Towards explicit calculation of terms in the Bias-Variance Decomposition (BVD).

Symposium on Data Science and Statistics

June 2-4, 2021

Neil A. Hamlett, D.Sc., MBA

Strategic Consulting Data Scientist, Uncertainty Research, LLC

neil.hamlett@uncertainty-research.science

https://www.linkedin.com/in/neil-hamlett-strategic-quant/

https://hamlett-neil-ur.github.io/

"Uncertainty is a key source of competitive advantage."

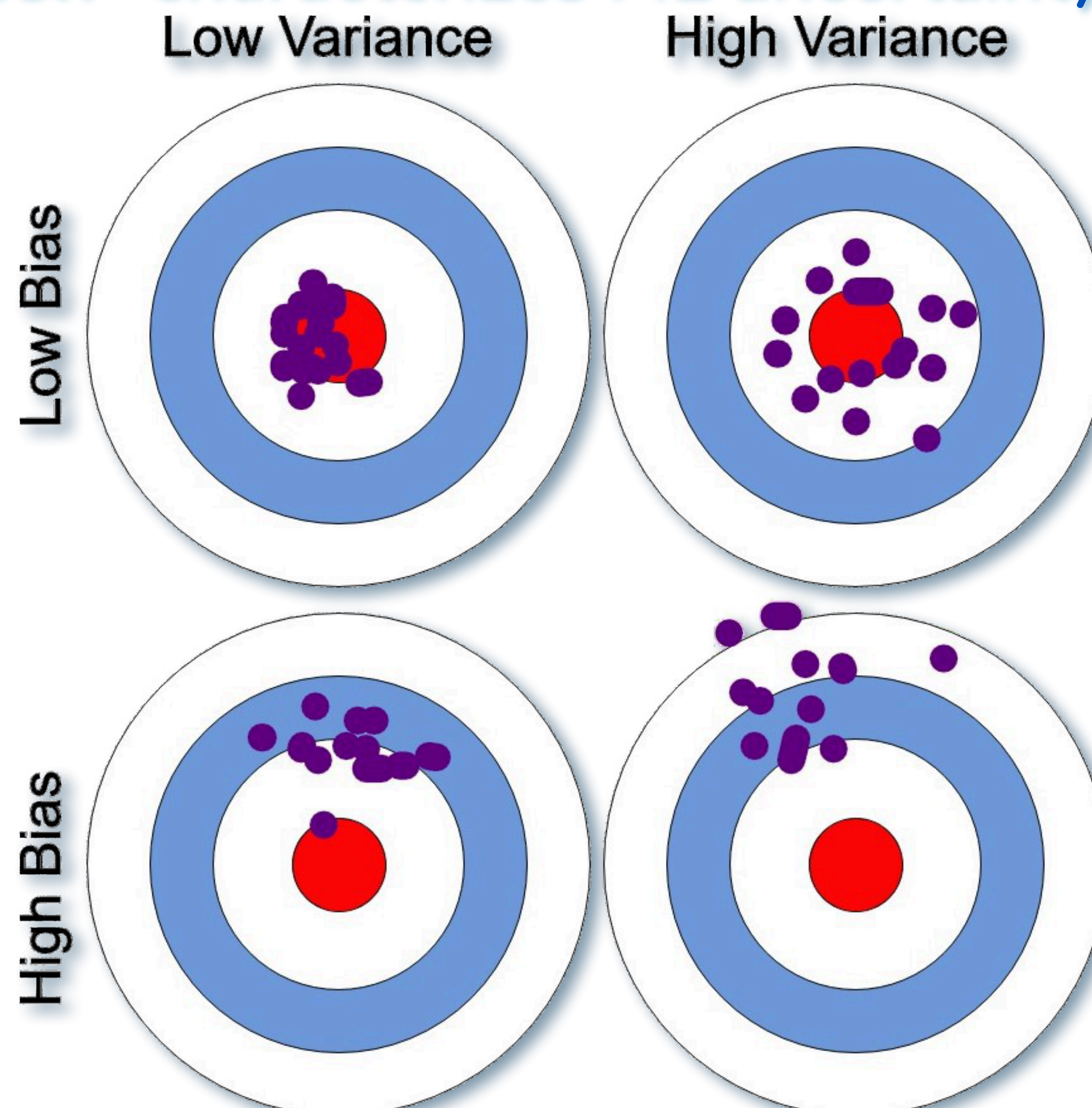
The *Bias-Variance Decomposition** characterizes ML uncertainty.

