

rego

**Stata module for decomposing
goodness of fit according to
Shapley and Owen values**



Frank Huettner and Marco Sunder
Department of Economics
University of Leipzig, Germany

Motivation

- Challenges in interpreting regression results by the analyst (the **quick and dirty** way)
 - Sign *econometrics*
 - Confusion of statistical significance and *economic* significance
 - Interaction terms
- Challenges to readers
 - Units of measurement
 - "Omitted coefficients"

Decompose

*R*²

Motivation

- Imagine regressor variables as “players”
 - Cooperative game theory: games with transferable utility
 - Regressors may form coalition – how to distribute the gains from cooperation?
 - **Shapley** (1953) and **Owen** (1977) **Values** as means to decompose goodness of fit
 - Axioms under which these values are unique solutions
- “**hierarchical partitioning**”
(Lindeman et al. 1980, Chevan & Sutherland 1991)
- Previous implementations:
 - Stata:** *shapley* (Kolenikov 2000)
 - R:** *relaimpo* package (Grömping 2006)

Illustrative calculation of the Shapley Value

- Assume the “full model” is...

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Illustrative calculation of the Shapley Value

- Assume the “full model” is...

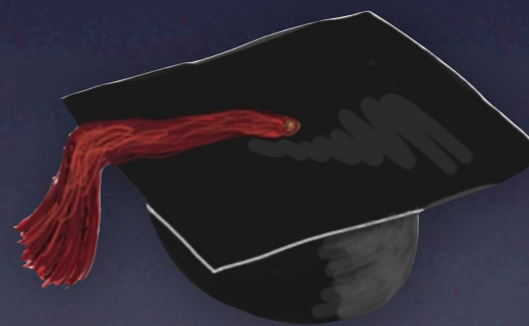
$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$



