Can Transient Institutions Correctly Interpret Small Negative Earnings Surprises in the Absence of Access to Management's Private Information?

Gang Hu, Bin Ke, and Yong Yu

***Journal of Accounting, Auditing and Finance* 33, 2018, 3-33.**

INTERNET APPENDIX

THE ALGORITHM OF MATCHING ABEL/NOSER'S INSTITUTION IDS WITH SPECTRUM'S INSTITUTION IDS

In this appendix we describe how we match Abel/Noser's institution IDs (denoted as clientcode) with Spectrum's institution IDs (denoted as mgrnum). The matching is performed over the Abel/Noser database's coverage period 1999-2005. There are 3,864,132 institution-firm-year-quarters in the Abel/Noser database and 13,089,377 institution-firm-year-quarters in Spectrum.

We first clean the observations in both databases to reduce potential measurement errors in the matching. For the Abel/Noser database, we delete the *clientcode-quarters* that report no stock trades in any month of a calendar quarter. This restriction deletes institutions that enter or exit the Abel/Noser database in the middle of a quarter or fail to file their stock trades for some months of a calendar quarter. To avoid unnecessary complications, we require all *firm-quarters* to satisfy the following restrictions for both databases:

(a) The firms must be publicly traded U.S. domestic firms that issue only one class of common stock traded on one of the three major stock exchanges.

(b) There are no stock splits, stock delistings, or IPO in the calendar quarter.

(c) The stock price at the beginning of the calendar quarter is greater than $5 and the number of common shares outstanding at the beginning of the calendar quarter is greater than one million shares.

(d) The net quarterly institutional ownership change should be more than 500 shares. Our match results are the same if we require a minimum of 100 shares.

(e) Each firm-quarter must have institutional trading data from both databases (though not necessarily by the same institutions).

Those sample restrictions reduce the Abel/Noser database to 2,677,564 institution-firm-quarters and reduce the Spectrum database to 6,862,411 institution-firm-quarters. The loss of observations is largely due to restrictions (a) and (d).

We match Abel/Noser's clientcodes with Spectrum's mgrnums as follows. For both databases, we compute each institution's quarterly change in stock ownership (in number of shares) for each stock (permno), denoted as ΔIO. For each pair of clientcode and mgrnum, we define N_ABEL as the total number of firm-quarters that have nonmissing ΔIO in the Abel/Noser database and N_BOTH as the total number of firm-quarters that have nonmissing ΔIO in both databases. By definition, N_BOTH is always less than or equal to N_ABEL. For each pair of clientcode and mgrnum, we define N_SAME_ΔIO as the number of firm-quarters in which the values of ΔIO from the two databases are identical. MATCH1 is N_SAME_ΔIO divided by N_ABEL and represents the percentage of a clientcode's firm quarters whose quarterly institutional ownership changes are equal to an mgrnum's quarterly institutional ownership changes in the same firm-quarters.

For each clientcode, we retain the mgrnum with the highest MATCH1 (denoted as mgrnum_1st and MATCH1_1st, respectively) and the mgrnum with the second highest MATCH1

(denoted as mgrnum_2nd and MATCH1_2nd, respectively). For example, assume that clientcode=001 has valid data in two firm-quarters, IBM-1999-Q1 and DELL-2000-Q2. For IBM-1999-Q1, mgrnum=100 and mgrnum=200 have the same ΔIO as clientcode=001. For DELL-2000-Q2, only mgrnum=100 has the same ΔIO. Then, for clientcode=001, mgrnum=100's MATCH1=100% (the highest), and mgrnum=200's MATCH1=50% (the second highest).

Even if a pair of clientcode and mgrnum belongs to the same institution, MATCH1 is unlikely to equal 100% for several reasons. First, the Spectrum database may not contain all the stock trades for all the institutions. Small institutions with less than $100 million in equity securities are not required to file the 13f with the SEC. Even if an institution is required to file the 13f, it is only required to disclose common stock positions greater than 10,000 shares or $200,000. Second, some institutions may report their stock positions under one clientcode in Abel/Noser but under several different mgrnums in Spectrum. For example, we find that Fidelity Management & Research (FMR Co.) and Fidelity International are separate companies in Spectrum (they report under different mgrnums) even though they are both subsidiaries of Fidelity Investments and share the same clientcode in the Abel/Noser database. We have corrected this inconsistency for Fidelity but we don't know if a similar problem exists for other institutions. Third, the stock holdings reported in Spectrum may not reflect the holdings exactly at the end of a calendar quarter and thus ΔIO in Spectrum may not be the same as ΔIO in the Abel/Noser database. Finally, the Abel/Noser database may not contain all the trading data for a given institution. Although we have confirmed that Abel/Noser did not filter data on their end, it is possible that their clients filtered out certain parts of the data before sending the data to Abel/Noser.

Because of these and other unknown reasons, MATCH1 could be low even if a pair of clientcode and mgrnum belongs to the same institution. To mitigate this problem, we also define an alternative matching score (denoted as MATCH2) that is identical to MATCH1 except that ΔIO in Spectrum is deemed identical to ΔIO in Abel/Noser as long as the absolute difference in ΔIO across the two databases is less than 10% of the absolute ΔIO from Abel/Noser.

