

Exploring the Stability of Software with Time-Series Cross-Sectional Data

Jukka Ruohonen, Sami Hyrynsalmi, Ville Leppänen
University of Turku, Finland

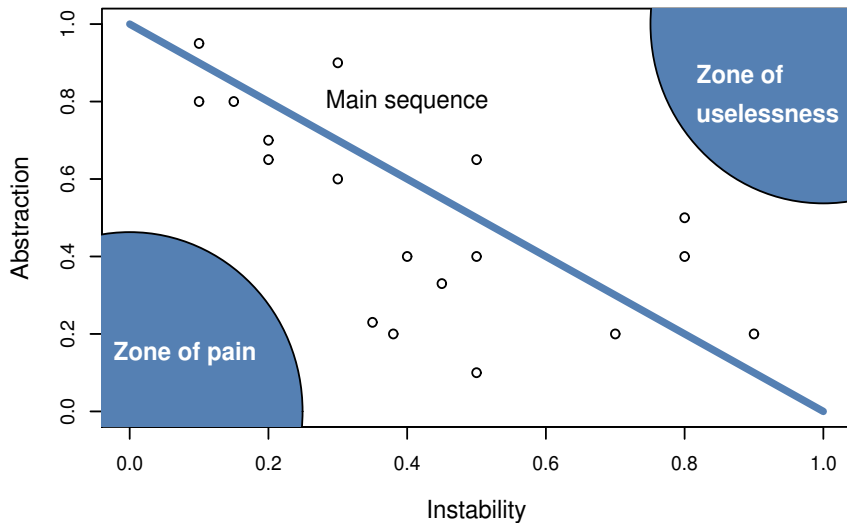
Objectives

- ▶ **Primary:** empirical evaluation of Robert C. Martin's object-oriented design principle with a case study
 - ⇒ Data: 73 versions and 14 packages from Vaadin
 - ⇒ Martin's argument does not hold (negative result)
- ▶ **Secondary:** explore the use of time-series cross-sectional (TSCS) modeling in the software metrics context
 - ⇒ Prolific because of the longitudinal dimension (cf. drift)
 - ⇒ Challenging to estimate robustly

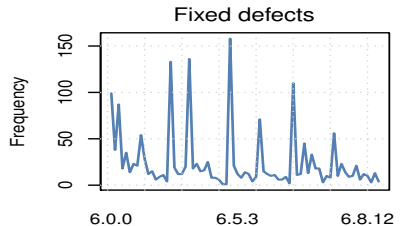
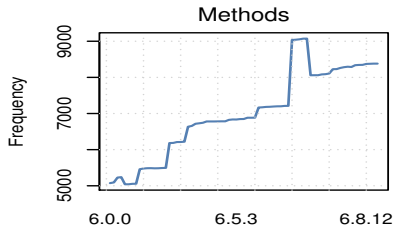
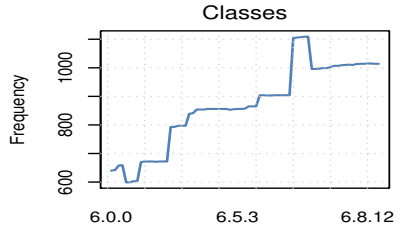
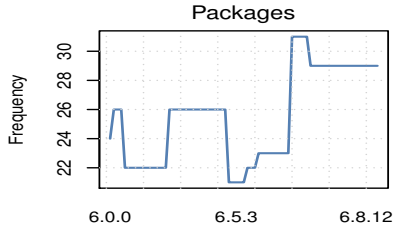
Martin's (1994) Metrics

- ▶ Object-oriented package-level design metrics
 - ⇒ Instability \equiv Many dependencies
 - ⇒ Related to the concept of logical stability (“ripples”)
- ▶ Stable (no coupling) classes and packages should be abstract
- ▶ A class (package) should depend towards stable counterparts
- ▶ Abstraction = $\frac{\text{Abstract classes}}{\text{All classes}}$
- ▶ Instability = $\frac{\text{Outward dependencies}}{\text{Inward dependencies} + \text{Outward dependencies}}$

Martin's Argument



The Longitudinal Dimension



Time-Series Cross-Sectional Estimation

- ▶ Two dimensions: versions (time) \times packages (cross-sections)
- ▶ Can be estimated with least squares and dummy variables
- ▶ Many statistical challenges
 - \implies Correlations across time and between cross-sections
- ▶ Encapsulation and inheritance were used as controls
- ▶ Martin's argument also requires aggregation
- ▶ Hierarchical estimation across architecture levels?

Thank you

Questions?