log wage



intelligence



schooling

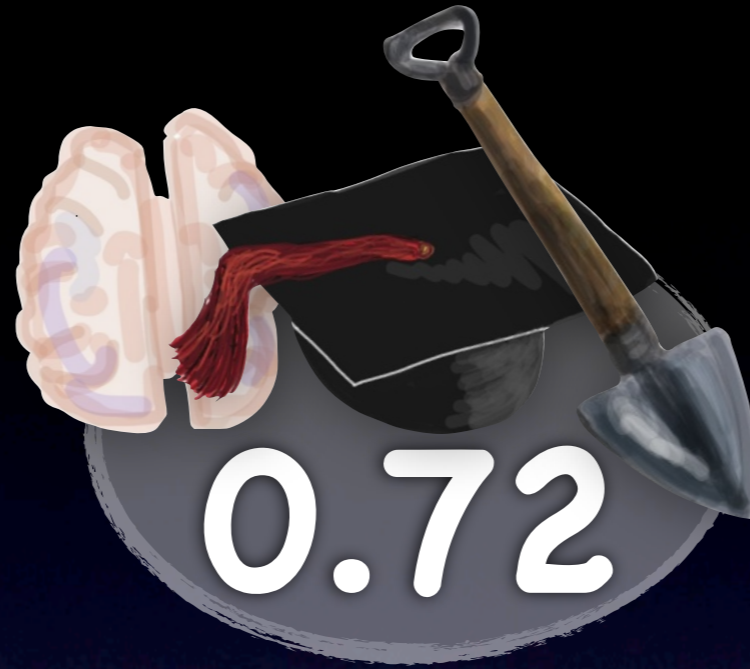


experience

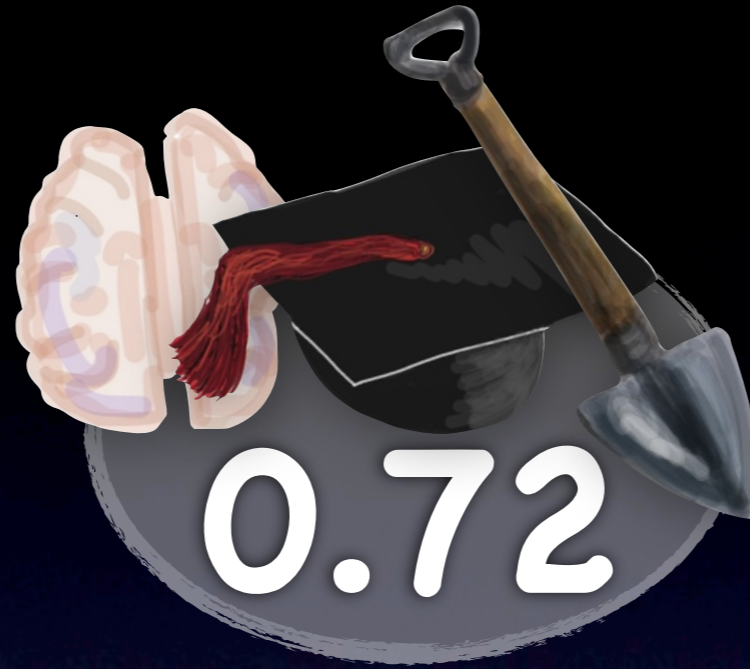
Worths in 2^k (sub-)models



Worths in 2^k (sub-)models



Worths in 2^k (sub-)models



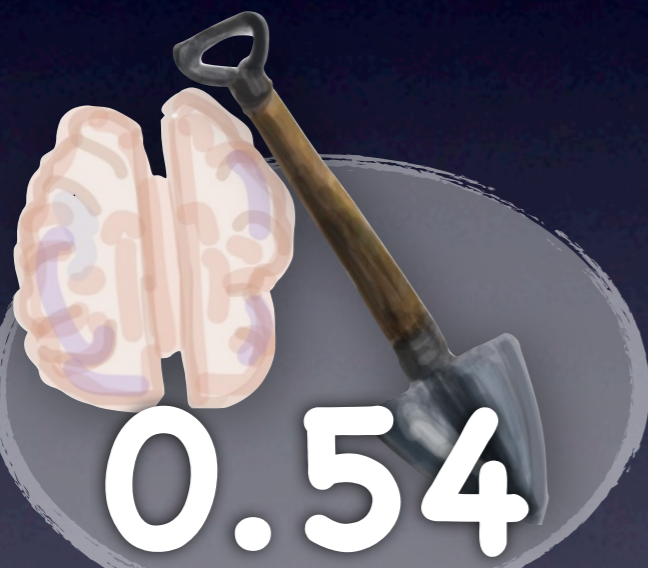
Worths in 2^k (sub-)models



0.72



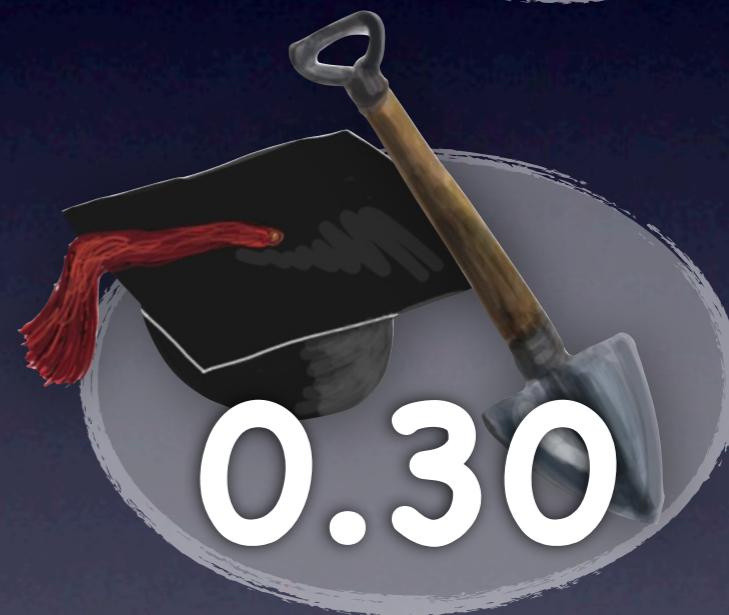
0.00



0.54



0.48



0.30

Worths in 2^k (sub-)models



0.72



0.00



0.54



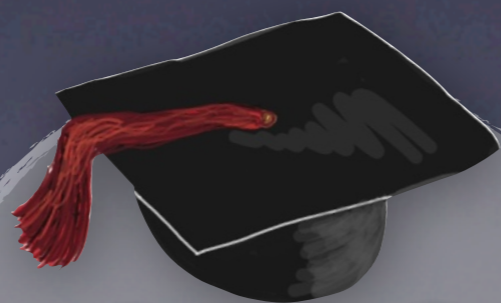
0.48



0.30



0.24



0.12




0.18

Worths in 2^k (sub-)models




$\Delta = 0.18$



0.72

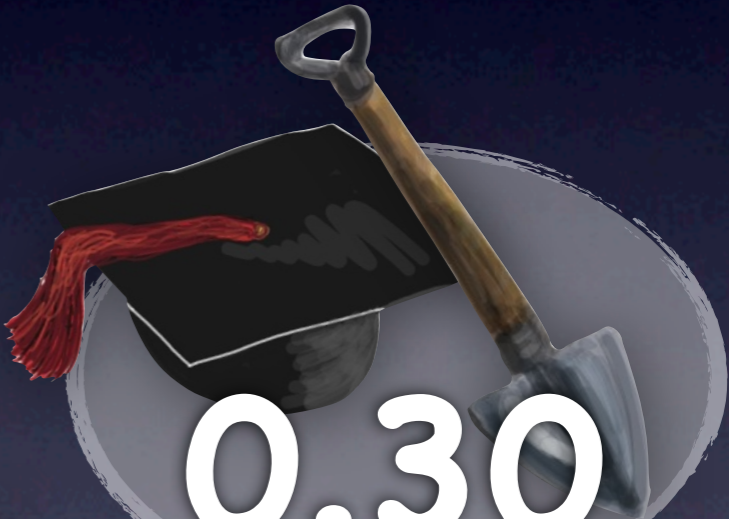
0.00



0.54



0.48



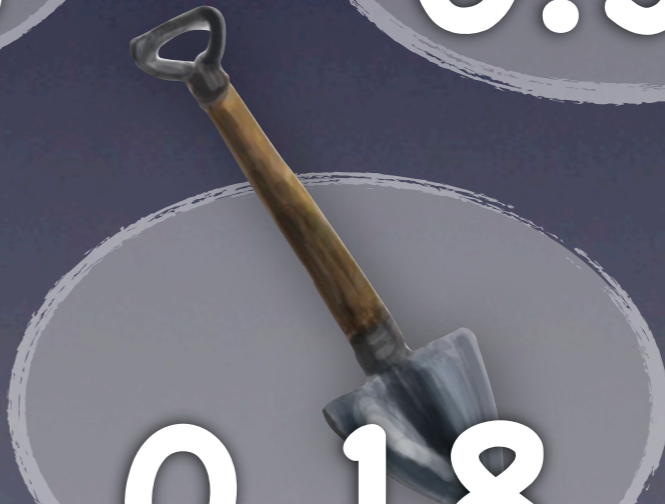
0.30



0.24



0.12



0.18

Marginal contributions (MC)

remove first

remove first

remove first







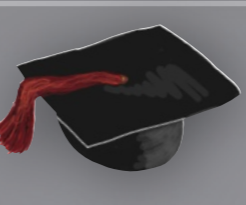

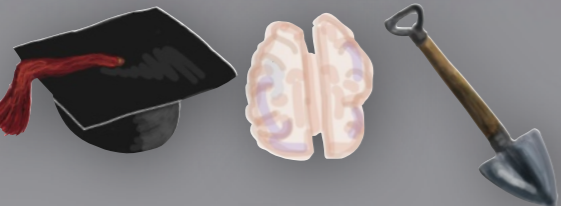





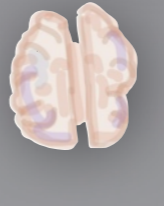

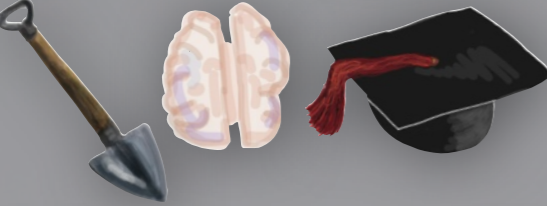