Irreducible Error. The extent to which variation in the data defies explanation by any model.

Bias. The extent to which *a given model* fails to explain the phenomenologically-based structure of the data.

Variance. The extent to which *a given model* explains the "structure" of the idiosyncratic error of a given sample. (Overfitting.)

$$
\begin{aligned}
Err(x_0) &= \sigma_\varepsilon^2 \\
&\quad + \left[\mathcal{E}\hat{f}(x_0) - f(x_0)\right]^2 \\
&\quad + \left[\hat{f}(x_0) - \mathcal{E}\hat{f}(x_0)\right]^2
\end{aligned}
$$



* T. Hastie, R. Tibshirian, J. Friedman, *Elements of Statistical Learning*, Springer, 2009, https://bityl.co/5qZD.

Metalog distributions provide a highly-flexible approach to analysis of residual errors from ML-model fits.

Metalog distributions' formulations are based on the quantile function $M_n(y)$ defined as

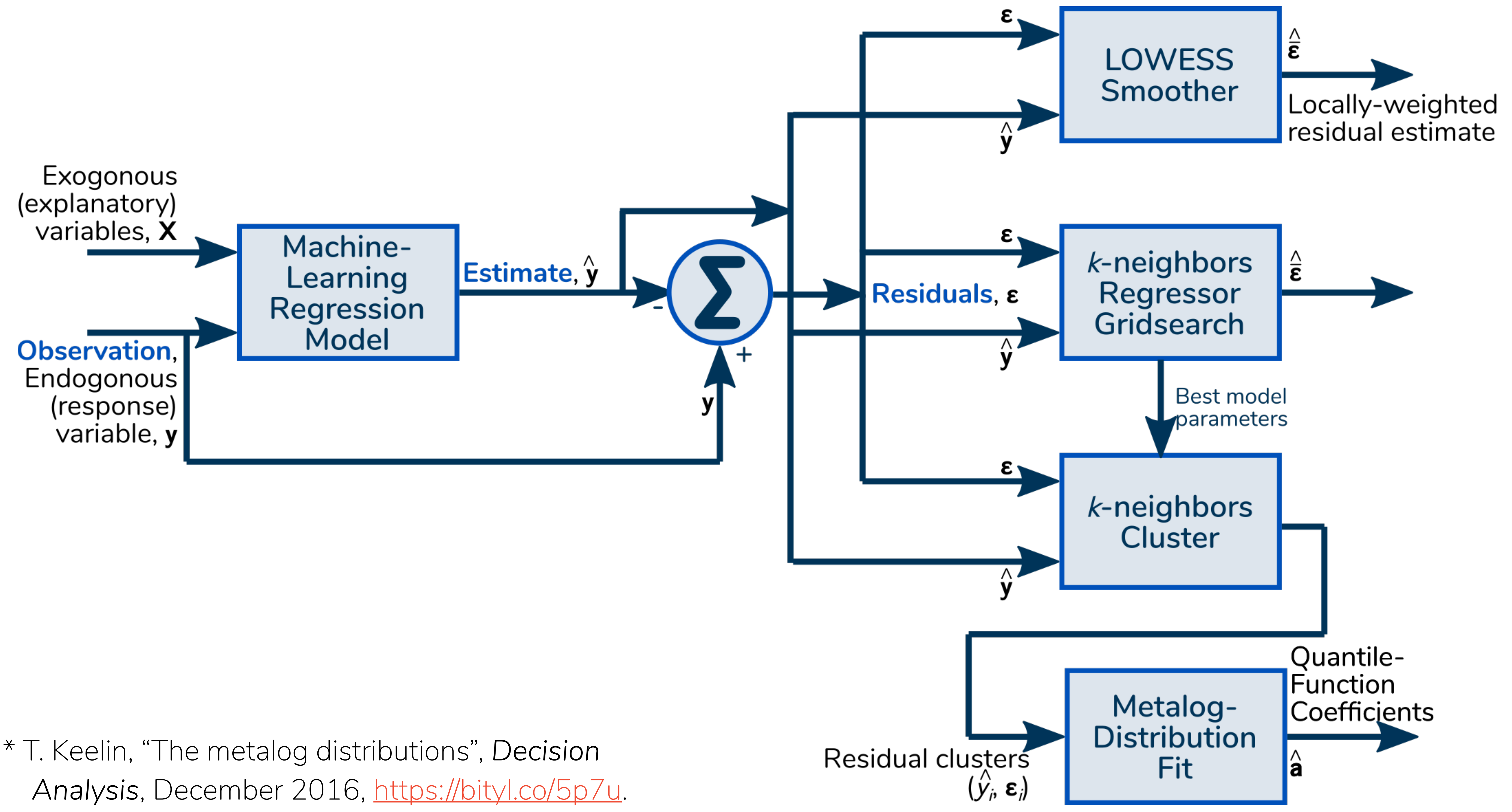
$$x = M_n(y) \Leftrightarrow y = Pr\{\mathscr{X} \leq x\} = P_{\mathscr{X}}(x).$$

