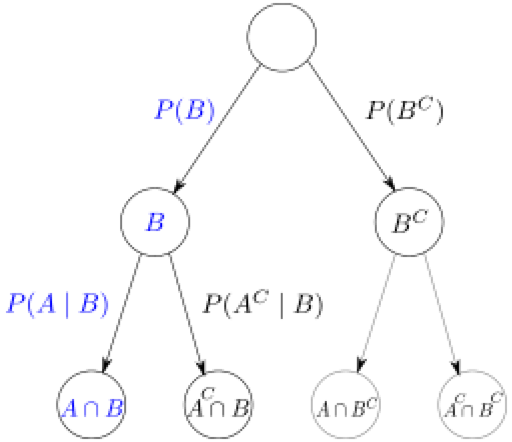
Adrian Raftery, Jennifer Hoeting, Chris Volinsky, Ian, Painter and Ka Yee Yeung (2017). BMA: Bayesian Model Averaging. R package version 3.18.7. https://CRAN.R-project.org/package=BMA

Fraunhofer ISI

Bayes' Theorem

- Definition of conditional probability:

$$P\,(A \mid B)\,P(B) = P(A \cap B)$$



- This implies:

$$P\,(A \mid B) \;=\; \frac{P(A \cap B)}{P(B)} \;=\; \frac{\frac{P(A \cap B)}{P(A)} \cdot P(A)}{P(B)} \;=\; \frac{P\,(B \mid A) \cdot P(A)}{P\,(B)}$$

- Bayes' Theorem results:

$$P(A \mid B) \;=\; \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

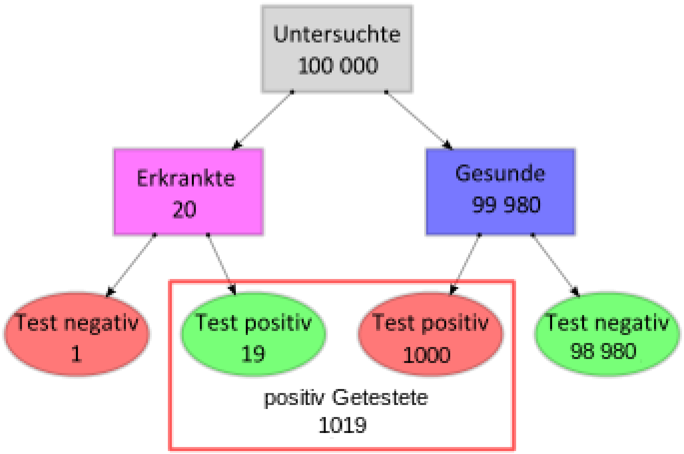
- Different generalisations exist:

$$P(A_i \mid B) \;=\; \frac{P(B \mid A_i) \cdot P(A_i)}{P(B)} \;=\; \frac{P\,(B \mid A_i) \cdot P(A_i)}{\sum_{j=1}^{N} P\,(B \mid A_j) \cdot P(A_j)}$$

Fraunhofer ISI

Example

- Medical (and statistical) tests can be wrong
- Let K denote the state Sickness and T the test result
- For sick people, the test is correct in 95 % of the cases, i.e. P(T|K) = 0.95
- The test yields "sick" also for 1% of the healthy people, i.e. P(T|~K) = 0.01.
- What is the probability that a tested person is actually sick?



$$P(K \mid T) = \frac{P(T \mid K)P(K)}{P(T \mid K)P(K) + P(T \mid K^c)P(K^c)} = \frac{0{,}95 \cdot 0{,}0002}{0{,}95 \cdot 0{,}0002 + 0{,}01 \cdot 0{,}9998} \approx 0{,}0186$$

- Despite the positive test result only about 2% are actually sick with this rare desease!
- Accuracy of tests is important for very rare diseases (e.g. HIV – a second test is obligatory in Germany)

Bayesian inference

- Let x denote a random variable from distribution $p(x|\theta)$ with unknown parameter θ
- Data points $\boldsymbol{x} = x_1, \ldots, x_n$ have been observed
- The prior distribution of the parameter is assumed to be $p(\theta|\alpha)$
- The probability to observe exactly the data $\boldsymbol{x}$ if θ had exactly the value θ is given by the likelihood function: $L(\theta\,|\boldsymbol{x}) = \mathrm{p}(\boldsymbol{x}|\theta)$
- **Result**: The Posterior distribution for the Parameter θ under the data is

$$p(\theta \mid \mathbf{X}, \alpha) = \frac{p(\mathbf{X} \mid \theta) p(\theta \mid \alpha)}{p(\mathbf{X} \mid \alpha)} \propto p(\mathbf{X} \mid \theta) p(\theta \mid \alpha)$$

where

$$p(\mathbf{X} \mid \alpha) = \int_\theta p(\mathbf{X} \mid \theta) p(\theta \mid \alpha) \, \mathrm{d}\theta$$

- Summary:

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}.$$

Application 4: Bayesian Model Averaging

- Linear models with several variables are very common:

$$y = \alpha_\gamma + X_\gamma \beta_\gamma + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I)$$

- Yet, the choice of variables can heavily influence results! Which model to choose?
- With K parameters 2^K models M_γ are possible!
- In the following: dependent variable: Interest in electric vehicles (EV); independent variables: Socio-economic attributes and other items.
- A single model suggests very low uncertainty:

```
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        -5.218129   0.623700  -8.366  < 2e-16 ***
Techaffin           0.770285   0.092990   8.284  < 2e-16 ***
WTPM                0.063270   0.009491   6.666 2.62e-11 ***
education           0.198189   0.064894   3.054  0.00226 **
autonomyEV          0.362730   0.051936   6.984 2.87e-12 ***
freedomICE         -0.323808   0.081642  -3.966 7.30e-05 ***
drivers_licenseyes  1.149443   0.275245   4.176 2.97e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, 382-401.
Wesche, Plötz, Dütschke (2015): *How to trigger mass market adoption of electric vehicles? Factors predicting interest in electric vehicles in Germany*. Submitted for publication