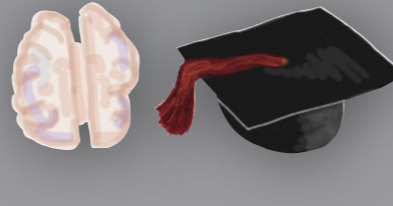
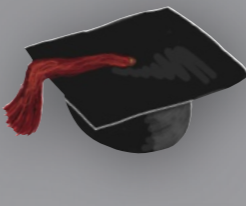

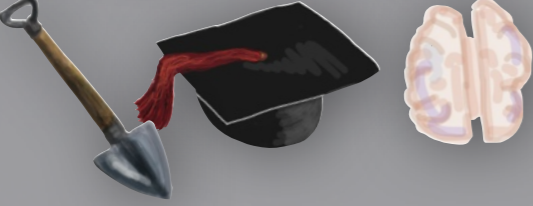
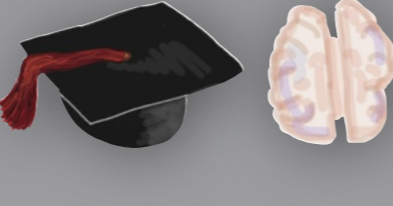


			

Marginal contributions (MC)

remove first

remove first

remove first

 0.72	 0.30	 0.18	 0.00
			
			
			
			
			

Marginal contributions (MC)

remove first

remove first

remove first

 0.72	 0.30	 0.18	 0.00
 0.72	 0.30	 0.12	 0.00
 0.72	 0.54	 0.18	 0.00
 0.72	 0.54	 0.24	 0.00
 0.72	 0.48	 0.12	 0.00
 0.72	 0.48	 0.24	 0.00

Marginal contributions (MC)

remove first

remove first

remove first

 0.72	 0.30 	 0.18	 0.00
 0.72	 0.30	 0.12 	 0.00
 0.72 	 0.54	 0.18	 0.00
 0.72 	 0.54	 0.24	 0.00
 0.72	 0.48	 0.12 	 0.00
 0.72	 0.48 	 0.24	 0.00

$$\text{Sh}(x_j) = \frac{1}{k!} \sum_{\text{all permutations}} \triangle_{MC}$$

$$\text{Sh}(x_j) = \frac{1}{k!} \sum_{\text{all permutations}} \triangle_{MC}$$

$$\text{Sh}(\text{🎓}) = \frac{1}{6} \cdot (0.12 + 0.12 + 0.18 + 0.18 + 0.12 + 0.24) = 0.16$$