The quantile function is estimated using a series

$$\hat{M}_n(y) = \sum_{\nu=1}^{\frac{n}{2}} \left(a_{2\nu-1}\left(y - \frac{1}{2}\right)^{2\nu-1} \right.$$
$$\left. + a_{2\nu}\left(y - \frac{1}{2}\right)^{2\nu-1} \ln\left(\frac{y}{1-y}\right)\right).$$

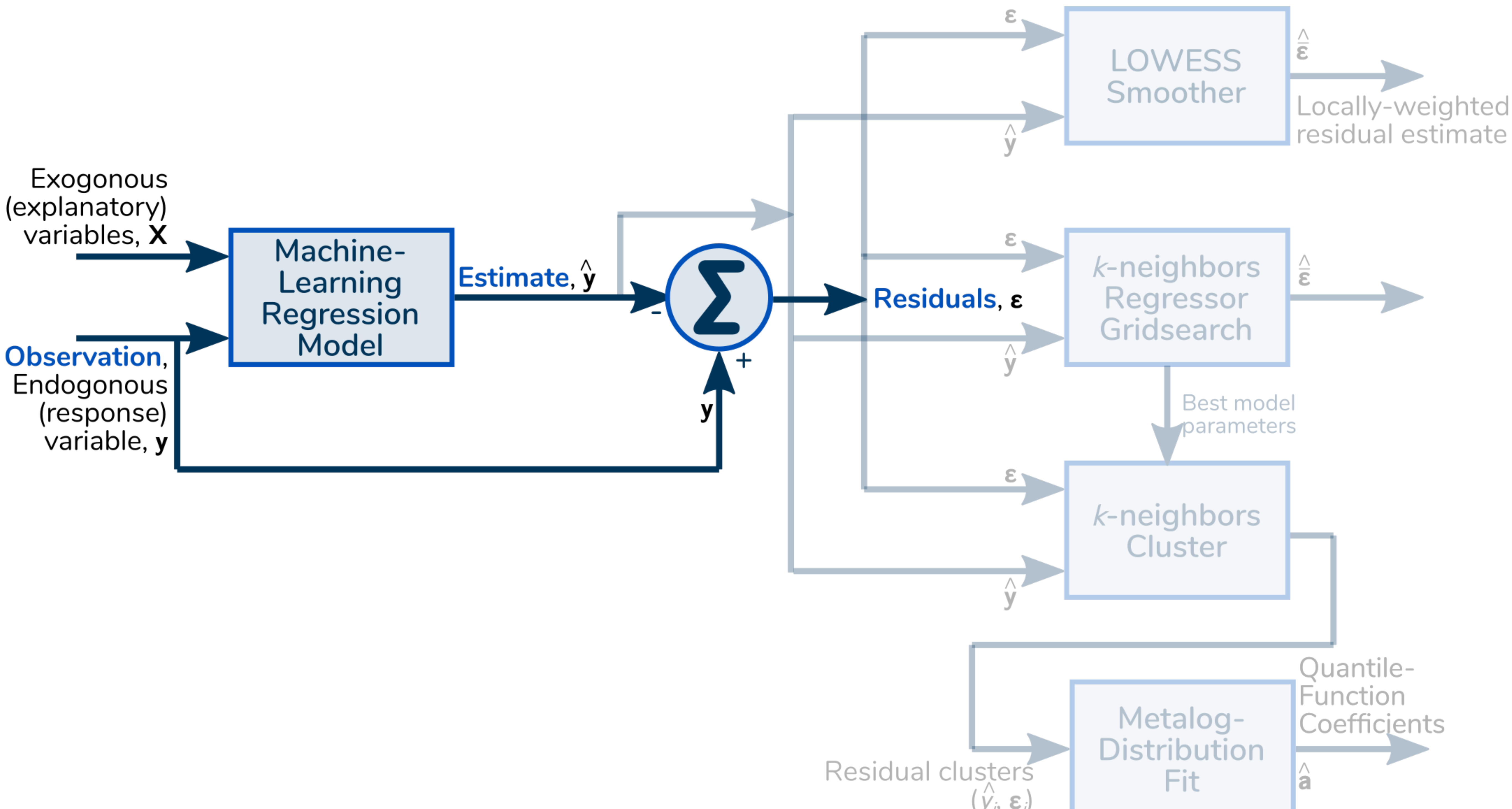
* T. W. Keelin, "The metalog distributions", *Decision Analysis*, December 2106, https://t.ly/WSTG.

