Cross-Sectional Momentum on S&P 500 (Kaggle)

Conglin BAO

Jan 21, 2026

Contents

0.1 Introduction

We analyze the Kaggle dataset "S&P 500 stock data" (camnugent/sandp500), which ~5 years of daily open, high, low, and close prices and trading volume (OHLCV) in `all_stocks_5yr.csv`. Our objectives are to:

1. Read raw stock price data using relative paths and perform necessary data cleaning.
2. Aggregate daily prices into monthly log returns and create a tidy dataset for analysis.
3. Construct a 12–1 cross-sectional momentum signal.
4. Form decile portfolios and evaluate the long–short (Q10–Q1) strategy.
5. Visualize cumulative performance with and without transaction costs.

All data processing, modeling, and visualization steps are implemented in separate, reproducible R scripts and sourced below.

The script below reads the raw data from the `data/raw/` directory, performs cleaning and aggregation, and produces a tidy monthly return dataset saved to `data/clean/`.

This step constructs the 12–1 momentum signal, forms decile portfolios, and fits a simple statistical model (monthly cross-sectional regression) to evaluate the relationship between momentum and subsequent returns.

0.1.1 Model output

We fit a monthly cross-sectional regression `ret1m ~ mom_12_1` and summarize the distribution of estimated slope coefficients (betas) across months.

Table 1: Summary of monthly cross-sectional regression slopes (ret1m ~ mom_12_1).

mean_beta	sd_beta	t_stat
0.0066	0.0657	0.6995

0.1.2 Table

Table 2: Sample coverage and momentum long–short performance (gross).

Sample start	Sample end	Unique tickers	Stock–month observations	Avg monthly long–short (log)	Annualized return	Annualized volatility	Sharpe ratio
2014-03	2018-02	505	29,507	0.45%	4.05%	17.38%	0.23

0.2 Table description:

The table summarizes the sample used in the 12–1 cross-sectional momentum test. We work from March 2014 to February 2018, covering 505 unique S&P-500 tickers and 29,507 stock–months after aggregating daily prices to monthly log returns (last trading day of each month). The equal-weight long–short portfolio (long top decile by momentum, short bottom decile) delivers an average monthly log return of 0.45%, which annualizes to 4.05% with 17.38% annualized volatility, yielding a Sharpe ratio of 0.23. These are gross results (no transaction costs) and reflect a short sample and a simple specification—no industry/size neutrality or turnover control—so they should be viewed as a reproducible baseline rather than an optimized strategy.

0.2.1 Figure

Cross–Sectional Momentum (12–1): Cumulative Return
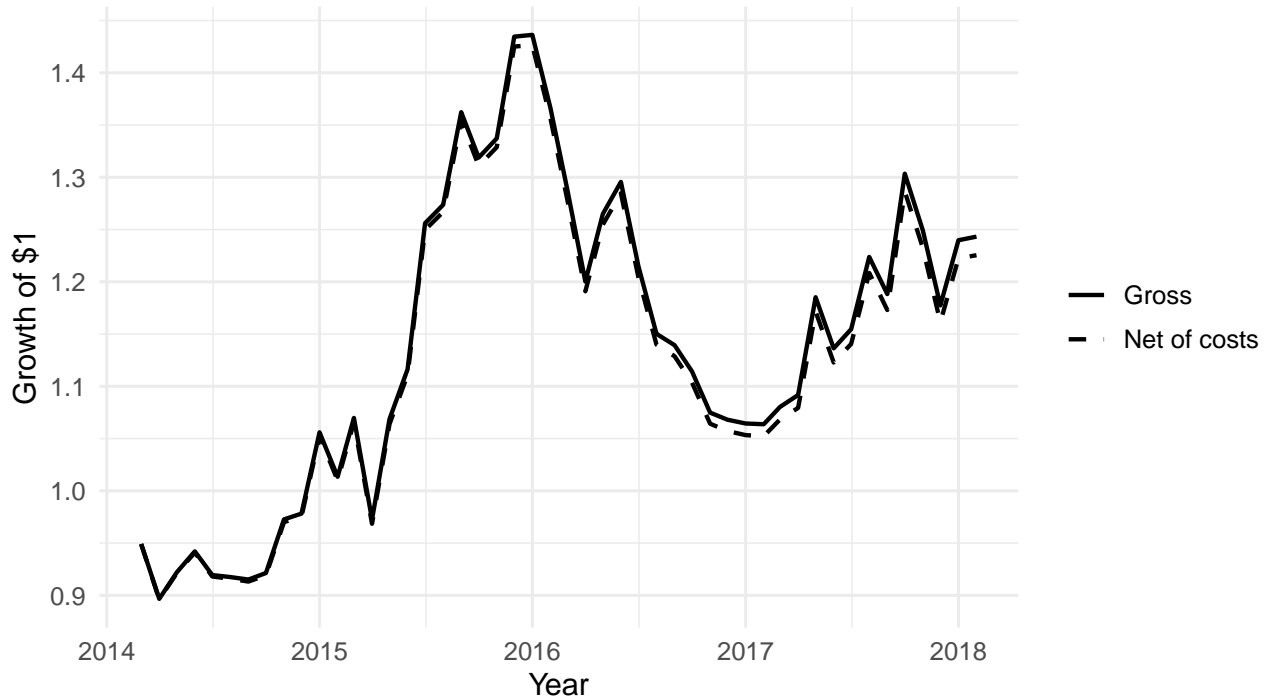
0.3 Figure description:

The figure plots the cumulative wealth of the monthly long–short momentum portfolio starting at $1. The gross curve (solid) and the net-of-costs curve (dashed; 3 bps per side, 50% monthly turnover) track each other closely—costs impose only a small drag under these assumptions. Performance dips in early 2014, rallies through 2015, reaches a local peak around early 2016, then experiences a drawdown into early 2017, followed by a partial recovery; by the end of the sample the portfolio stands near $1.24 gross with a slightly lower net value. This pattern is consistent with momentum cycles over the period and highlights that costs matter but are not dominant with the simple settings used here.