Manual Intervention and Expert Input Guide

Shaikh & Tonak (1994) Replication and Extension

September 28, 2025

Abstract

This document outlines the key areas in the replication and extension of Shaikh & Tonak (1994) that require manual intervention, expert judgment, or methodological choices. While the project strives for maximum automation and fidelity, certain aspects of the data and methodology necessitate human input.

1 Historical Replication (Phase 1)

1.1 Handling of the 1973 Capacity Utilization Gap

- **Issue:** The original book data for capacity utilization ('u') shows a value of 0.0 for the year 1973. This is a data anomaly that makes the profit rate calculation mathematically impossible for that year.
- **Intervention:** An expert decision was made to handle this anomaly.
- **Action Taken:** The value for 1973 was interpolated as the midpoint between the 1972 and 1974 values. This is a methodological choice to create a continuous series, but it's important to note that this is an alteration of the raw data. The original 'u=0.0' is preserved in the source data files for full transparency.

2 Modern Extension (Phase 2)

2.1 Industry Classification Correspondence (SIC to NAICS)

- **Issue:** The historical data is based on the Standard Industrial Classification (SIC) system, while modern data uses the North American Industry Classification System (NAICS).
- **Intervention:** A mapping between SIC and NAICS codes is required to align the historical and modern datasets.
- **Action Taken:** A correspondence file (e.g., `config/industry_correspondences/st_naics_corr` was created. This file requires expert review, especially for broad categories like "Services," where the mapping can be ambiguous.

2.2 Assumptions for Modern Surplus Product (SP)

- **Issue:** A direct, one-to-one modern equivalent of the book's SP calculation is not available due to changes in NIPA accounting.
- **Intervention:** An assumption is needed to scale a modern data series (like corporate profits) to be comparable to the historical SP.
- **Action Taken:** A default annual nominal growth assumption (e.g., 3%) is used to scale modern corporate profits. This is a significant modeling choice and should be treated as an expert-editable parameter.

2.3 Estimation of Modern Capital Stock (K)

- **Issue:** Similar to SP, a direct modern equivalent of the book's capital stock series ('K' and 'KK') is not straightforward.
- **Intervention:** A method must be chosen to estimate the modern capital stock in a way that is consistent with the historical data.
- **Action Taken:** The modern capital stock is estimated based on the median historical ratio of Capital to Surplus Product (K/SP). This is a reasonable but coarse approximation and represents another key expert-editable parameter.

2.4 Definitional Ambiguities

- **Issue:** Some variables in the book, like the capital growth rate ('gK'), are not defined with sufficient precision to be unambiguously replicated.
- **Intervention:** An interpretation of the likely definition is required, based on the available text and data.
- **Action Taken:** The replication proceeds with the most plausible interpretation (e.g., that 'gK' is based on 'I/K'), but this is flagged as an area where the replication may not be "perfect" until more precise definitions are discovered.

3 Conclusion

The final dataset is a product of both automated, faithful replication and necessary, documented human interventions. This guide serves as a transparent record of these interventions, allowing future researchers to understand the methodological choices made and to explore alternative assumptions by modifying the relevant configuration files and parameters.