Metalog distributions* lead to analytical expressions for residual errors from ML-model fits.

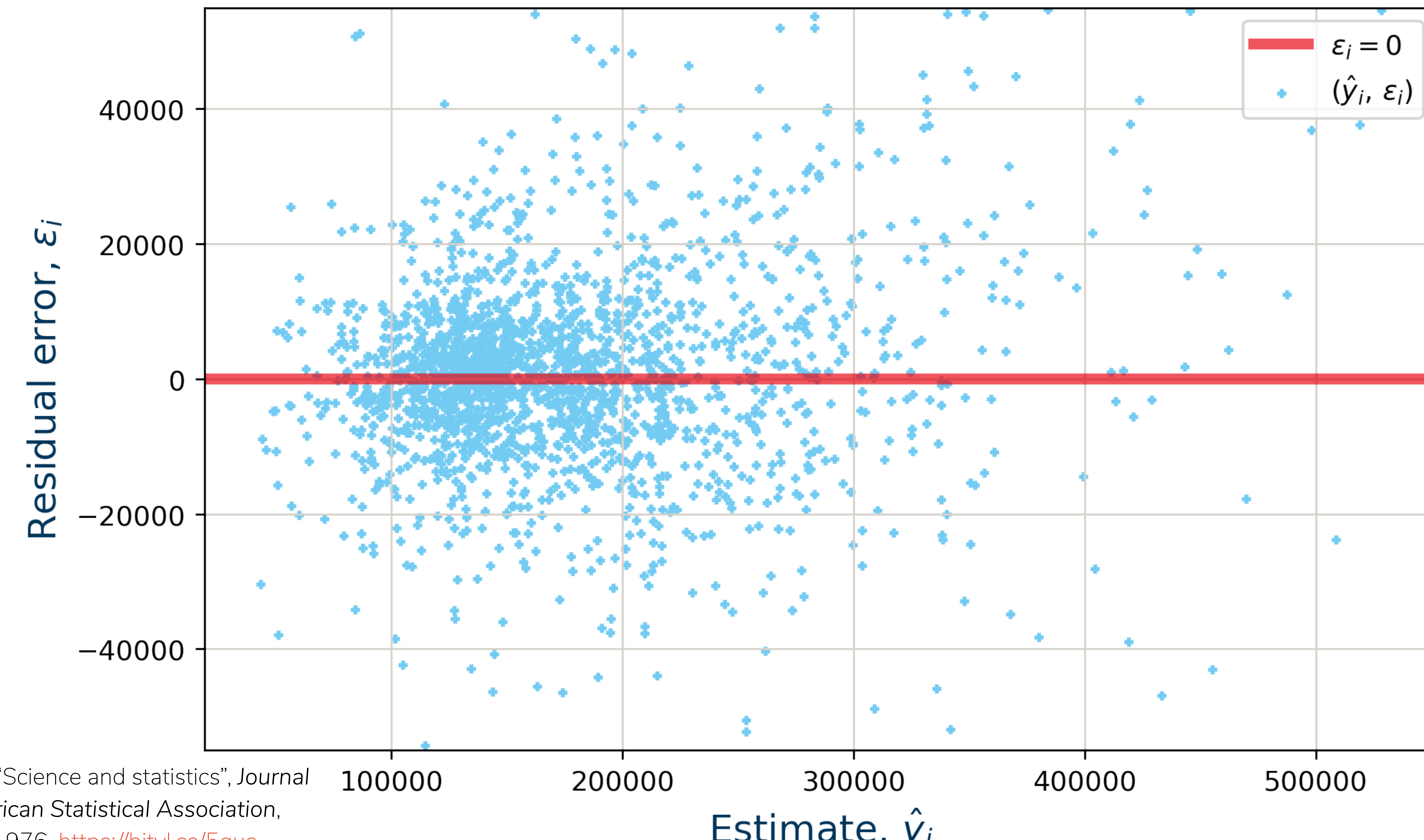


* T. Keelin, "The metalog distributions", *Decision Analysis*, December 2016, https://bityl.co/5p7u.

