Final Project

Jonathan Cai, Keyi Chen, Adam Aldad, and Jean-Sébastien Gaultier

March 3, 2024

1 Introduction

In our final project, we replicated the first two tables of "Factor Demand and Factor Returns" by Cameron Peng and Chen Wang found here. The paper covers the persistence of factor demand and reveals the prevalence of factor rebalancing; We focus on the paper's discussion of factor demands. Table 1 summarizes a sample of US domestic equity mutual funds from 1980 to 2019, and table 2 summarizes the distribution of factor betas for mutual funds.

2 Replicating Table 1

2.1 Retrieving the Data

We pulled our data from WRDS' Monthly Total Net Assets, Returns, and Net Asset Values table found here. The paper only uses US domestic equity mutual funds in their analysis. Accordingly, we pulled data from WRDS' Style attributes for each fund table in order to filter the data. Our quarterly fund holdings data was pulled from the Thomson-Reuters Mutual Fund Holdings (s12) dataset

2.2 Cleaning the Data

We found this part of the replication process challenging - first, finding an optimal way to filter the data to only "US domestic equity" took various trials and errors. However, we discovered the **crsp_obj_cd** column of the **crsp.fund_style** table to be the best way to achieve this filter. Furthermore, when using the fund-level identifier **wficn**, we discovered a handful of occurrences where one **crsp_fundno** matches with multiple **wficn**. We suspected it could have to do something with delisting / merging of funds, and ultimately decided to drop these samples based on the descriptions in the paper. After obtaining the appropriate **wficn** values, we computed the yearly returns by first computing each fund's monthly returns. At this point, we ran into another issue: in our attempt replicate the paper's use of **mtna**, we noticed that not all **mtna** values are available. To solve this issue, we decided to use simple average instead because we expected different share classes of a given mutual fund to have similar returns. We then merged the TNA and yearly return information and got the following table:

Yearly Returns and Year End TNA

	wficn	year	crsp_{tna}	yret
0	100001	1990	169.57	0.03
1	100001	1991	330.03	0.30
2	100001	1992	596.27	0.06
3	100001	1993	857.67	0.06
4	100001	1994	876.19	-0.01

...

We then followed a similar process when preparing the S12 data. As such, we ran into tangentially similar issues: missing TNA values, minor discrepancies between **mflink1** and **mflink2**, and some

Dometic Equity

	wficn	year	assets	useq$_{tna}$
0	100001.00	1990	16957.00	161803.10
1	100001.00	1991	33003.00	314952.40
2	100001.00	1992	59627.00	578201.50
3	100001.00	1993	84286.00	821482.00
4	100001.00	1994	92961.00	896403.43

...

troubles with filtering the data to Domestic Equity. After solving these issues through various methods, we formed a table describing the S12 TNA data:

Finally, we merge the CRSP and S12 data, and ultimately create a close replication of table 1 of the paper.

3 Replicating Table 2

Then, we moved on to replicating table 2. We began by using the same merged CRSP and S12 data that we computed in table 1. We found much more success in replicating this table; the main challenge of this project was re-creating the original dataset. To obtain the Fama French Factor returns, we pulled factor returns **df_ff** from Kenneth R. French's Website. We then merged the CRSP dataset with a Fama-French dataset based on dates, calculated investment flow for each unique 'wficn' identifier as the percentage change in total net assets adjusted for returns, and merged this flow information back into the original dataset.