$$\text{Sh}(x_j) = \frac{1}{k!} \sum_{\text{all permutations}} \text{MC}$$

$$\text{Sh}(\text{🎓}) = \frac{1}{6} \cdot (0.12 + 0.12 + 0.18 + 0.18 + 0.12 + 0.24) = 0.16$$

$$\text{Sh}(\text{🧠}) = \frac{1}{6} \cdot (0.42 + 0.42 + 0.36 + 0.24 + 0.36 + 0.24) = 0.34$$

$$\text{Sh}(\text{🛠️}) = \frac{1}{6} \cdot (0.18 + 0.18 + 0.18 + 0.30 + 0.24 + 0.24) = 0.22$$

$$\text{Sh}(x_j) = \frac{1}{k!} \sum_{\text{all permutations}} \text{MC}$$

$$\text{Sh}(\text{🎓}) = \frac{1}{6} \cdot (0.12 + 0.12 + 0.18 + 0.18 + 0.12 + 0.24) = 0.16$$

$$\text{Sh}(\text{🧠}) = \frac{1}{6} \cdot (0.42 + 0.42 + 0.36 + 0.24 + 0.36 + 0.24) = 0.34$$

$$\text{Sh}(\text{🛠️}) = \frac{1}{6} \cdot (0.18 + 0.18 + 0.18 + 0.30 + 0.24 + 0.24) = 0.22$$

$$\Sigma = 0.72$$

Axioms

Axioms

1) **Efficiency**

GOF of full model is decomposed among the regressors

Axioms

1) **Efficiency**

GOF of full model is decomposed among the regressors

2) **Monotonicity**

Increase in R^2 must not decrease the value

Axioms

1) **Efficiency**

GOF of full model is decomposed among the regressors

2) **Monotonicity**

Increase in R^2 must not decrease the value

3) **Equal treatment**

Perfect substitutes (in terms of GOF) receive the same value

Axioms

1) **Efficiency**

GOF of full model is decomposed among the regressors

2) **Monotonicity**

Increase in R^2 must not decrease the value

3) **Equal treatment**

Perfect substitutes (in terms of GOF) receive the same value

The **Shapley Value** is the **only** value that satisfies these properties (Young 1985).


A priori grouped regressors

- Analyst may believe that some regressor variables belong together, e.g.: polynomial terms, region dummies
- Regressor variables are partitioned: $\mathcal{G} = \{G_1, \dots, G_\ell, \dots, G_\gamma\}$
- When thinking about marginal contributions of variables from G_ℓ , variables of $G_{q \neq \ell}$ must be completely absent or present
- This limits the set of admissible model permutations

👉 **Owen Value**

- E.g.:  "nature",  "nurture"

4 permutations respect \mathcal{G}

 0.72 	 0.30 	 0.18 	 0.00
 0.72 	 0.30 	 0.12 	 0.00
 0.72 	 0.54 	 0.18 	 0.00
 0.72 	 0.54 	 0.24 	 0.00
 0.72 	 0.48 	 0.12 	 0.00
 0.72 	 0.48 	 0.24 	 0.00

4 permutations respect \mathcal{G}

 0.72 	 0.30 	 0.18 	 0.00
 0.72 	 0.30 	 0.12 	 0.00
 0.72 	 0.54 	 0.18 	 0.00
 0.72 	 0.54 	 0.24 	 0.00
 0.72 	 0.48 	 0.12 	 0.00
 0.72 	 0.48 	 0.24 	 0.00

4 permutations respect \mathcal{G}

 0.72  	  0.30	 0.18	\emptyset 0.00
  0.72 	  0.30	 0.12	\emptyset 0.00
   0.72	  0.54	 0.18	\emptyset 0.00
   0.72	  0.54	 0.24	\emptyset 0.00
   0.72	  0.48	 0.12	\emptyset 0.00
   0.72	  0.48	 0.24	\emptyset 0.00

$$Ow(\text{brain}) = 0.33 \quad Ow(\text{mortarboard}) = 0.165 \quad Ow(\text{shovel}) = 0.225$$

Axioms

Axioms

1) **Efficiency***

Axioms

- 1) **Efficiency***
- 2) **Monotonicity***

Axioms

- 1) **Efficiency***
- 2) **Monotonicity***
- 3) **Equal treatment of players**

Axioms

- 1) **Efficiency***
- 2) **Monotonicity***
- 3) **Equal treatment of players**
- 4) **Equal treatment of groups**

Axioms

- 1) **Efficiency***
- 2) **Monotonicity***
- 3) **Equal treatment of players**
- 4) **Equal treatment of groups**

The **Owen Value** is the **only** value that satisfies these properties (Khmelnitskaya and Yanovskaya 2007).

Stata implementation

- With grouping, “large” models become possible
- User decides for which groups to calculate Owen values
- R^2 calculated from covariance structure of the data
- Syntax uses “\” to indicate group boundaries in *varlist*
- Computation in Mata
- Bootstrapping option



help rego

Title

rego — Linear regression with Shapley and Owen decomposition of R-squared (for Stata 9+)

Syntax

```
rego varlist [if] [in] [, options]
```

*options**description*

Main

detail

calculate Owen values for individual variables

vce(*vcetype*)determine what standard errors are computed: `ols`, `robust`, `cluster clustervar`.noperc

do not display Owen/Shapley values as percentages of overall R-squared

force

perform calculations even if it would take a lot of time

Bootstrap

bsreps

number of bootstrap replications

level

confidence level, in %

`xi` (Stata 9+) and time series operators (Stata 11+) may be used, but factor variables are not supported.