ML focuses on fitting statistical models to explanatory data.

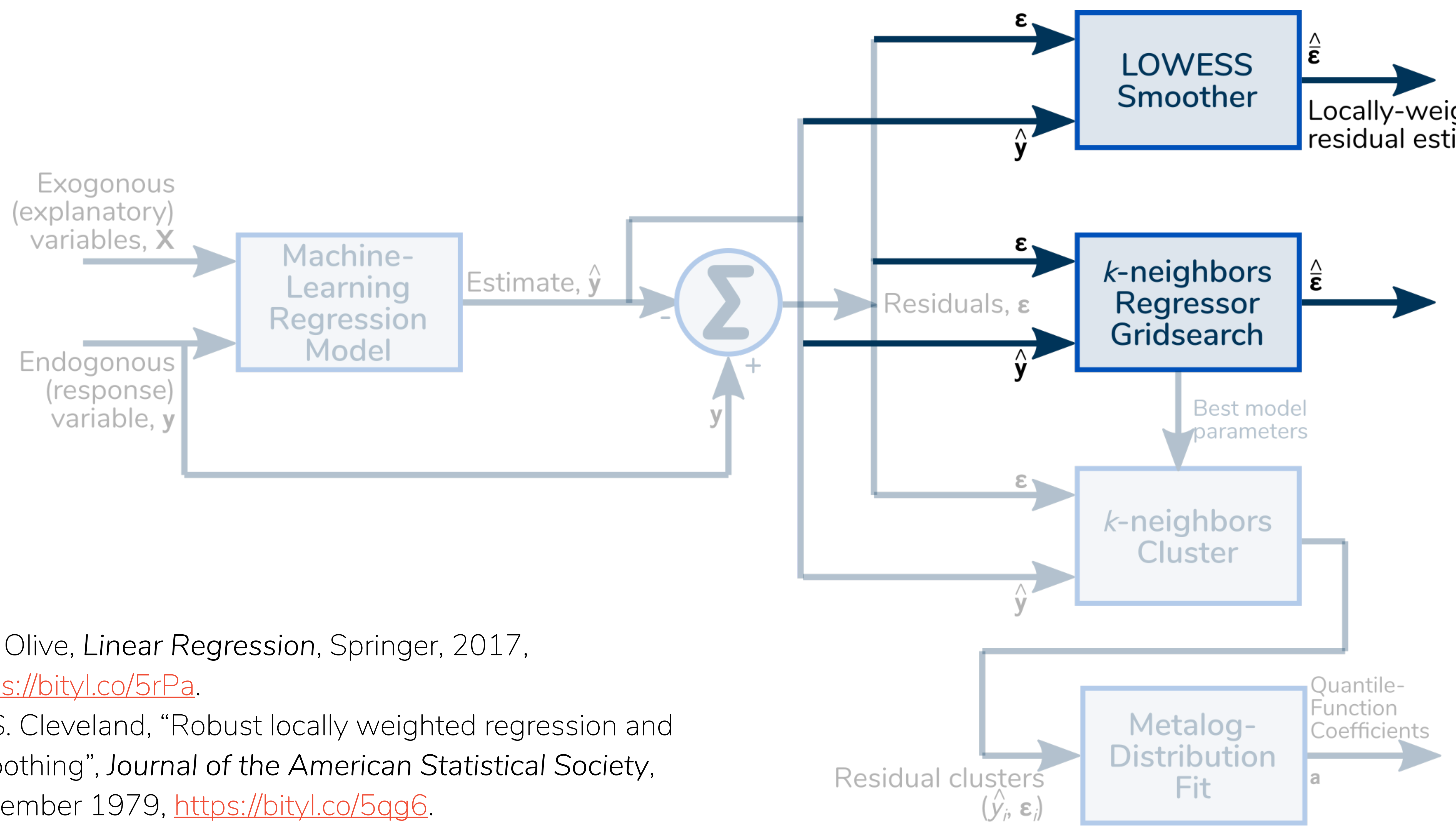


Estimates from any predictive model are accompanied by unexplained variation.*



* G. E. P. Box, "Science and statistics", *Journal of the American Statistical Association*, December 1976, https://bityl.co/5qua.

