Replicating the Nozawa Corporate Bond Portfolios from He, Kelly, and Manela (2017)

Jesse Yan & Gabriel Ran

March 12, 2025

Abstract

This report documents our attempt to replicate the decile-sorted corporate bond portfolios originally constructed by Nozawa, as described in He, Kelly, and Manela (2017). We used the WRDS Bond Returns Database in combination with the latest WRDS (TRACE) Duration-Matched Treasury Returns (Open Source Bond Pricing, June 2024 release) contributed by Yoshio Nozawa to estimate yield spreads and replicate the subsequent returns. Notably, while the original HKM dataset covers 1970–2012, Nozawa's duration-matched Treasuries data is available from 2002 to 2023, limiting our replication to the overlapping period (2002–2012). Despite challenges such as missing yield data, negative amounts outstanding, and partial coverage of duration-matched Treasuries, our final decile returns closely track the original benchmark's direction and ordering.

1 Introduction

This project aims to reproduce the corporate bond decile portfolios sorted by yield spreads, as introduced by Nozawa and detailed in He, Kelly, and Manela (2017). Our primary data sources are:

- **WRDS Bond Returns Database** for corporate bond yield, return, and amount outstanding information, and
- **Latest WRDS (TRACE) Duration-Matched Treasury Returns** (Open Source Bond Pricing, June 2024 release) contributed by Yoshio Nozawa, which provides U.S. Treasury bond yields and returns aligned with the duration of each corporate bond.

While He, Kelly, and Manela (HKM) originally span 1970–2012, the duration-matched Treasury dataset starts in 2002 and continues through 2023. Consequently, the only overlapping interval between the HKM dataset and Nozawa's updated duration-matched data is roughly 2002–2012, restricting our replication to that period.

1.1 Data Coverage and Challenges

We observed that no corporate bonds had a corresponding yield in the first available month (August 2002) to compare with the duration-matched Treasury yield, so we started our replication from September 2002 onward. In other months, missing yield data occurred because some bonds had no trading volume; we were unable to calculate yields for those bonds given the limited methodological guidance in the literature. Forward-filling the yields did not help; it in fact made our replication worse. Additionally, some bonds exhibited negative amounts outstanding, which required data cleaning. As a result, while the replicated portfolios follow the overall direction and trend of the benchmark from 2002 to 2012, exact numerical matching was not achieved.

2 Methodology

2.1 Yield Spread Calculation

To calculate the yield spread for each corporate bond, we subtracted the duration-matched YTM (yield to maturity) of the Treasuries (obtained from Open Source Bond Pricing) from the corporate bond yield (obtained from WRDS). We then formed deciles based on these yield spreads each month. Not much other filtering was needed, as Nozawa already filtered according to his own requirements when generating the duration-matched Treasury YTM.

2.2 Portfolio Construction and Returns

After assigning each bond to a decile, we computed the subsequent month's return for that bond. Each decile's return was calculated as a weighted average of its constituent bonds' returns, using the bond's amount outstanding as the weight. Since no bonds had duration-matched Treasury yields in August 2002, the first month of return replication began in September 2002. The final sample ends in 2012 to match the last available year of the HKM data.

3 Data Overview

In this section, we include a time-series plot of the average corporate bond yield and the average duration-matched Treasury yield over the same sample period, highlighting the spread between the two. We also show a snapshot of the benchmark returns and the replication returns data. Each table displays the first 10 and last 10 rows of the respective datasets.

