

Monthly TRACE Data README*

Alexander Dickerson^a, Cesare Robotti,^{b,†} and Giulio Rossetti^b

^a*UNSW Business School, Sydney, NSW 2052, Australia*

^b*Warwick Business School, Coventry, CV4 7AL, United Kingdom*

July 11, 2024

ABSTRACT

README for the monthly TRACE data available on openbondassetpricing.com.

*The companion website to this README, openbondassetpricing.com, contains updated corporate bond factor data, replication code, and market-microstructure adjusted TRACE bond prices and returns.

A WRDS MMN Corrected Data

Please download the data at <https://openbondassetpricing.com/>. Generally, lower case bond-related characteristics (signals) are *not* adjusted for MMN. The MMN-adjustment can be understood by reading Dickerson, Robotti, and Rossetti (2023) and Bartram, Grinblatt, and Nozawa (2020). Upper case variables *are* adjusted for MMN. The variables `exretn_t+1`, `exretnc_dur_t+1`, `exretnc_t+1` and `bond_ret_t+1` are sampled from $t : t + 1$, i.e., one-month ahead. These variables are not signals/characteristics and are not changed/augmented for MMN. Only *signals* in lower case, without the $t + 1$ suffix are adjusted. The remaining variables are observable to the investor at the end of month t .

NOTE: *always* use the return variables with the suffix $t + 1$ as the month-ahead ex ante bond returns. Do not use the MMN-adjusted returns as future bond returns, these are to be used as *signals* only. When forming your portfolios, only use the MMN-adjusted bond return as the short-term reversal signal, never as the return that is realized ex ante.

July 2024 Addition: inclusion of `permno` and `issuer_cusip` identifiers and two dummies: `zcb` (1 if Zero Coupon, 0 otherwise) and `conv` (1 if Convertible, 0 otherwise).

June 2024 Addition: duration adjusted returns from Binsbergen et al. (2023).

April 2024 Addition: included two dummies, `size_ig` and `size_jk`, encoded as 1 if the bond is in the BAML/ICE indices or not. If you only want the BAML/ICE constituent bonds, only keep the bonds where these column variables = 1.

The variables are as follows (not always in order as they appear in the dataframe):

1. `date`: monthly date.
2. `cusip`: bond CUSIP.
3. `issuer_cusip`: issuer CUSIP.
4. `permno`: firm-level identifier from CRSP.
5. `exretn_t+1`: one-month ahead bond return in excess of the one-month risk-free rate of return.

6. `exretnc_t+1`: one-month ahead bond return in excess of a maturity-matched portfolio of U.S. Treasury Bond returns.
7. `exretnc_dur_t+1`: one-month ahead bond return in excess of a duration-matched portfolio of U.S. Treasury Bond returns.
8. `bond_ret_t+1`: one-month ahead bond return.
9. `bond_ret`: current-month bond return.
10. `exretnc`: current-month bond return in excess of the one-month risk-free rate of return.
11. `exretnc_dur`: current-month bond return in excess of a duration-matched portfolio of U.S. Treasury Bond returns.
12. `rating`: bond-level S&P rating, 1 (AAA) to 22 (D).
13. `cs`: MMN-biased bond credit spread computed as `bond_yield` in excess of a duration-matched portfolio of U.S. Treasury Bond yields.
14. `bond_yield`: bond equivalent yield.
15. `bond_amount_out`: bond amount outstanding.
16. `offering_amt`: bond offering amount.
17. `bondprc`: MMN-biased bond price.
18. `perc_par`: bond price (MMN-biased) as a percentage of par.
19. `tmt`: bond time-to-maturity.
20. `duration`: bond modified duration (computed with the MMN-biased price).
21. `ind_num_17`: Fama French 17 industry classification.
22. `sic_code`: Industry SIC code.
23. `mom6_1`: Cumulative sum of bond returns over 6-months skipping the prior month as in Jostova, Nikolova, Philipov, and Stahel (2013)

24. `ltrev48_12`: Cumulative sum of bond returns over 48-months skipping the prior 12-months as in Bali, Subrahmanyam, and Wen (2021).
25. `BOND_RET`: MMN-adjusted bond return.
26. `ILLIQ`: MMN-adjusted bond illiquidity as per Bao, Pan, and Wang (2011).
27. `var95`: 24-month minimum expanding up to 36-month rolling historical VaR-95%.
28. `n_trades_month`: Number of bond trades in the month.
29. `BOND_YIELD`: MMN-adjusted bond equivalent yield.
30. `CS`: MMN-adjusted bond credit spread.
31. `BONDPRC`: MMN-adjusted bond clean price.
32. `PRFULL`: MMN-adjusted bond dirty (full) price.
33. `DURATION`: MMN-adjusted bond modified duration.
34. `CONVEXITY`: MMN-adjusted bond convexity.
35. `bond_value`: Bond clean unadjusted price multiplied by bond amount outstanding.
36. `BOND_VALUE`: MMN-adjusted bond clean unadjusted price multiplied by bond amount outstanding.
37. `size_ig`: dummy for the investment grade bonds in the BAML/ICE indices.
38. `size_jk`: dummy for the junk grade bonds in the BAML/ICE indices.