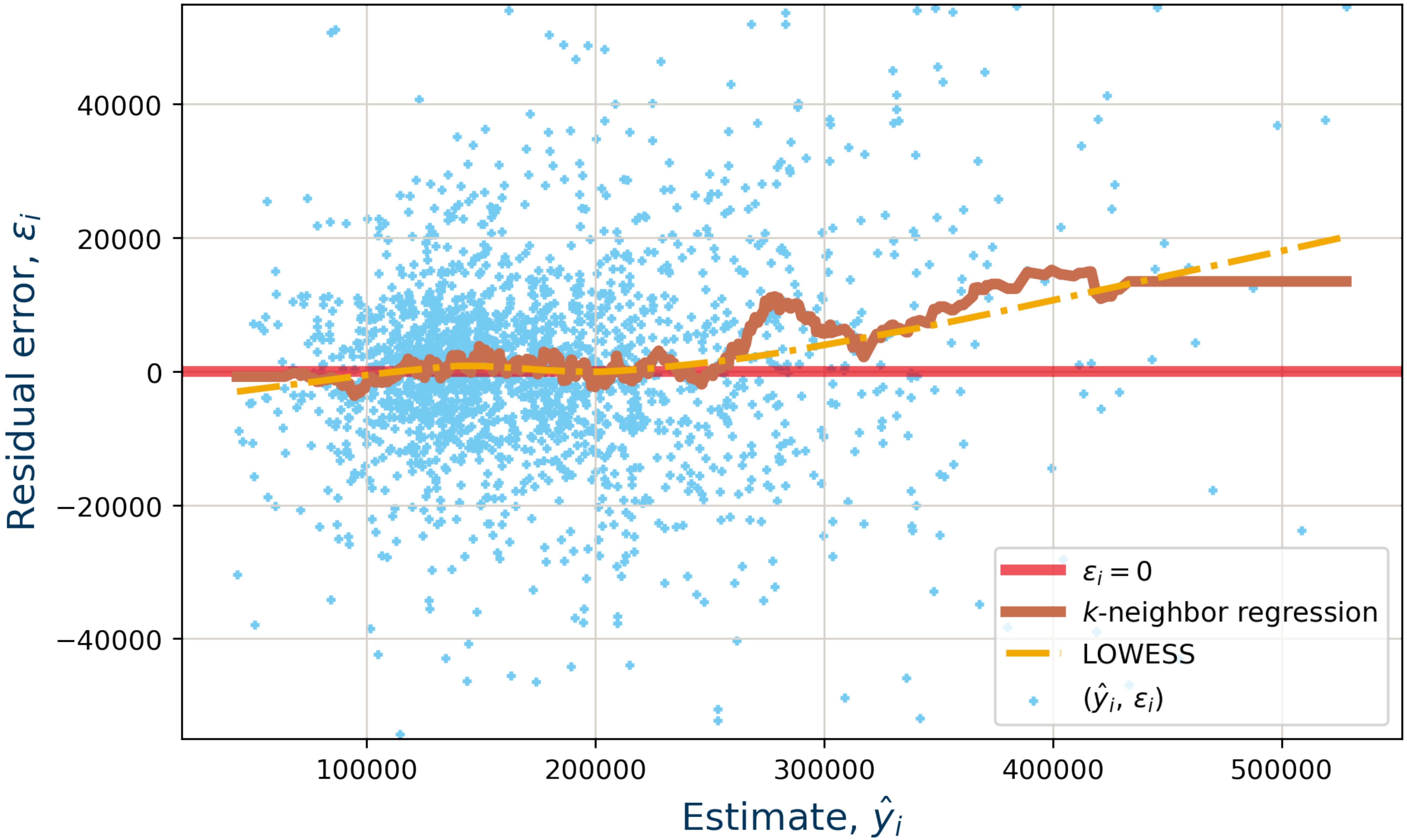
Locally-Weighted Scatterplot Smoothing (LOWESS)*† characterizes bias.