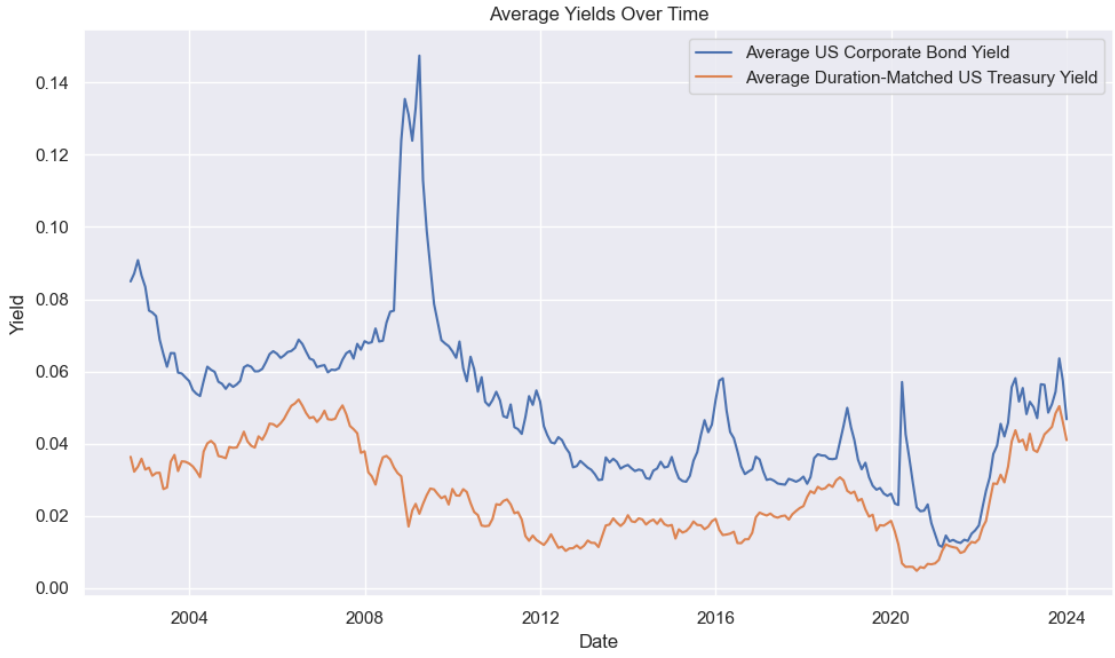

Figure 1: Average US Corporate Bond Yield and Duration-Matched US Treasury Yield Over Time

This figure plots the monthly average corporate bond yield and the average duration-matched US Treasury yield from 2002 to 2023. The gap between the two lines illustrates the yield spread driving our decile construction.

Table 1: Benchmark US Corporate Bond Returns

	11	12	13	14	15	16	17	18	19	20
2002-09-30	0.0228	0.0262	0.0132	0.0127	0.0103	0.0045	0.0026	-0.0006	0.0048	-0.0180
2002-10-31	0.0002	-0.0115	-0.0036	0.0019	0.0120	0.0052	-0.0015	0.0080	-0.0097	0.0084
2002-11-30	-0.0017	0.0130	0.0125	0.0240	0.0165	0.0215	0.0312	0.0339	0.0439	0.0438
2002-12-31	0.0201	0.0256	0.0310	0.0237	0.0192	0.0130	0.0209	0.0109	0.0025	0.0211
2003-01-31	0.0017	0.0040	0.0087	0.0092	0.0087	0.0126	0.0092	0.0091	0.0104	0.0460
2003-02-28	0.0154	0.0191	0.0183	0.0157	0.0140	0.0100	0.0082	0.0151	0.0126	0.0325
2003-03-31	0.0005	-0.0024	0.0041	0.0037	0.0116	0.0098	0.0073	0.0124	0.0168	0.0544
2003-04-30	0.0046	0.0123	0.0131	0.0121	0.0204	0.0212	0.0172	0.0313	0.0435	0.0968
2003-05-31	0.0328	0.0311	0.0304	0.0173	0.0202	0.0125	0.0123	0.0075	0.0049	0.0174
2003-06-30	-0.0006	-0.0011	-0.0028	0.0110	0.0031	0.0100	0.0077	0.0159	0.0219	0.0315
...	...	...	...	...	...	...	...	...	...	...
2011-03-31	0.0010	-0.0005	-0.0001	-0.0010	0.0018	0.0040	0.0047	0.0053	0.0046	0.0065
2011-04-30	0.0087	0.0149	0.0149	0.0198	0.0168	0.0126	0.0092	0.0092	0.0092	0.0126
2011-05-31	0.0103	0.0159	0.0165	0.0163	0.0123	0.0133	0.0067	0.0054	0.0043	0.0084
2011-06-30	-0.0008	-0.0052	-0.0052	-0.0061	-0.0073	-0.0058	-0.0025	-0.0029	-0.0001	-0.0044
2011-07-31	0.0122	0.0220	0.0206	0.0223	0.0168	0.0134	0.0085	0.0103	0.0081	0.0044
2011-08-31	0.0126	0.0274	0.0195	0.0140	-0.0033	-0.0095	-0.0122	-0.0105	-0.0101	-0.0286
2011-09-30	0.0036	0.0196	0.0064	0.0089	-0.0067	-0.0055	-0.0059	-0.0083	-0.0073	-0.0202
2011-10-31	0.0030	0.0044	0.0107	0.0122	0.0178	0.0222	0.0300	0.0301	0.0342	0.0508
2011-11-30	-0.0007	-0.0041	-0.0106	-0.0074	-0.0164	-0.0189	-0.0299	-0.0116	-0.0197	-0.0149
2011-12-31	0.0057	0.0141	0.0214	0.0194	0.0273	0.0307	0.0253	0.0189	0.0210	0.0270