Description

`rego` uses results from `regress` and decomposes the share of explained variance (measured by R-squared) into contributions by individual regressor variables or groups of regressor variables. The former case is calculated as Owen value, the latter as Shapley value (for groups). In the special case that each group consists of only one regressor, both quantities coincide. If not suppressed by the `noperc` option, the contributions are displayed as percentages of overall R-squared.

Remarks

Groups of regressors are specified with the help of the backslash ("`\`") symbol. It must be placed in the varlist after the last variable of each group. If no groups are defined, `rego` assumes that each regressor belongs to a group of its own and calculates the "traditional" Shapley value.

Instead of using the `detail` option after the comma, the user may also specify "`(detail)`" (without quotation marks)

auto dataset: Shapley values

```
. rego mpg weight length headroom price
```

Gr	Regressor	Coef.	Std.Err.	P> t	Std.Coeff.	Shapley %R2
1	weight	-.0030149 *	.0017844	0.096	-0.4050	41.0462
2	length	-.0942241	.0588767	0.114	-0.3626	39.5705
3	headroom	-.1387698	.5611849	0.805	-0.0203	8.1060
4	price	-.0001795	.0001706	0.296	-0.0915	11.2773
-	Intercept	49.6308	6.375722	0.000		
	Observations	74				
	Overall R2	0.66678				
	Root MSE	3.435139				
	F-stat. Model	34.51741 ***		0.000		
	Log Likelihood	-193.7332				

auto dataset: Owen values

```
. rego mpg \ weight length (detail) \ headroom \ price
```

Gr	Regressor	Coef.	Std.Err.	P> t	Ind. %R2	Group %R2
1	weight	-.0030149 *	.0017844	0.096	36.8744	72.2732
	length	-.0942241	.0588767	0.114	35.3988	
2	headroom	-.1387698	.5611849	0.805		11.8612
3	price	-.0001795	.0001706	0.296		15.8656
-	Intercept	49.6308	6.375722	0.000		
	Observations	74				
	Overall R2	0.66678				
	Root MSE	3.435139				
	F-stat. Model	34.51741 ***		0.000		
	Log Likelihood	-193.7332				

auto dataset: Bootstrap Owen values

```
. rego mpg \ weight length \ headroom \ price, detail noperc bs(2000) level(90)
```

