

TECHNICAL LEVERAGE ANALYSIS IN THE PYTHON ECOSYSTEM

*Revision under review at the Journal of Empirical Software Engineering (EMSE)

WHAT IS TECHNICAL LEVERAGE?

A novel metric for measuring dependencies: Ratio between the **dependency codes** and the **original codes** in a software package (Massacci and Pashchenko, ICSE'2021).

$$\text{Technical Leverage} = \frac{\text{LOC of Dependency code}}{\text{LOC of Own code}}$$



DATASET



21,205
package versions
in PyPI

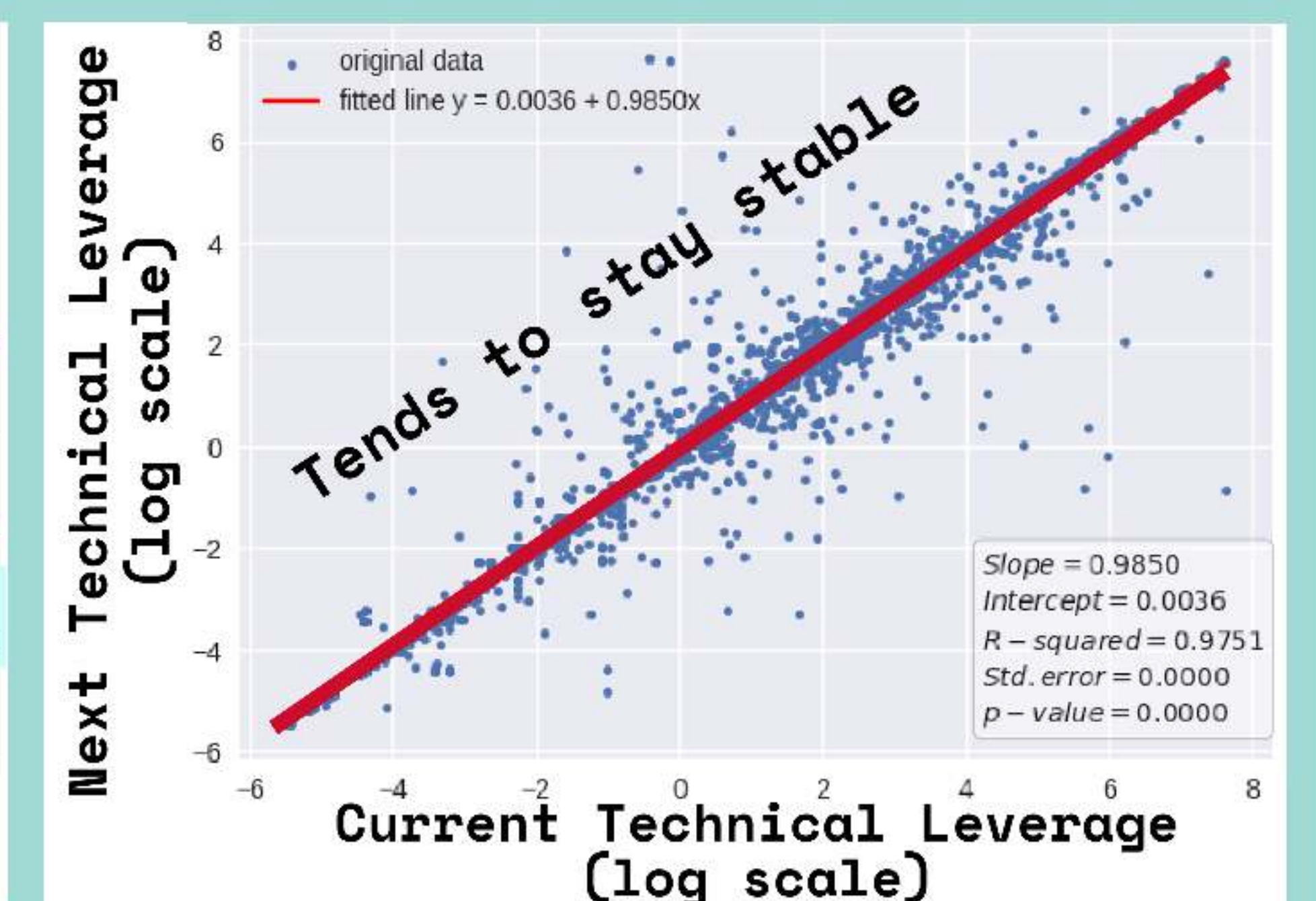
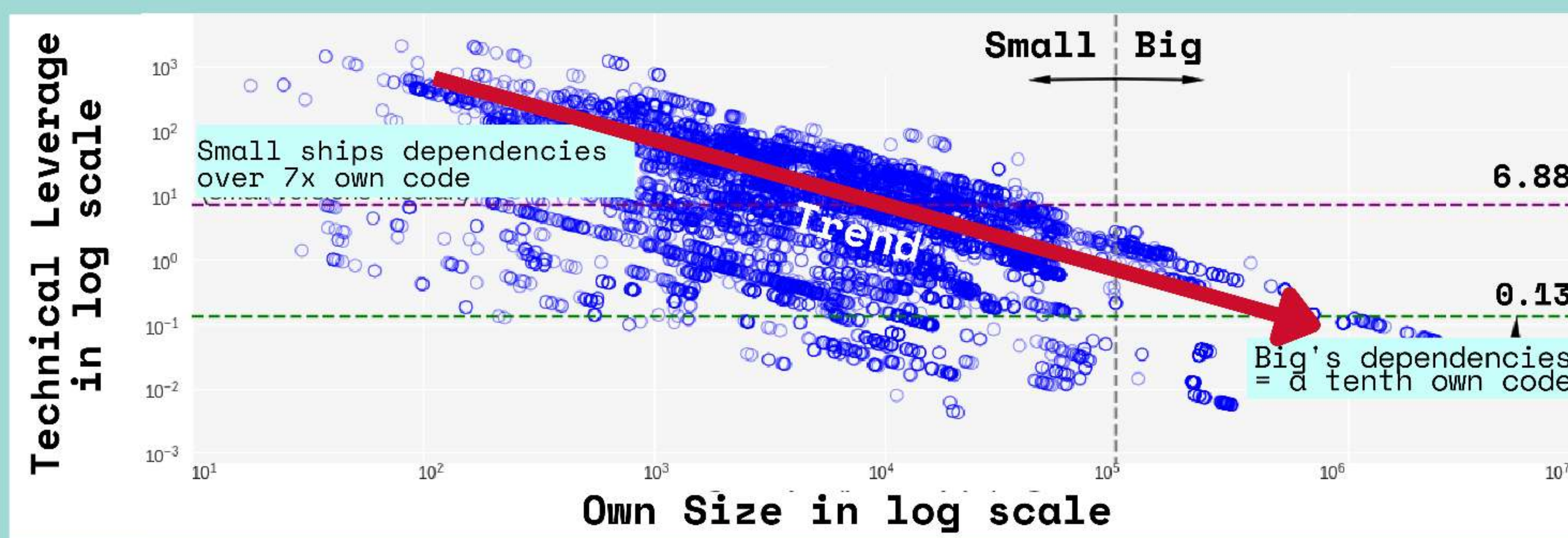
FROM