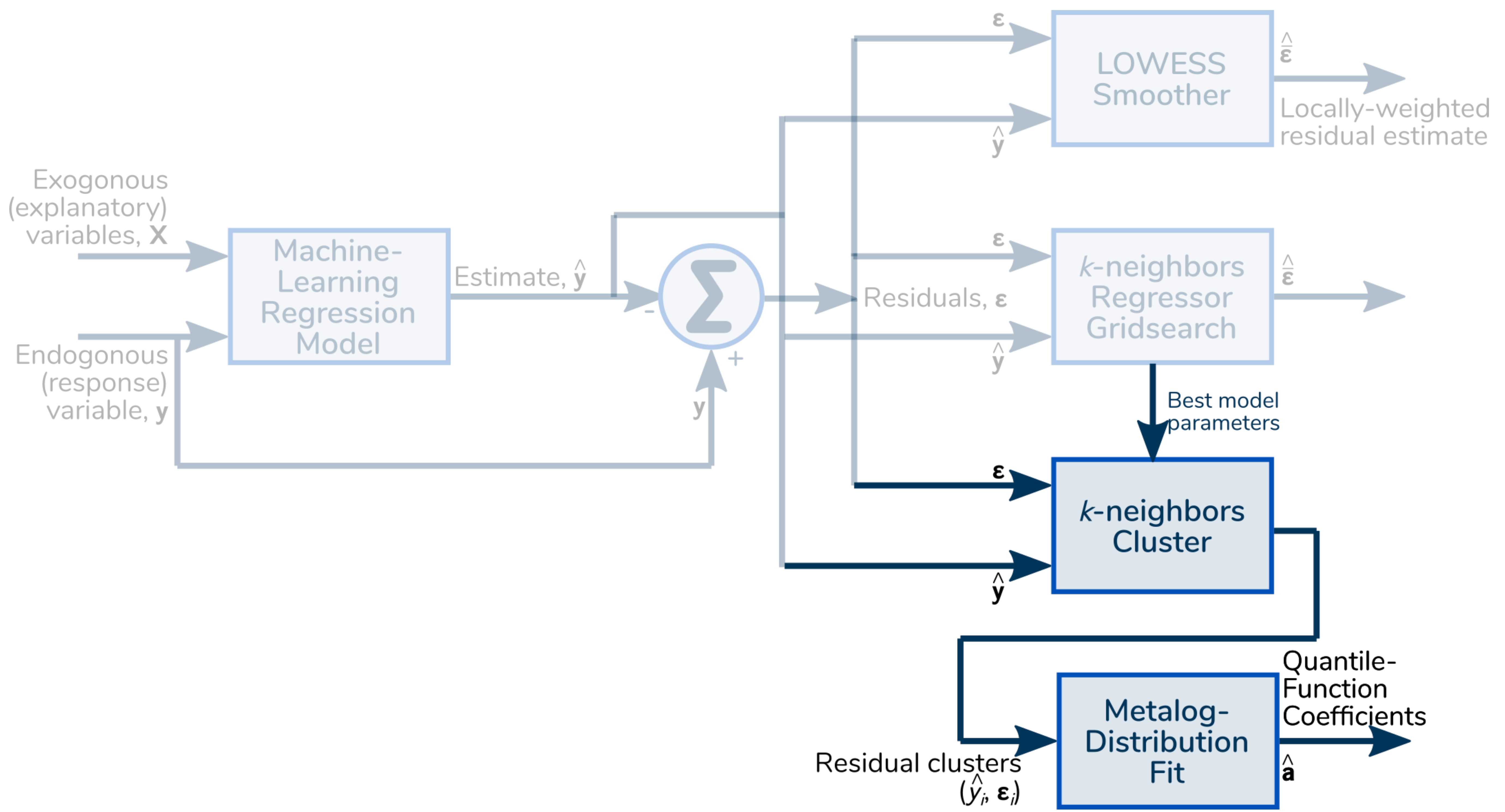
* D. J. Olive, *Linear Regression*, Springer, 2017, https://bityl.co/5rPa.

† W. S. Cleveland, "Robust locally weighted regression and smoothing", *Journal of the American Statistical Society*, December 1979, https://bityl.co/5qg6.