Bootstrapping

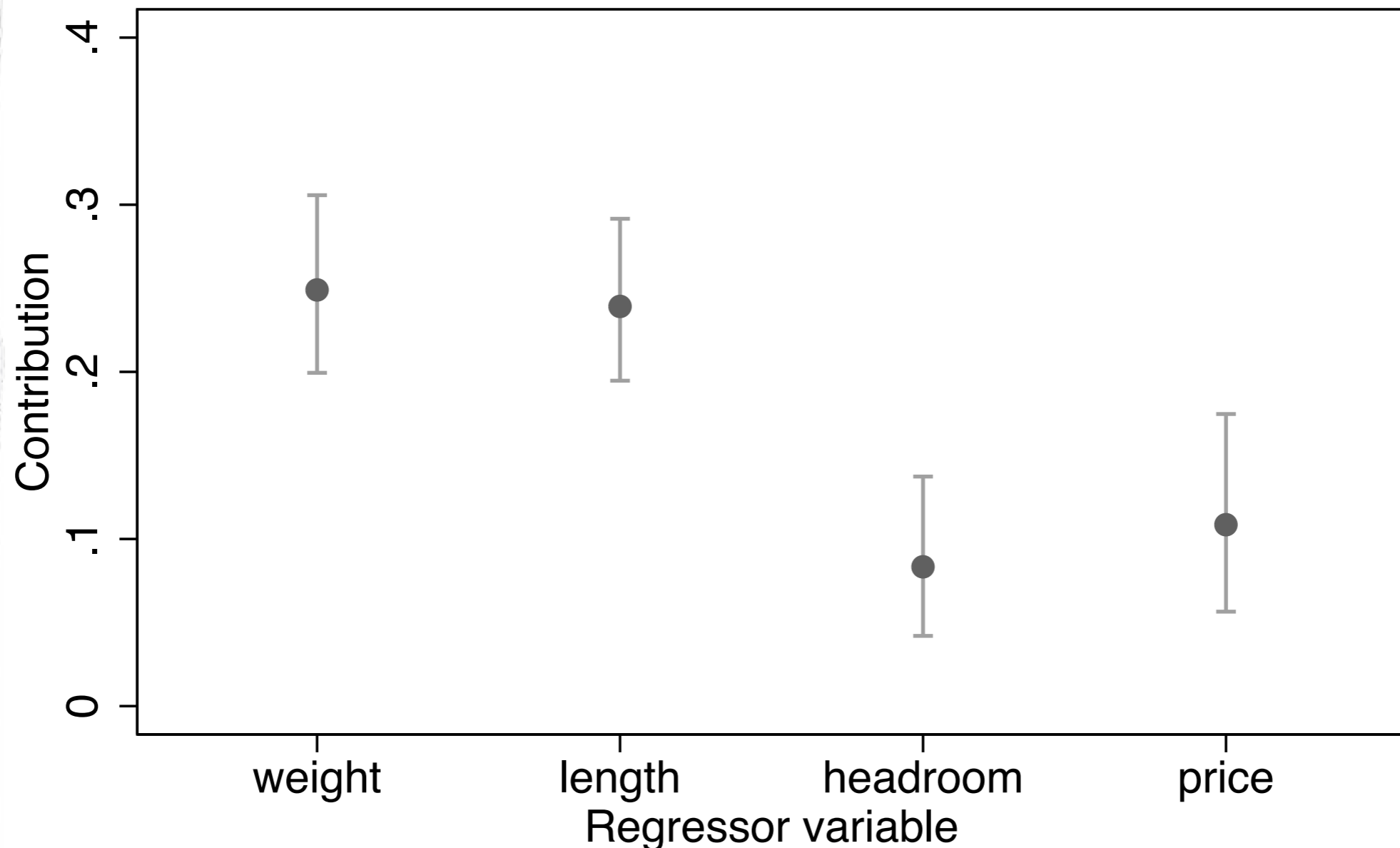
```
..... 1000
..... 2000
```

Bootstrap percentile confidence intervals for **REGO**
Approximate level of confidence: **90 %**

Gr	Regressor	Individual R2			Group R2		
		Median	CI lower	upper	Median	CI lower	upper
1	weight	0.250	0.198	0.305	0.492	0.414	0.568
	length	0.238	0.192	0.291			
2	headroom	0.081	0.041	0.140	0.081	0.041	0.140
3	price	0.109	0.056	0.176	0.109	0.056	0.176

auto dataset: Bootstrap Owen values

Absolute contributions to R-squared
90% percentile confidence intervals with median



Wage regression: German male employees

OLS regression results with decomposition of R^2 (in %)

Group	Regressor	Coef.	R^2 decomposition (%)	
			Owen	Group
1	SCT	0.789 *	3.0	33.2
	SCT \times EDUC	-0.048 *	8.3	
	EDUC	0.103 ***	21.9	
2	EXPER	0.025 ***	7.0	11.0
	(EXPER) ² /100	-0.041 ***	4.0	
3	TENURE	0.017 ***	9.3	14.3
	(TENURE) ² /100	-0.029 **	5.0	
4	MARRIED	0.084 ***	5.0	5.0
5	Firm size	(3 dummies) ***		14.7
6	Industry	(6 dummies) ***		5.5
7	Region	(14 dummies) ***		16.2
Observations		850		
Full model R^2		0.501		

Remark: */**/** denotes statistical significance at the 10% / 5% / 1% level for individual variables (t-test) or groups of dummy variables (F-test), based on the heteroscedasticity-robust covariance matrix.