B Data Filters and Descriptions

B.1 TRACE (WRDS) Bond Database

The Wharton Research Data Services (WRDS) Bond Database is a pre-processed monthly bond database that uses the Enhanced TRACE and Mergent FISD bond databases. It was introduced by WRDS in April 2017. The data is publicly available. After logging in to WRDS, the data is available here. We use the version of the WRDS dataset that has a sample end date of 2022:09.

WRDS bond returns. The WRDS data team provides us with three different bond return variables: `RET_EOM` (returns are computed using bond prices that land on any day of the month), `RET_L5M` (a bond must trade on the last five days of the month), and `RET_LDM` (a bond must trade on the last day of the month). For the results based on the WRDS Bond Database, we always use `RET_L5M`, i.e., a return is valid if the bond trades on the last five days of month t and month $t - 1$. However, the publicly available data we use from WRDS, imposes a data filter which sets any bond return that is greater than 100% to 100%, i.e., the data is truncated/trimmed at this level. Although this does not make any material difference whatsoever to asset pricing results, we carefully address the issue below.

WRDS bond returns truncation correction. Download the WRDS Bond returns truncation correction file. We carefully adjust for the truncation of bonds with returns greater than +100% imposed by WRDS, by setting any bond return which is truncated to the return observed in the ICE database, i.e., if the WRDS bond return is equal to 100% (truncated), we set this value to the bond return from ICE as the ‘true’ bond return. If the ICE return is missing, we set the value to the return computed from the TRACE data itself. These adjustments do not make any material difference to the robustness results. In total we identify only 94 cases where the truncation occurs, and we are able to address 91 of them. The remaining 3 cases are removed.

WRDS bond filters. To align the data to the Bank of America Merrill Lynch (BAML) corporate bond database provided by the Intercontinental Exchange (ICE), we follow Andreani, Palhares, and Richardson (2023) and use the following filters (all using data provided by WRDS),

1. Remove investment (IG) rated bonds that have less than USD 150 million outstanding prior to, and including, November 2004, and less than USD 250 million after November 2004.
2. Remove non-investment grade (HY) rated bonds that have less than USD 100 million outstanding prior to, and including, September 2016, and less than USD 250 million after September 2016.

We merge the WRDS data to the Mergent FISD database (also publicly available via the WRDS data platform) and apply the following filters which are all standard in the literature,

1. Only keep bonds that are issued by firms domiciled in the United States of America, `COUNTRY_DOMICILE == 'USA'`.
2. Remove bonds that are private placements, `PRIVATE_PLACEMENT == 'N'`.
3. Only keep bonds that are traded in U.S. Dollars, `FOREIGN_CURRENCY == 'N'`.
4. Bonds that trade under the 144A Rule are discarded, `RULE_144A == 'N'`.
5. Remove all asset-backed bonds, `ASSET_BACKED == 'N'`.
6. Remove convertible bonds, `CONVERTIBLE == 'N'`.
7. Only keep bonds with a fixed or zero coupon payment structure, i.e., remove bonds with a floating (variable) coupon, `COUPON_TYPE != 'V'`.
8. Remove bonds that are equity linked, agency-backed, U.S. Government, and mortgage-backed, based on their `BOND_TYPE`.
9. Remove bonds that have a “non-standard” interest payment structure or bonds not caught by the variable coupon filter (`COUPON_TYPE`). This affects a tiny fraction of bonds ($\sim 0.10\%$ or 142 bonds) of the FISD data file. We remove bonds that have an `INTEREST_FREQUENCY` equal to -1 (N/A), 13 (Variable Coupon), 14 (Bi-Monthly), and 15 and 16 (undocumented by FISD). Additional information on `INTEREST_FREQUENCY` is available on Page 60 of 67 of the FISD Data Dictionary 2012 document.
10. Remove a small fraction of bonds that do not have the required (and crucial information) to compute accrued interest. Bonds that do not have a valid `DATED_DATE` are removed (3,051 bonds). The `DATED_DATE` variable is the date from which bond interest accrues. Bonds without a valid `INTEREST_FREQUENCY`, `DAY_COUNT_BASIS`, `OFFERING_DATE`, `COUPON_TYPE`, and `COUPON` are also removed (425 bonds in total).

For bonds with missing amount outstanding information in the file, we set the amount outstanding equal to the face value at issuance.

