Linear Regression Through the Origin

Dyussenov Nuraly
Supervisor: Jozsef Mala

December 15, 2023

Abstract

This thesis explores the properties of simple linear regression trough the origin with standard OLS assumptions, and effectively compares it to the full model with an intercept term included. The goal is to summarize the existing literature about RTO and identify conditions favoring RTO's performance.

The last section of the paper focuses on comparison of RTO with a full model, which includes intercept. For the practical comparison, we assume that underlying true relationship is $y_i = \beta_1 x_i + \beta_0$, where β_0, β_1 are intercept and slope terms, and x_i, y_i are i-th regressor and regressand, respectively. Then we add normally distributed, centered error terms to fulfill the OLS assumptions. To compare two models, several factors are observed: Akaike Information Criterion, Bayesian Information Criterion and sum of squared deviations of fitted values from values on true underlying model of a generated dataset (later reffered to as SSD).

The study reveals that, on average, RTO provides a superior fit [1] if an intercept of an underlying true model, from where the data comes from, is small. Augmenting OLS assumptions with a zero true intercept validates RTO as generally preferred model, as it is unbiased with lower variance. Despite RTO's bias in datasets with intercept terms, its simplicity mitigates bias for very small intercept values. In contrast, the full OLS model, with an intercept estimate, amplifies overall model variability:

In summary, this thesis offers practical insights into optimal conditions for RTO application, contributing to both theoretical understanding and applied statistical modeling guidance.

---

[1] In terms of AIC, BIC, and SSD