Wage regression: German male employees

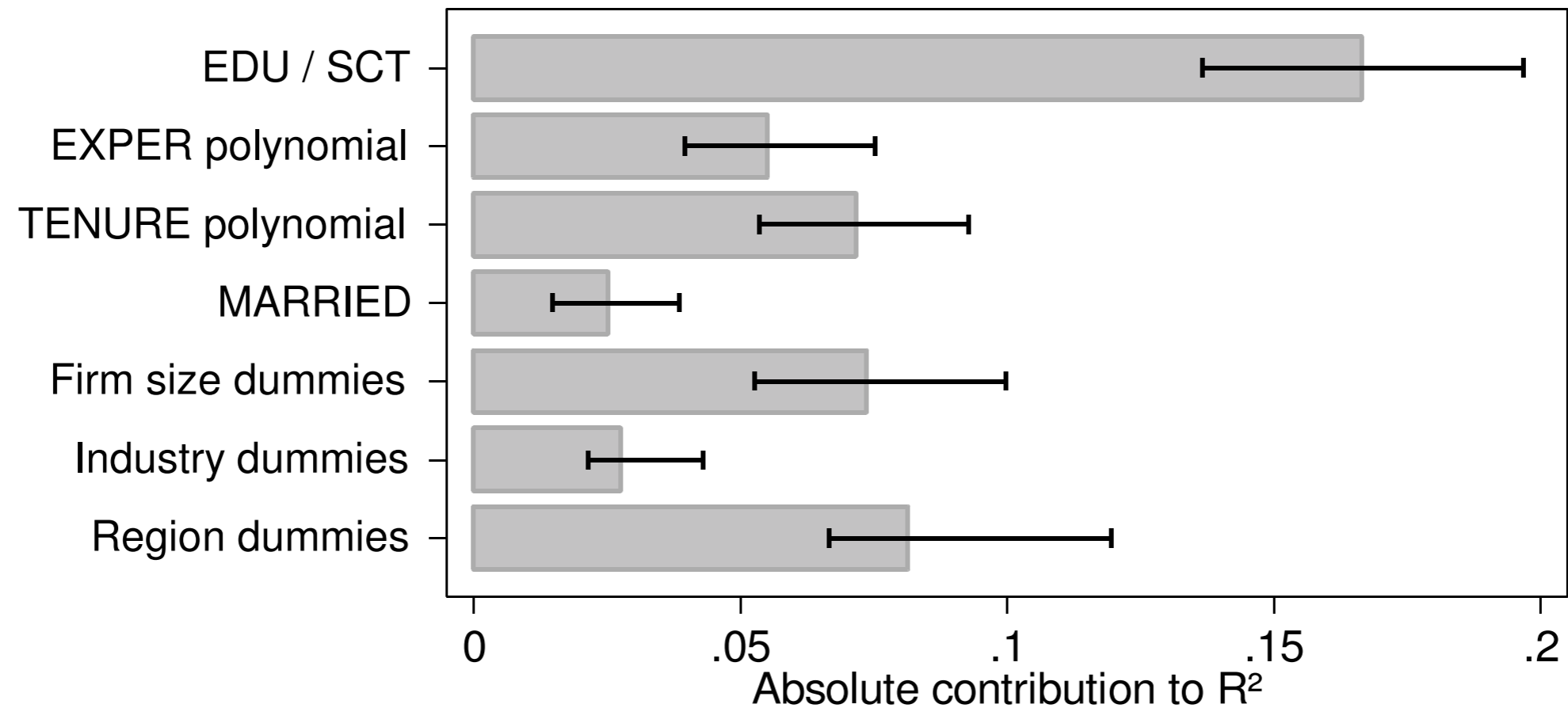


FIG 1. *Decomposition results for groups, with 90% bootstrap confidence intervals, based on 5000 bootstrap replications.*