This table provides a similar head/tail snapshot of the benchmark returns from He, Kelly, and Manela (2017). It serves as the comparison point for evaluating our replicated portfolios, restricted to the overlapping 2002–2012 period.

Table 2: Replication Portfolio Returns

	11	12	13	14	15	16	17	18	19	20
2002-09-30	0.0162	0.0219	0.0212	0.0203	0.0181	0.0131	0.0049	-0.0019	-0.0269	-0.0456
2002-10-30	-0.0063	-0.0103	-0.0084	-0.0093	-0.0025	-0.0044	-0.0197	0.0003	0.0072	0.0054
2002-11-30	-0.0066	0.0034	0.0007	0.0080	0.0127	0.0272	0.0433	0.0605	0.0783	0.1582
2002-12-30	0.0152	0.0219	0.0212	0.0197	0.0241	0.0252	0.0226	0.0183	0.0077	0.0392
2003-01-31	0.0015	0.0016	0.0057	0.0077	0.0135	0.0141	0.0120	0.0149	0.0359	0.1070
2003-02-28	0.0120	0.0162	0.0145	0.0186	0.0184	0.0170	0.0176	0.0129	0.0151	0.0307
2003-03-28	-0.0033	-0.0016	0.0125	0.0013	0.0027	0.0104	0.0120	0.0097	0.0275	0.0643
2003-04-30	0.0013	0.0071	0.0128	0.0164	0.0231	0.0315	0.0317	0.0442	0.0581	0.1516
2003-05-30	0.0169	0.0259	0.0312	0.0326	0.0331	0.0350	0.0255	0.0173	0.0062	0.0506
2003-06-30	-0.0008	-0.0018	-0.0009	-0.0003	0.0039	0.0111	0.0101	0.0188	0.0223	0.0645
...	...	...	...	...	...	...	...	...	...	...
2011-03-28	-0.0001	-0.0005	0.0002	0.0002	0.0005	-0.0009	0.0026	0.0079	0.0055	0.0047
2011-04-30	0.0065	0.0107	0.0130	0.0149	0.0176	0.0224	0.0219	0.0182	0.0141	0.0179
2011-05-30	0.0041	0.0099	0.0139	0.0144	0.0160	0.0150	0.0138	0.0125	0.0066	0.0033
2011-06-30	-0.0005	-0.0007	-0.0030	-0.0044	-0.0064	-0.0062	-0.0076	-0.0074	-0.0067	-0.0121
2011-07-30	0.0059	0.0150	0.0209	0.0230	0.0221	0.0210	0.0160	0.0134	0.0140	0.0065
2011-08-31	0.0014	0.0140	0.0200	0.0182	0.0114	0.0018	-0.0064	-0.0180	-0.0265	-0.0506
2011-09-30	-0.0013	0.0077	0.0124	0.0079	0.0010	-0.0085	-0.0187	-0.0262	-0.0232	-0.0483
2011-10-30	0.0029	0.0043	0.0088	0.0127	0.0212	0.0268	0.0354	0.0432	0.0490	0.0771
2011-11-30	-0.0026	-0.0047	-0.0086	-0.0119	-0.0152	-0.0193	-0.0251	-0.0334	-0.0180	-0.0347
2011-12-30	0.0055	0.0100	0.0195	0.0219	0.0256	0.0275	0.0222	0.0277	0.0322	0.0244

This table displays a head/tail snapshot of the replication portfolio returns, which are monthly from September 2002 onward. The yield spreads and returns are based on combining WRDS corporate bond data and duration-matched Treasury yields from Nozawa's dataset.

4 Summary Statistics

In addition to the decile analysis, we computed summary statistics for both the benchmark and the replicated returns, in addition to the overall data. The tables below present the mean, standard deviation, cumulative return, and the period over which the data are available (start and end dates) for each decile. Although the overall directional trends are similar between the benchmark and replication summaries, small differences in magnitude are evident. These discrepancies likely arise from missing yield data and other data cleaning challenges.

