Rediscover Predictability: Information from the Relative Prices of Long-term and Short-term Dividends

Alec Zhao, Zihao Liu, Yu Guo, Coco Qu

March 2, 2024

Abstract

In this project, we endeavor to replicate and validate the findings of Li Wang (2023) in the study "Rediscover Predictability: Information from the Relative Prices of Long-term and Short-term Dividends." The core objective is to reproduce the results presented in Table 1, which provides a statistical summary of the price ratio (pr_t) and the log price-dividend ratio (pd_t) , and Table 2, which details the regression outcomes predicting the 12-month returns of the S&P 500 index using pr_t, pd_t, and their residuals.

The replication employs data from Bloomberg for S&P 500 futures and utilizes the Fama-Bliss database to closely follow the original study's methodology. We reconstruct the statistical relationships and analyze the regression coefficients (β) and the adjusted R-squared values to assess the predictability of stock returns based on the relative pricing of dividends. Our efforts focus on matching the precision of the original work, acknowledging the absence of standard errors in our regression outputs.

1 Introduction

This project replicates the pivotal findings in the article. Our objective is to replicate Table 1 and Table 2. Utilizing Bloomberg data for S&P 500 futures and zero-coupon yields from the Federal Reserve, we not only include analysis from January 1988 to June 2017 time frame, but also include data up to January 2024. This replication effort not only underscores the importance of empirical validation in financial research but also confronts the challenges of data availability and methodological adaptation inherent in such scholarly pursuits.

2 Replication Tables

Table 1 is a statistical summary of pr_t, pd_t. pr_t is price ratio of long- to short-term dividend prices, which is calculated by

$$pr_t = ln\left(\frac{\text{Price of Long-term Dividends}}{\text{Price of Short-term Dividends}}\right) = \ln\left(\frac{P^{T+}}{P^{T-}}\right) \tag{1}$$

where, P^{T+} is the multiplication of price of ZCB (zero coupon bonds) that matures in T periods and futures price that is the Q-expectation of future stock price. P^{T-} is the portion of the index value that is not accounted for by the futures component, which is S&P 500 stock prices minus P^{T+}

pd_t is log price-dividend ratio, which is calculated by

$$pd_t = \ln\left(\frac{P_t}{D_t}\right) \tag{2}$$

Here is the replicated Table1 in the article

	obs	mean	std	min	25%	50%	75%	max	ρ	Correlation
pr	354	3.760	1.024	2.348	2.926	3.323	4.657	6.179	0.962	0.137
pd	354	3.868	0.304	3.241	3.602	3.887	4.034	4.551	0.990	0.137

Table 1: Summary Statistics

Here is the Table1 with updated numbers, which means using data until 2024-01-31

	obs	mean	std	min	25%	50%	75%	max	ρ	Correlation
pr	433	3.857	1.058	2.348	2.994	3.515	4.697	7.104	0.958	0.192
pd	433	3.904	0.291	3.241	3.763	3.928	4.089	4.551	0.988	0.192

Table 1: Updated Summary Statistics

Table 2 reports the results of the regression that predicts the 12-month S&P 500 index return via pr_t, pd_t, and their residuals after projecting on them.

Here is the replicated Table2 in the article

	pr_t	pd_t	ϵ_t^{pr}	ϵ_t^{pd}
β	0.465	-0.500	0.012	0.203
R^2	0.054	0.106	0.155	0.120

Table 2: One-year Return Prediction

Here is the Table2 with updated numbers, which means using data until 2024-01-31

	pr_t	pd_t	ϵ_t^{pr}	ϵ_t^{pd}
β	0.465	-0.500	-0.008	0.208
R^2	0.054	0.106	0.208	0.124

Table 2: Updated One-year Return Prediction

3 Additional Summary Statistics

We also created summary statistics similar to Table1 for five source data: S&P500 index, dividend yield, 1-year future price, 1-year zero-coupon yields, and the corresponding discounting factor.

	obs	mean	std	min	25%	50%	75%	max	ρ
Dividend Yield	354	2.187	0.666	1.056	1.770	2.051	2.727	3.914	0.988
Index	354	1076.526	542.016	257.070	536.238	1117.885	1379.128	2423.410	0.996
Futures	354	1077.047	540.954	257.050	536.800	1120.375	1382.875	2421.000	0.996
1-Year Yield	354	3.489	2.637	0.099	0.663	3.655	5.542	9.658	0.996
1-Year Discount Factor	354	0.966	0.025	0.908	0.946	0.964	0.993	0.999	0.996

Table 3: Summary Statistics for Source Data

4 Data Sources

We calculated pr_t using the S&P500 indexes and its 1-year expiry future prices extracted from Bloomberg, as well as the 1-year zero-coupon yields extracted from the Federal Reserve's website.

$pd_t$is computed via the S&P500 dividend yield from Bloomberg. These are monthly data from Jan 1988 to Jun 2017 for replication and Jan 1988 to Jan 2024 for creating up-to-date tables.

5 Successes and Challenges

In our endeavor to replicate the seminal study, we successfully adhered to the paper's logic and replication methodology. Our access to and automation of data extraction were particularly noteworthy; we were able to seamlessly retrieve Bloomberg data, mirroring the process utilized by the original authors. Moreover, our replication of pd_t revealed only minimal discrepancies in comparison to the published results. Significantly, we achieved complete automation of the replication and table generation processes, all operationalized through a single 'doit' command. The thoroughness of our replication was affirmed by the successful passage of all unit tests within reasonable tolerance levels.

Despite these successes, we encountered several obstacles. Most notably, we were unable to access the Fama-Bliss database, which was the source of the 1-year zero-coupon yields in the original research. To circumvent this, we sourced the data from the Federal Reserve's website, although this introduced some methodological differences. The yields from the Federal Reserve are spot rates, which may not align with the non-standard rates used in the Fama-Bliss database, potentially leading to discrepancies in our calculated pr_t and the ultimate findings. Furthermore, the lack of a singular Bloomberg formula to extract all 1-year expiry S&P500 future prices from January 1988 to January 2024 posed a significant challenge. We were compelled to merge data extracted using two distinct Bloomberg formulas since one formula was limited to data only up to August 1997. This data fusion was necessary to ensure a complete dataset for the required period.