Model-Based Recursive Partitioning: Ideas, Theory, and Implementation

Achim Zeileis, Florian Wickelmaier, Julia Kopf, Carolin Strobl

http://statmath.wu.ac.at/~zeileis/

Overview

- Motivation: Trees and leaves
- Theory
  - Model estimation
  - Tests for parameter instability
  - Segmentation
  - Pruning
- Applications
  - Linear regression: Costly journals
  - Bradley-Terry model: A model model
  - Rasch model: Differential item functioning
- Implementation
  - Generic object-oriented infrastructure
  - Model plugins
  - Methods

Motivation: Trees

Breiman (2001, *Statistical Science*) distinguishes two cultures of statistical modeling:

- **Data models**
  - Stochastic models, typically parametric.
  - Predominant modeling strategy in the social sciences.
  - Regression models are workhorse for empirical analyses.
- **Algorithmic models**
  - Flexible models, data-generating process unknown.
  - Example: Regression trees model dependent variable Y by "learning" a partition w.r.t explanatory variables $Z_1, \dots, Z_l$.
  - Few applications in the social sciences.

Motivation: Leaves

Examples for trees: CART and C4.5 in statistical and machine learning, respectively.

Key features: Predictive power in nonlinear regression relationships, *and* interpretability (enhanced by visualization), i.e., no "black box".

Typically: Simple models for univariate Y, e.g., mean or proportion.

Idea: More complex models for multivariate Y, e.g., multivariate normal model, regression models, etc.

Here: Synthesis of parametric data models and algorithmic tree models.

Recursive partitioning

Base algorithm:

1. Fit model for Y.
2. Assess association of Y and each Z_j.
3. Split sample along the Z_{j^*} with strongest association: Choose breakpoint with highest improvement of the model fit.
4. Repeat steps 1–3 recursively in the subsamples until some stopping criterion is met.

Here: Segmentation (3) of parametric models (1) with additive objective function using parameter instability tests (2) and associated statistical significance (4).

1. Model estimation

Models: $\mathcal{M}(Y, \theta)$ with (potentially) multivariate observations $Y \in \mathcal{Y}$ and k-dimensional parameter vector $\theta \in \Theta$.

Parameter estimation: $\widehat{\theta}$ by optimization of objective function $\Psi(Y, \theta)$ for n observations Y_i ($i = 1, \dots, n$):

$$\widehat{\theta} \quad = \quad \underset{\theta \in \Theta}{\text{argmin}} \sum_{i=1}^{n} \Psi(Y_i, \theta).$$

Special cases: Maximum likelihood (ML), weighted and ordinary least squares (OLS and WLS), quasi-ML, and other M-estimators.

Central limit theorem: If there is a true parameter θ_0 and given certain weak regularity conditions, $\hat{\theta}$ is asymptotically normal with mean θ_0 and sandwich-type covariance.

2. Tests for parameter instability

Estimating function: Model deviations can be captured by

$$\psi(Y_i, \widehat{\theta}) = \frac{\partial \Psi(Y, \theta)}{\partial \theta} \Big|_{Y_i, \widehat{\theta}}$$

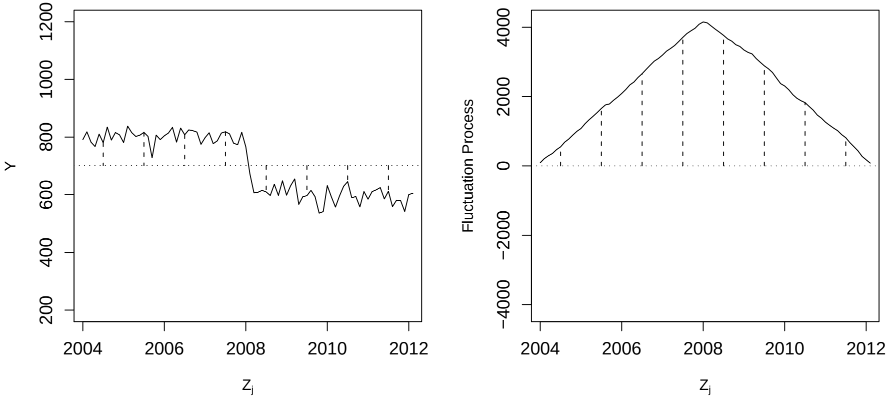
also known as *score function* or contributions to the *gradient*.

Fluctuation processes: Systematic changes in parameters over the variables $Z = (Z_1, \dots, Z_l)$ can be assessed by cumulative sums of the empirical estimating functions.

Fluctuation tests: Aggregate process to test statistics.

- Andrews' supLM test for numerical Z_j,
- χ^2-type test for categorical Z_j.

2. Tests for parameter instability



3. Segmentation

Goal: Split model into $b = 1, \dots, B$ segments along the partitioning variable Z_j associated with the highest parameter instability. Local optimization of

$$\sum_b \sum_{i \in I_b} \Psi(Y_i, \theta_b).$$

Here: $B = 2$, binary partitioning.

4. Pruning

Pruning: Avoid overfitting.

Pre-pruning: Internal stopping criterion. Stop splitting when there is no significant parameter instability.

Post-pruning: Grow large tree and prune splits that do not improve the model fit (e.g., via crossvalidation or information criteria).

Here: Pre-pruning based on Bonferroni-corrected *p* values of the fluctuation tests.