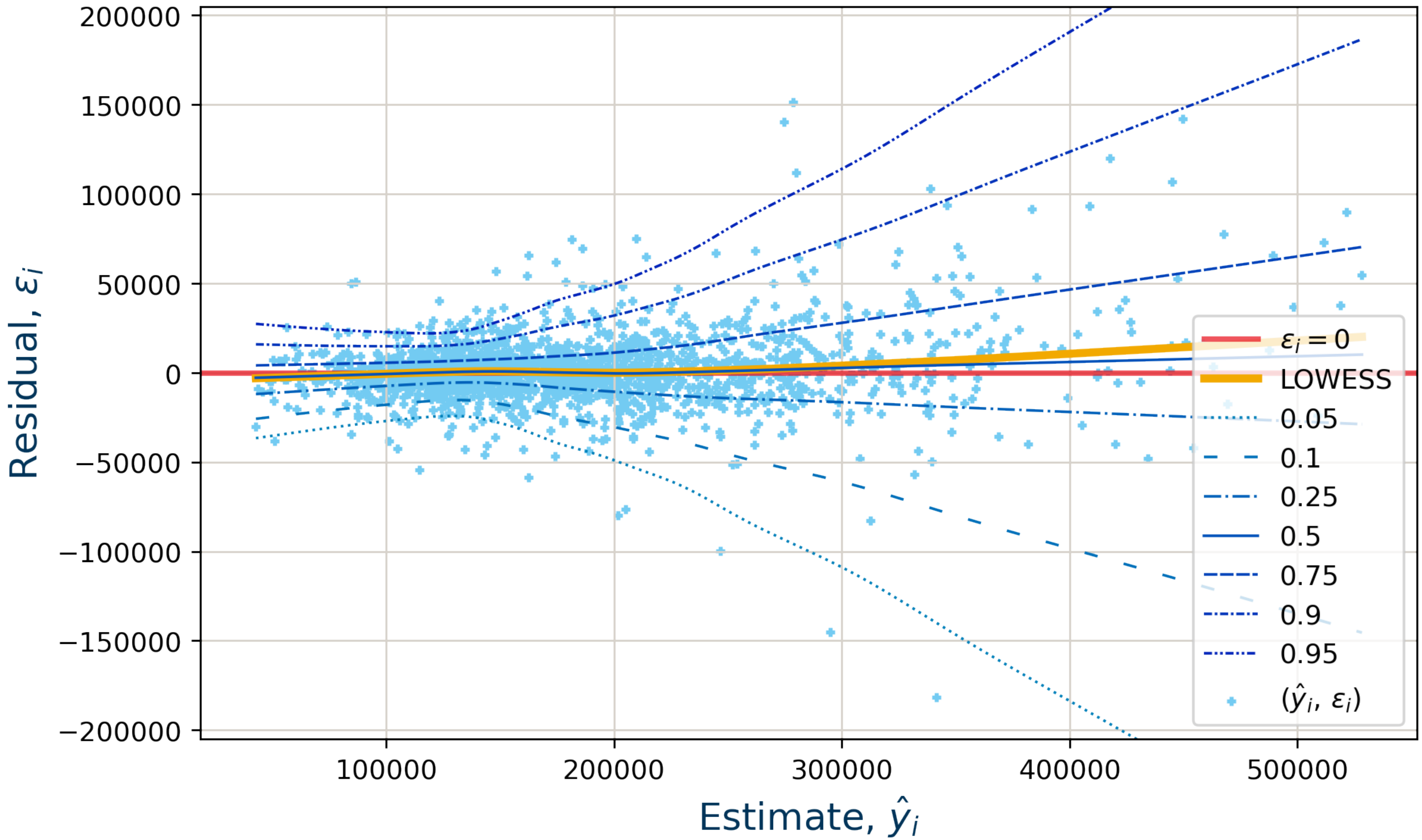
The *k*-neighbors regression ML method approximates LOWESS.



k-neighbor clustering provides a residual-plot neighborhood within which to locally fit a metalog distribution.



Local metalog fits to ML residuals produce conditional probability distributions.



Where to from here?

- Use metalog conditional densities to calculate statistics that quantify the difference between residuals for training, test data sets.
  - Kolmogorov-Smirnov test.
  - Kullback-Leibler divergence.
  - Area Under Receiver Operating Characteristic.

- Calculate irreducible uncertainty.
  - Begin with Cramer-Rao Lower Bound (CRLB) as a proxy.

- Apply method to decision-analysis case studies.

Uncertainty Research, LLC

Neil A. Hamlett, D.Sc., MBA
Strategic Consulting Data Scientist, Uncertainty Research, LLC
neil.hamlett@uncertainty-research.science
https://www.linkedin.com/in/neil-hamlett-strategic-quant/
https://hamlett-neil-ur.github.io/