482
different top
Python packages

ANALYSIS PROCEDURE

1. Statistical Analysis
2. Mathematical Estimations
3. Simulations

DATA IN DETAILS



RQ1

How is the Python ecosystem regarding technical leverage and developers' behavior?

As in Java, Python developers also tend to ship a lot of other people's code.

RQ2

How does the technical leverage metric change across versions in a package?

If you are highly leveraged, you will stay so.

PROBABILITIES OF GETTING SAFE PACKAGE VERSIONS

Package Group	Standard Calculation #VulnVersion/#Versions	Our Formula for What Developers Actually Experience	
		Not considering downloads	Considering downloads
No dependencies (TL = 0)	78.85%	89.41%	78.60%
Below industry avg. (0 < TL <= 4)	71.89%	81.01%	77.05%
Above industry avg. (TL > 4)	68.93%	89.91%	81.65%

RQ3

How does the technical leverage metric affect the risk of having vulnerabilities in Python ecosystem?

The **CHANCE** of getting a **SAFE PACKAGE VERSION** is **HIGHER** than just reporting the percentage of vulnerable versions.

FUTURE WORKS

1. How do packages' security states evolve over time?
2. How to do security MSR research with the available knowledge at a certain point in time?



RANINDYA PARAMITHA¹
ranindya.paramitha@unitn.it

(1) Università di Trento, Italy; (2) Vrije Universiteit Amsterdam, The Netherlands

FABIO MASSACCI^{1,2}
fabio.massacci@ieee.org

