Chair of Financial Management and Capital Markets
TUM School of Management
Technical University of Munich

# From Paper to Code: Implementing Asset Pricing Anomalies with LSEG/Refinitiv Data in Python

Keywords: Empirical Asset Pricing, Return Predictors, Automation, Datastream, LSEG/Refinitiv Workspace, Python Programming, Error Handling

## Project description

Understanding which firm characteristics predict stock returns is at the heart of modern empirical asset pricing research. Over the past decades, hundreds of so-called “anomalies” or cross-sectional return predictors have been proposed in academic papers. However, reproducing these predictors in a consistent and scalable way is technically demanding, as it requires carefully following detailed variable definitions, combining multiple large data sources, and implementing efficient calculation routines.

In this project, the student will develop a Python-based environment that systematically codes and computes a broad set of cross-sectional return predictors using data from LSEG/Refinitiv (e.g. Datastream, IBES, Worldscope). The starting point will be a comprehensive PDF that lists a wide range of anomalies together with precise instructions on how each predictor is constructed (for example, which balance sheet, income statement, analyst estimates, or price variables enter the formula and how they need to be transformed). Based on this document, the student will translate the textual variable definitions into robust Python code.

A key focus of the project will be designing a clean and modular code structure that avoids redundancy and allows for efficient parallelization. Together with the supervisors, the student will define reusable building blocks for frequently used operations (such as lagging variables, winsorizing, forming ratios, or computing rolling statistics) and design a pipeline that can handle many predictors in a unified way. The student will be provided with the necessary baseline datasets, including company fundamentals, analyst estimates, and stock price information, and will learn how to link and process these data sources at scale.

The ultimate goal is to create a Python environment that enables streamlined, fast recalculation of all implemented predictors with minimal manual intervention when the underlying data are updated. This includes transparent configuration options (e.g. sample definitions, frequency choices, lag structures) and clear documentation so that future researchers and student teams can easily extend the library with new anomalies or adapt it to different research projects.

Beyond the technical implementation, this IDP offers valuable preparation for careers at the intersection of finance, data, and technology. The skills acquired—working with large financial datasets, implementing academic factor definitions, writing production-ready Python code, and collaborating on a structured research software project—are directly relevant for positions as quantitative analysts or data scientists in asset management, investment banks, hedge funds, FinTechs, and consulting firms, as well as for a potential future research career in empirical finance.

Chair of Financial Management and Capital Markets
TUM School of Management
Technical University of Munich

## What we are looking for

- Strong analytical and project management skills
- Determination and passion for your areas of expertise
- Good Python programming skills
- Interest to work at the intersection of finance and IT
- 1 or 2 persons

## What we offer

- Knowledge in quantitative finance, corporate finance and machine learning
- Kick-off session including introduction to relevant finance and/or business topics
- Experience with IDPs
- Open dialogue and support
- Access to prime capital markets databases (Bloomberg, Datastream, Thomson Reuters, etc)
- Potential for publication and/or evaluation of future use cases
- Both single and group projects are possible

## Confidentiality of Code and Data

This IDP is based on proprietary financial data and internal research infrastructure provided by the chair. The Python code, scripts and documentation developed in this IDP are **confidential** and become the property of the chair. They **must not be shared** with third parties and **must not be published** in any public or semi-public repository (e.g. GitHub, GitLab, Bitbucket, Kaggle, etc.).

By participating in this project, you agree to comply with these confidentiality requirements.

## Interested?

Please send an e-mail with CV, academic transcript and your preference for this project to sebastian.mueller.hn@tum.de and to garvin.kruthof@tum.de.

## Questions?

In case of any (e.g. topic related) questions, please contact Prof. Dr. Sebastian Müller (sebastian.mueller.hn@tum.de) and/or Garvin Kruthof (garvin.kruthof@tum.de).