Costly journals

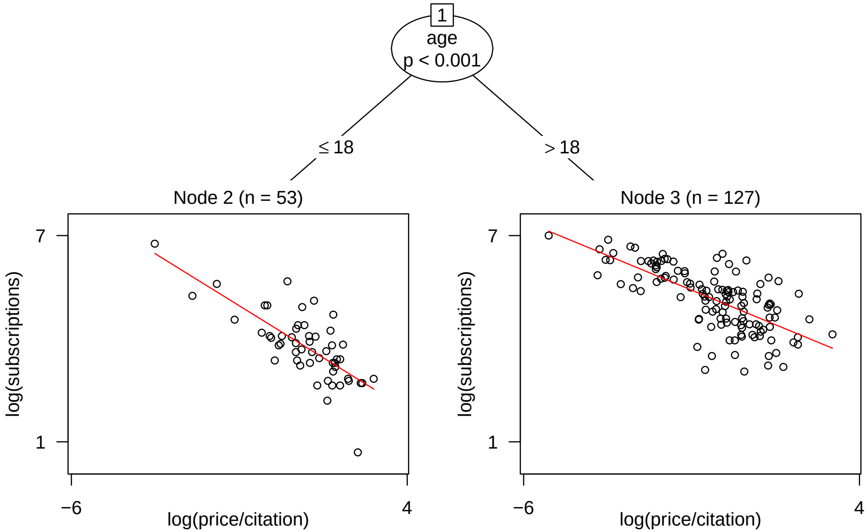
Task: Price elasticity of demand for economics journals.

Source: Bergstrom (2001, *Journal of Economic Perspectives*) "Free Labor for Costly Journals?", used in Stock & Watson (2007), *Introduction to Econometrics*.

Model: Linear regression via OLS.

- Demand: Number of US library subscriptions.
- Price: Average price per citation.
- Log-log-specification: Demand explained by price.
- Further variables without obvious relationship: Age (in years), number of characters per page, society (factor).

Costly journals



Costly journals

Recursive partitioning:

	Regressors		Partitioning variables				
	(Const.)	log(Pr./Cit.)	Price	Cit.	Age	Chars	Society
1	4.766	-0.533	3.280	5.261	42.198	7.436	6.562
	< 0.001	< 0.001	0.660	0.988	< 0.001	0.830	0.922
2	4.353	-0.605	0.650	3.726	5.613	1.751	3.342
	< 0.001	< 0.001	0.998	0.998	0.935	1.000	1.000
3	5.011	-0.403	0.608	6.839	5.987	2.782	3.370
	< 0.001	< 0.001	0.999	0.894	0.960	1.000	1.000

(Wald tests for regressors, parameter instability tests for partitioning variables.)

Implementation

Software provided in R system for statistical computing and graphics. Available under General Public License from Comprehensive R Archive Network at `http://CRAN.R-project.org/`:

- Trees/recursive partytioning: **party**.
- Structural change inference: **strucchange**.
- Bradley-Terry/Rasch trees: **psychotree**.

Implementation

Generic infrastructure:

- Model-based recursive partitioning in function `mob()` from package **party**.
- Takes care of all steps except model fitting (i.e., estimation of θ by minimization Ψ).
- Leverages inference methods from package **strucchange**.
- Models can be plugged in using object-oriented approach: Model objects and methods.
- Visualization can be customized by panel functions (in **grid**).

Implementation

Model plugins:

- Model fitting function (à la `lm()` or – even better – `lm.fit()`) that returns a classed object (like "`lm`").
- Required methods: `estfun()`, `weights()`, `reweight()` (at least for 0/1 weights), and extractor for objective function (e.g., `deviance()` or `logLik()`).
- Optional methods (reused if available): `print()`, `predict()`, `coef()`, `summary()`, `residuals()`.
- Additional glue: S4 "`StatModel`" objects (**modeltools** package). Separate data handling (in particular, formula processing) from model fitting. (Hopefully facilitated in future versions.)

Implementation

Examples:

- `bttree()`: Interface for `mob()` with `btReg.fit()`.
- `raschtree()`: Interface for `mob()` with `RaschModel.fit()`.

Summary

Model-based recursive partitioning:

- Synthesis of classical parametric data models and algorithmic tree models.
- Based on modern class of parameter instability tests.
- Aims to minimize clearly defined objective function by greedy forward search.
- Can be applied general class of parametric models.
- Alternative to traditional means of model specification, especially for variables with unknown association.
- Object-oriented implementation freely available: Extension for new models requires some coding but not too extensive if interfaced model is well designed.

References

Zeileis A, Hornik K (2007). "Generalized M-Fluctuation Tests for Parameter Instability." *Statistica Neerlandica*, **61**(4), 488–508. `doi:10.1111/j.1467-9574.2007.00371.x`

Zeileis A, Hothorn T, Hornik K (2008). "Model-Based Recursive Partitioning." *Journal of Computational and Graphical Statistics*, **17**(2), 492–514. `doi:10.1198/106186008X319331`

Kleiber C, Zeileis A (2008). *Applied Econometrics with R*. Springer-Verlag, New York. URL `http://CRAN.R-project.org/package=AER`

Strobl C, Wickelmaier F, Zeileis A (2010). "Accounting for Individual Differences in Bradley-Terry Models by Means of Recursive Partitioning." *Journal of Educational and Behavioral Statistics*, Forthcoming. URL `http://statmath.wu.ac.at/~zeileis/papers/Strobl+Wickelmaier+Zeileis-2010.pdf`