Table 3: Benchmark Summary Statistics

Portfolio	Mean	Std	Cumulative Return	Start Date	End Date
11	0.0056	0.0101	0.4366	2002-09-30	2011-11-30
12	0.0069	0.0119	0.5549	2002-09-30	2011-11-30
13	0.0063	0.0129	0.4924	2002-09-30	2011-11-30
14	0.0066	0.0128	0.5263	2002-09-30	2011-11-30
15	0.0058	0.0133	0.4473	2002-09-30	2011-11-30
16	0.0056	0.0122	0.4305	2002-09-30	2011-11-30
17	0.0059	0.0109	0.4624	2002-09-30	2011-11-30
18	0.0073	0.0111	0.5953	2002-09-30	2011-11-30
19	0.0080	0.0154	0.6651	2002-09-30	2011-11-30
20	0.0136	0.0293	1.3411	2002-09-30	2011-11-30
Overall	0.0072	0.0112	0.5831	2002-09-30	2011-11-30

This table provides summary statistics for the benchmark US corporate bond returns. The statistics indicate overall trends that align with the replicated data, albeit with slight differences in average returns and volatility.

Table 4: Replication Summary Statistics

Portfolio	Mean	Std	Cumulative Return	Start Date	End Date
11	0.0019	0.0089	0.1294	2002-09-30	2011-11-30
12	0.0045	0.0123	0.3359	2002-09-30	2011-11-30
13	0.0055	0.0136	0.4231	2002-09-30	2011-11-30
14	0.0059	0.0144	0.4559	2002-09-30	2011-11-30
15	0.0075	0.0154	0.6134	2002-09-30	2011-11-30
16	0.0089	0.0183	0.7618	2002-09-30	2011-11-30
17	0.0089	0.0215	0.7571	2002-09-30	2011-11-30
18	0.0102	0.0252	0.8997	2002-09-30	2011-11-30
19	0.0104	0.0330	0.8883	2002-09-30	2011-11-30
20	0.0201	0.0668	2.1824	2002-09-30	2011-11-30
Overall	0.0084	0.0189	0.7017	2002-09-30	2011-11-30

This table shows summary statistics for the replicated decile returns. Although these statistics capture the general direction of the benchmark data, differences in the mean, standard deviation, and cumulative return suggest that the replication is not a perfect match, likely due to data limitations.

5 Replication Analysis

Below is a table of the key replication metrics, treating each decile's replicated returns relative to the benchmark. Although we achieved strong alignment in trends and directions, some differences emerged due to missing yield data and partial coverage of duration-matched Treasuries.

Table 5: Decile Replication Analysis

Portfolio	Correlation	R Squared	Slope	Intercept	Mae	Rmse	Tracking Error
11	0.8263	0.6828	0.9393	0.0038	0.0040	0.0057	0.0057
12	0.9118	0.8313	0.8796	0.0029	0.0036	0.0048	0.0051
13	0.9636	0.9286	0.9194	0.0012	0.0024	0.0034	0.0036
14	0.9445	0.8921	0.8351	0.0017	0.0030	0.0042	0.0048
15	0.9057	0.8203	0.7842	-0.0001	0.0044	0.0056	0.0065
16	0.8047	0.6476	0.5368	0.0008	0.0054	0.0072	0.0111
17	0.8496	0.7218	0.4306	0.0021	0.0038	0.0057	0.0134
18	0.9344	0.8732	0.4101	0.0031	0.0031	0.0039	0.0153
19	0.9433	0.8898	0.4399	0.0034	0.0034	0.0051	0.0190
20	0.9443	0.8917	0.4146	0.0053	0.0069	0.0096	0.0400

Correlations, R^2, regression slope, intercept, MAE, RMSE, and tracking error for decile portfolios. High correlations and low errors suggest a solid replication, though exact agreement was not always possible.

6 Results and Discussion

6.1 Cumulative Returns of Decile Portfolios

Figure 2 shows the cumulative returns for each of the ten deciles. Although our replication focuses on the 2002–2012 overlap, we extended the plot for illustrative purposes using all available data points from Nozawa's dataset. Notably, portfolios with the **lowest yield spreads** (e.g., Decile 11) tend to exhibit more stable, steady growth, reflecting a relatively lower risk profile. Conversely, portfolios with the **highest yield spreads** (e.g., Decile 20) show greater volatility—achieving higher peaks but also larger drawdowns, aligning with the higher-risk nature of high-spread bonds.