Enhanced TRACE (TRACE) data

TRACE provides intraday bond clean prices, trading volumes, and buy-and-sell indicators. We apply the standard bond filtering procedure used by Binsbergen, Nozawa, and Schwert (2023).

TRACE bond filters. We apply the following filters in cleaning the intraday TRACE data, for the pre-2012 database (see Dick-Nielsen 2014 for detailed descriptions of the TRACE data changes).

1. Keep all trades that have less than two days to settlement, `days_to_sttl_ct == '002'`, `days_to_sttl_ct == '001'`, `days_to_sttl_ct == '000'` or `days_to_sttl_ct == 'None'`.
2. Remove trade records with the 'when-issued' indicator, `wis_fl != 'Y'`.
3. Remove trade records with the 'locked-in' indicator, `lckd_in_ind != 'Y'`.
4. Keep trade records which do not have special conditions, `sale_cndtn_cd == 'None'` or `sale_cndtn_cd == '@'`.

Thereafter, we clean the bond trades for reversals, corrections, and cancellations in the standard manner as prescribed by Dick-Nielsen (2009) and Dick-Nielsen (2014). The end-of-day bond clean price is the volume-weighted price of all eligible trades within each day d of month t , that register par volume equal or greater than \$10,000, `entrd_vol_qt \geq $10,000`. We are careful to *not* filter out bond trades (`rptd_pr`) that have a price of less than \$5 or greater than \$1,000. Imposing this filter introduces look-ahead to the out-of-sample portfolio results that use monthly data.¹

Bank of America Merrill Lynch (BAML) database

The BAML data is provided by the Intercontinental Exchange (ICE) and provides daily bond price quotes, accrued interest, and a host of pre-computed corporate bond characteristics such as the bond option-adjusted credit spread (OAS), the asset swap spread, duration, convexity, and bond returns in excess of a portfolio of duration-matched Treasuries. The ICE sample spans the time period 1997:01 to 2022:12 and includes constituent bonds from the ICE Bank of America High Yield (H0A0) and Investment Grade (C0A0) Corporate Bond Indices. To align the ICE sample with the WRDS sample, we use a sample end date of 2022:09.

¹We thank Avaniidhar Subrahmanyam for making us aware of this fact.

ICE bond filters. We follow Binsbergen et al. (2023) and take the last quote of each month to form the bond-month panel. We then merge the ICE data to the Mergent FISD database (which has had the same filters applied to it as discussed above). The following ICE-specific filters are then applied:

1. Only include corporate bonds, `Ind_Lvl_1 == 'corporate'`
2. Only include bonds issued by U.S. firms, `Country == 'US'`
3. Only include corporate bonds denominated in U.S. Dollars, `Currency == 'USD'`

BAML/ICE bond returns. Total bond returns are computed in a standard manner in ICE, and no assumptions about the timing of the last trading day of the month are made because the data is quote based, i.e., there is always a valid quote at month-end to compute a bond return. This means that each bond return is computed using a price quote at exactly the end of the month, each and every month. This introduces homogeneity into the bond returns because prices are sampled at exactly the same time each month. ICE only provides bid-side pricing, meaning bid-ask bias is inherently not present in the monthly sampled prices, returns and credit spreads. The monthly ICE return variable is (as denoted in the original database), is `trr_mtd_loc`, which is the month-to-date return on the last business day of month t .

References

- Andreani, Martina, Diogo Palhares, and Scott Richardson, 2023, Computing corporate bond returns: A word (or two) of caution, *Review of Accounting Studies* forthcoming.
- Bali, Turan G, Avanidhar Subrahmanyam, and Quan Wen, 2021, Long-term reversals in the corporate bond market, *Journal of Financial Economics* 139, 656–677.
- Bao, Jack, Jun Pan, and Jiang Wang, 2011, The illiquidity of corporate bonds, *Journal of Finance* 66, 911–946.
- Bartram, Söhnke M., Mark Grinblatt, and Yoshio Nozawa, 2020, Book-to-market, mispricing, and the cross-section of corporate bond returns, Technical report, National Bureau of Economic Research.
- Binsbergen, Jules H. van, Yoshio Nozawa, and Michael Schwert, 2023, Duration-based valuation of corporate bonds, Working Paper.
- Dick-Nielsen, Jens, 2009, Liquidity biases in TRACE, *Journal of Fixed Income* 19, 43–55.
- Dick-Nielsen, Jens, 2014, How to clean Enhanced TRACE data, Working Paper.
- Dickerson, Alexander, Cesare Robotti, and Giulio Rossetti, 2023, Return-based anomalies in corporate bonds: Are they there?, Working Paper, Warwick Business School.
- Jostova, Gergana, Stanislava Nikolova, Alexander Philipov, and Christof W Stahel, 2013, Momentum in corporate bond returns, *Review of Financial Studies* 26, 1649–1693.