Wage regression: German male employees

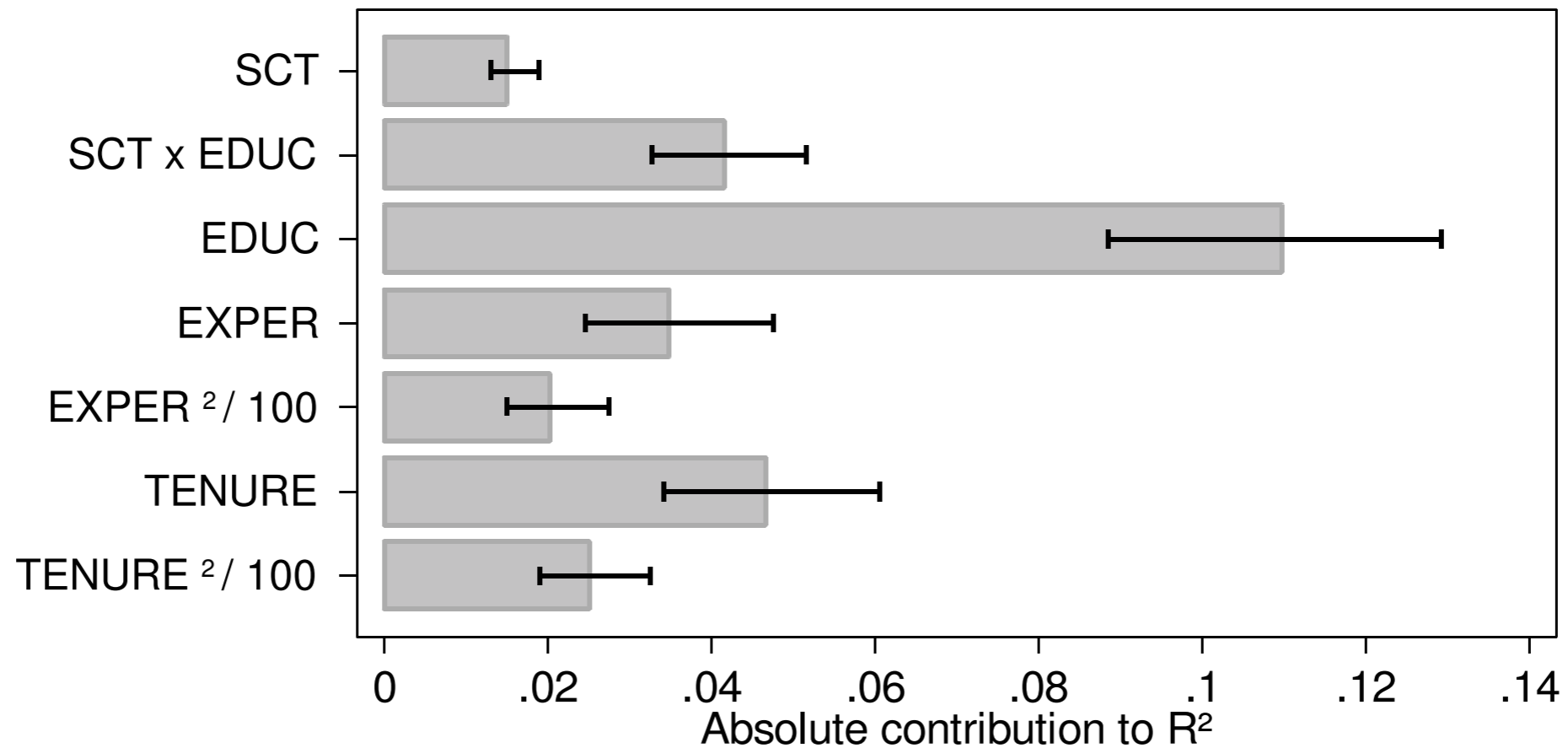


FIG 2. Owen value decomposition results for 'human capital' variables, with 90% bootstrap confidence intervals, based on 5000 bootstrap replications.

Limits

- Only OLS with R^2 decomposition at this time
- Does not yet accept factor variables or weights
- ...

Limits

- Only OLS with R^2 decomposition at this time
- Does not yet accept factor variables or weights
- ...

Possible extensions?

- Decomposition of other measures (e.g., AIC)
- More levels of aggregation
- Essential regressors / fixed effects model (*econ*)

Thank you!

Thank you!

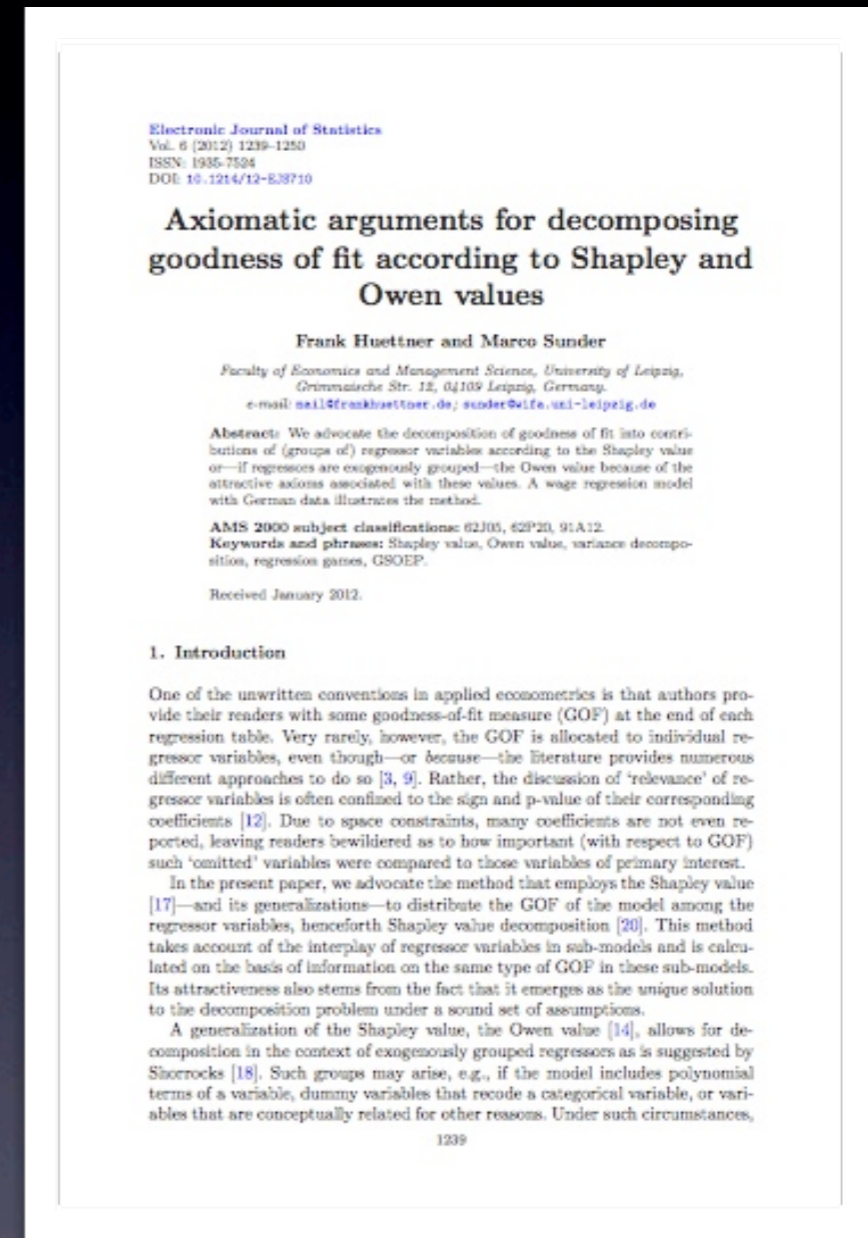


`www.uni-leipzig.de/~rego`

Thank you!



www.uni-leipzig.de/~rego



Huettner & Sunder (2012)
Electronic Journal of Statistics