

Mutual Funds Links (MFLINKS)

Wharton Research Data Services (WRDS)

January 2013

and Wharton Financial Institution Center (WFIC)

MFLINKS provide a reliable means to join the mutual fund performance and expense information in the CRSP Mutual Fund Database (MFDB) with the equity holdings information in the Thomson-Reuters Mutual Fund ownership database (TFN/CDA S12).

MFLINKS allows a researcher to gather details on holdings for either particular funds or fund families and groups of funds at specific points in time. MFLINKS has been developed with Professor Russ Wermers from the University of Maryland's Robert H. Smith School of Business, as a contributor and ongoing consultant to the project. Professor Wermers was one of the first researchers to use linked mutual fund data. A good example of how MFLINKS can be used is his article, "Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses" (Journal of Finance, 2000). It decomposes mutual fund performance into its various components based upon actual holdings.

Coverage

MFLINKS provides links between the 98% of the domestic equity funds in CRSP and Thomson-Reuters. The database consists of two tables that provide the mapping between CRSP and Thomson-Reuters mutual fund identifiers, to a common Wharton Financial Institution Center Number (WFICN), which is a unique and permanent fund portfolio identifier. There are 23,340 WFICN-CRSP_FUNDNO pairs in MFLINKS Table 1, including 4,600+ new links added in the 2013 update. There are 8,476 FUNDNO-WFICN pairs in MFLINKS Table 2, including 1,850 new links added in the 2013 update.

Using MFLINKS

Joining the CRSP MFDB and TFN/CDA data is a two-step process. The first step involves translating the CRSP_FUNDNOs in the CRSP MFDB set to WFICN identifiers. The second step uses the WFICNs to link to FUNDNO in TFN/CDA. It is also possible to reverse the linking, i.e. from FUNDNO-FDATE to WFICN and then CRSP_FUNDNO.

Note that FUNDNO in TFN/CDA is not the same as CRSP_FUNDNO in CRSP and they do not represent the same entity. In addition, CRSP_FUNDNO is at the share class level, while TFN/CDA FUNDNO is at the portfolio level.

MFLINKS Table1 (*mfl.mflink1*)

CRSP_FUNDNO	Fund Identifier used in CRSP MFDB
WFICN	Wharton Financial Institution Center Number

Each linked CRSP_FUNDNO has a single WFICN, but a WFICN may have more than one CRSP_FUNDNO. The latter covers cases where funds are offered as different share classes but have the exact same portfolio of holdings (and are reported as a single entity for the purpose of filing the fund's holdings at the SEC).

Not all CRSP_FUNDNOs in the CRSP MFDB set are given a WFICN link. MFLINKS concentrates on a target universe comprised of domestic equity funds, exchange traded funds, and target-date equity funds. These are defined by objective codes and stock holding percentages in MFDB (to identify domestic equity funds). Approximately 92% of the funds, and 96% of the assets, in the target universe have been linked. Some foreign, bond, and money market funds have also been linked. The unlinked domestic equity funds are mainly small, defunct funds where adequate information for linking is not available. New funds are also less likely to be linked, mainly because they are not yet in the TFN/CDA data.

MFLINKS Table 2 (*mfl.mflink2*)

WFICN	Wharton Financial Institution Center Number
FUNDNO	TFN/CDA Portfolio ID
FUNDNO_ID	Fund Number (FUNDNO) Re-Use Counter
FUNDNAME	Fund Portfolio Name
FDATE	File/Vintage Date
RDATE	Holdings Report Date
MGRCOAB	Management Company Abbreviation
ASSETS	End of Quarter Assets (x\$10,000)
NUM_HOLDINGS	Number of Portfolio Holdings
IOC	Investment Objective Code
PRDATE	Prior Report Date
COUNTRY	Fund Domicile Country

The information in MFLinks Table 2 comes from the *tfn.s12type1* file which has two date variables. FDATE is the file date on which the fund appears in the Thomson database. RDATE is the report date on which the stock holdings for that fund are valid. In this file, there are often multiple FDATES for each RDATE, which is due to staleness of data.

Additional Benefits of MFLINKS

The MFLINKS Table 2 solves a significant problem with the TFN/CDA set in cases where FUNDNOs have been re-used or changed arbitrarily. In essence, WFICN serves as a permanent identifier for finding and tracking holdings in the TFN/CDA set. FUNDNO cannot be used with complete confidence because of examples such as below:

Example 1: Re-used FUNDNO

WFICN	FundNo	Fund ReUse	FDATE	RDATE	Thomson-Reuters Fund Name	Assets (x\$10,000)	# Holdings	Investment Objective
100008	230	1	31-Dec-90	31-Oct-90	ABT SECURITY INCOME	457	6	4
100008	230	1	31-Mar-91	31-Mar-91	ABT SECURITY INCOME	420	6	4
105364	230	2	30-Jun-95	31-Mar-95	CALVERT WRLD VAL-CAP ACC	398	110	3
105364	230	2	30-Sep-95	31-Mar-95	CALVERT WRLD VAL-CAP ACC	398	110	3
105364	230	2	31-Dec-95	30-Sep-95	CALVERT WRLD VAL-CAP ACC	1609	154	3

Example 1 illustrates the reuse of FUNDNO 230 for two different funds. Notice the gap in

the reporting dates between the old and new funds, which not only have different names, but also different assets, holdings, and investment objective code. To avoid linking errors, it is necessary to use FUNDNO-FDATE when pairing it with a particular WFICN, as FUNDNO 230 could be correctly used for two different funds.