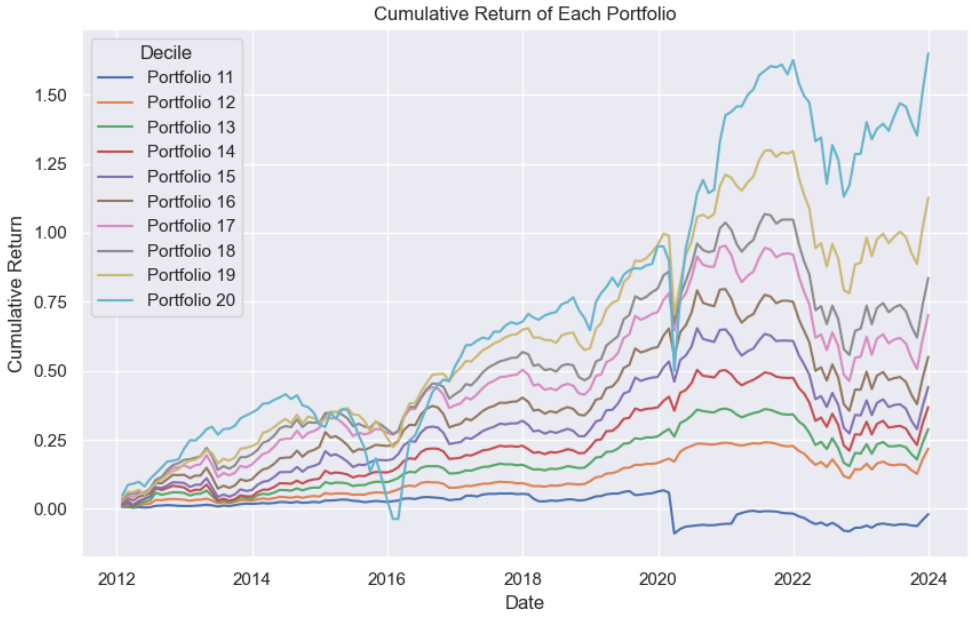

Figure 2: Cumulative Returns of Replicated Decile Portfolios

This figure illustrates the cumulative returns for each yield-spread decile over time. Portfolios in lower deciles (lower spreads) show steadier returns and less volatility, while higher-spread deciles exhibit higher peaks and more pronounced drawdowns. The ordering confirms the risk-return relationship typically associated with yield spreads.

6.2 Overall Assessment

Despite missing yield data and necessary data cleaning (e.g., handling negative amounts outstanding), we largely succeeded in replicating the decile returns for the overlapping 2002–2012 period. The correlation and regression statistics (Table 5) indicate that our replicated returns closely track the benchmark. However, the summary statistics (Tables 3 and 4) reveal small discrepancies in the mean, standard deviation, and cumulative returns between the benchmark and replicated data. These differences suggest that while the replication captures the overall directional trend and risk-return profile, exact numerical matching was not achieved due to data limitations.

7 Conclusion

Our replication of Nozawa's corporate bond portfolios from He, Kelly, and Manela (2017) shows strong alignment in direction and decile ranking during the overlapping 2002–2012 window. The combination of WRDS corporate bond data and Nozawa's duration-matched Treasury yields allowed us to construct yield spreads and subsequent returns starting in September 2002. Key challenges included missing yield data, negative amounts outstanding, and limited data coverage relative to the broader 1970–2012 HKM sample. Overall, however, the high correlations, low error measures, and similar decile trajectories indicate a largely successful replication effort, even though slight differences in summary statistics point to some inherent data challenges.

8 References

He, Z., B. Kelly, and A. Manela (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics 126*(1), 1–35.

Nozawa, Y. (2017). What drives the cross-section of credit spreads?: A variance decomposition approach. *The Journal of Finance 72*(5), 2045–2072.

Open Source Bond Asset Pricing (2025). Open source bond asset pricing. https://openbondassetpricing.com/. Accessed: March 12, 2025.

Wharton Research Data Services (WRDS) (2025). Crsp bond returns database. Wharton Research Data Services (WRDS). Accessed: March 12, 2025.