Example 2: Arbitrary change in FUNDNO

In this case, FUNDNO 23000 became FUNDNO 32110 in 1993 in the TFN/CDA set when the name was changed, even though it is the same fund, and is treated as such in the MFDB set.

WFICN	FundNo	Fund ReUse	FDATE	RDATE	Thomson-Reuters Fund Name	Assets (x\$10,000)	# Holdings	Investment Objective
101151	23000	2	31-Mar-93	31-Mar-93	FSP-UTILITIES PORTFOLIO	12365	25	4
101151	32110	1	30-Jun-93	30-Jun-93	ISP-UTILITIES	16252	58	4
101151	32110	1	30-Sep-93	30-Sep-93	INVESCO STRAT-UTILITIES	21055	72	4
101151	32110	1	31-Dec-93	31-Dec-93	INVESCO STRAT-UTILITIES	17445	66	4

Out of a total of 8,476 FUNDNO-WFICN pairs in MFLINKS Table 2, there are 81 cases where a FUNDNO was re-used and 538 cases where a FUNDNO was changed arbitrarily.

Example 3: Multiple CRSP_FUNDNOs for the Same Fund

Two or more CRSP_FUNDNOs may have the same manager and portfolio and should be grouped under a single WFICN. In most of these cases, each CRSP_FUNDNO represents a different fund class with a unique expense and load structure. MFLINKS Table 1 can be used separately or together with Table 2 to study such funds. For example, MFLINKS Table 1 along with the MFDB set can be used to study how different fee structures affect realized investor returns, as well as mutual funds flows.

WFICN	CRSP FundNo	CRSP Fund Name	First Offer Date	CRSP Obj Code	TNA in 2006	Expense Ratio	12b-1 Fee	Front-End Load
101048	11345	Fidelity Advisor Equity Growth Fund; Class A Shares	19960903	EDYG	1160.5	0.0113	0.0025	0.058
101048	11290	Fidelity Advisor Equity Growth Fund; Class B Shares	19960101	EDYG	433.7	0.0194	0.0100	.
101048	11335	Fidelity Advisor Equity Growth Fund; Class C Shares	19971103	EDYG	290.7	0.0192	0.0100	.
101048	11289	Fidelity Advisor Equity Growth Fund; Class T Shares	19920910	EDYG	2924.1	0.0131	0.0050	0.035
101048	11288	Fidelity Advisor Equity Growth Fund; Institutional Class	19831122	EDYG	1186.4	0.0078	.	.

Here, the same basic fund is offered under five different share classes. It seems that there are no loads on funds with both the highest and lowest management expenses.

For these funds, the corresponding MFLINKS Table 2 entry is:

WFIGN	FundNo	Fund ReUse	FDATE	RDATE	Thomson-Reuters Fund Name	Assets ¹ (x\$10,000)
101048	21777	1	31-Dec-91	30-Nov-91	FIDELITY ADV EQ PORT-GR	6877
101048	21777	1	31-Mar-92	30-Nov-91	FIDELITY ADV EQ PORT-GR	6877
...						
101048	21777	1	30-Jun-06	31-May-06	FIDELITY ADVR EQUITY GRO	709522
101048	21777	1	30-Sep-06	31-Aug-06	FIDELITY ADVR EQUITY GRO	629000
101048	21777	1	31-Dec-06	30-Nov-06	FIDELITY ADVR EQUITY GRO	638514
...						
101048	21777	1	30-Jun-10	30-Jun-10	FIDELITY ADVR EQUITY GRO	271441
101048	21777	1	30-Sep-10	30-Sep-10	FIDELITY ADVR EQUITY GRO	295398
...						

The names used in the two mutual fund sets are close but not exactly the same. To confirm that the five CRSP_FUNDNOs in Example 3 share a single portfolio of holdings in the TFN/CDA set, we see that the sum of CRSP_FUNDNOs' assets in December 2006 (2006, quarter 4) is \$6 billion, close to the \$6.38 billion reported in November 2006 (RDATE) in TFN/CDA. In other words, at the end of 2006, the sum of TNAs from CRSP MFDB almost matches the TFN/CDA value at \$6.3 billion.

Using SAS to link MFLINKS Table 2 with the Thomson-Reuters (CDA/S12) Files

The logic for linking the MFLINKS Table 2 with Thomson data files is illustrated in the two examples below. In the first example, MFLINKS Table 2 is linked to *tfn.s12type3* and in the second example it is linked to *tfn.s12* which in addition to *tfn.s12type3* dataset has information on the stock holdings of the funds.

SAS example 1: Linking to *tfn.s12type3*

```
proc sql;
create table mydata as
select a.wfign, a.rdate, b.*
from mfl.mflink2 as a, tfn.s12type3 as b
where a.fundno=b.fundno and a.fdate=b.fdate;
quit;
```

SAS example 2: Linking to *tfn.s12*

```
proc sql;
create table mydata as
select a.wfign, b.*
from mfl.mflink2 as a, tfn.s12 as b
where a.fundno=b.fundno and a.fdate=b.fdate;
quit;
```

¹ Note: Assets in TFN/CDA are in \$10,000 units, while TNA in CRSP are in \$1 million units.