We follow two basic principles to match clientcodes with mgrnums. First, there must be a reasonable number of common firm-quarters (i.e., N_BOTH) for each pair of clientcode and mgrnum so that the matching scores MATCH1 and MATCH2 are reliable. Second, MATCH1_1st (MATCH2_1st) should be as high as possible while MATCH1_2nd (MATCH2_2nd) should be as low as possible. We use the following four sequential iterations to identify 103 matched pairs of clientcode-mgrnum. The stringency of the matching conditions declines from iteration 1 to iteration 4. However, it is important to note that our inferences are qualitatively similar if we restrict our sample to the matched institutions from iteration 1 only (untabulated).

Iteration 1 requires the following matching conditions:

- N_BOTH>=100 firm quarters.
- MATCH1_1st>=10%.
- MATCH2_1st>=20%.
- MATCH1_2nd<=1%.
- MATCH2_2nd<=5%.

We selected the cutoffs for iteration 1 based on the empirical distributions of the variables. Iteration 1 results in 62 matched pairs of clientcode-mgrnum.

Iteration 2 relaxes some of the matching conditions in iteration 1 as follows:

- N_BOTH>=100 firm quarters.
- MATCH1_1st>=5%.
- MATCH2_1st>=15%.
- MATCH1_2nd<=1%.
- MATCH2_2nd<=5%.

Iteration 2 results in 24 additional matched pairs of clientcode-mgrnum.

Iteration 3 repeats iterations 1 and 2 after deleting the 84 matched pairs of clientcode-mgrnum from iterations 1 and 2. However, iteration 3 results in no additional matched pairs of clientcode-mgrnum.

For the remaining unmatched clientcodes and mgrnums, iteration 4 requires that N_BOTH>=100 firm quarters and either MATCH1_1st>=5% or MATCH2_1st>=10%. Those conditions result in a total of 32 pairs of clientcode-mgrnum, from which we retained 17 additional matched pairs of clientcode-mgrnum whose MATCH1_1st dominates MATCH1_2nd or MATCH2_1st dominates MATCH2_2nd.

Table 1A shows the overlap of firm-quarters between Abel/Noser and Spectrum for the 103 matched institutions as a whole and by iteration. For the 103 matched institutions as a whole, on average 31.4% of the combined firm-quarters from both databases have institutional trading data in Abel/Noser only, 10.7% in Spectrum only, and 57.8% in both. The high percentage of firm quarters that have only institutional data from Abel/Noser is not surprising because as noted above, Abel/Noser should contain all the institutional trades while Spectrum tend to report only large institutions' stock positions greater than a certain threshold.

Table 2A reports descriptive statistics of MATCH1 and MATCH2 for the 103 matched institutions. By construction, MATCH2_1st is always no smaller than MATCH1_1st and

MATCH1_1st (MATCH2_1st) is always higher than MATCH1_2nd (MATCH2_2nd). For all 103 matched institutions, the mean MATCH2_1st is 31.46% versus a mean MATCH1_1st of 19.24%. As neither MATCH1_1st nor MATCH2_1st is close to 100%, one may question the accuracy of our matching algorithm.

We check the reasonableness of our matching algorithm in two ways. First, we use a large Abel/Noser institution whose clientcode is known to us to gauge the quality of our matching procedures. Our algorithm successfully matched this institution's clientcode in Abel/Noser with its mgrnum in Spectrum. However, this institution's MATCH1_1st is only 20% and MATCH2_1st is only 55% while MATCH1_2nd and MATCH2_2nd are both 0%. Those values are comparable to the mean values reported in Table 2A. This evidence suggests that a successful match does not require either MATCH1_1st or MATCH2_1st to be 100%.

Second, we know that mgrnum in Spectrum is unique and thus our matching algorithm should not be able to find pairs of mgrnums within Spectrum with high MATCH1_1st and MATCH2_1st. To check if this is the case, we use the same matching algorithm to find the best match for each mgrnum in Spectrum over the period 1999-2005. There are 3,272 unique institutions in Spectrum (mgrnum) that satisfy our pre-match sample screening described above. Out of the 3,272 mgrnums, 2,730 mgrnums have at least 1 firm-quarter that has the same ΔIO as another mgrnum. After imposing the requirement that N_BOTH>=100, we identified 1,638 pairs of matched mgrnums. Table 3A reports the result of this pseudo match. The largest MATCH1_1st is only 1.19% while the largest MATCH2_1st is only 4.74%. By comparison, the minimum MATCH1_1st is 3% while the minimum MATCH2_1st is 6% for the 103 institutions in our sample (untabulated). Hence, the results in Tables 2A and 3A suggest that our matching algorithm is unlikely to cause a mismatch between clientcodes with mgrnums.

Table 1A. The mean (median) number and percentage of firm-quarters that are in Abel/Noser only, in Spectrum only, and in both databases for the 103 matched pairs of clientcode-mgrnum

Iteration	Firm-quarters that have institutional trading data in Abel/Noser only		Firm-quarters that have institutional trading data in Spectum only		Firm-quarters that have institutional trading data in both databases	
	Number of firm-quarters	As a percentage of total firm-quarters from both databases	Number of firm-quarters	As a percentage of total firm-quarters from both databases	Number of firm-quarters	As a percentage of total firm-quarters from both databases
1	1,999 (566)	29.1 (28.0)	493 (97)	7.3 (3.1)	4,640 (1,418)	63.6 (65.1)
2	1,868 (711)	30.2 (27.2)	1,189 (148)	13.5 (8.3)	6,100 (1,425)	56.4 (55.2)
3	N/A	N/A	N/A	N/A	N/A	N/A
4	1,838 (1,615)	41.6 (39.6)	1,503 (403)	19.6 (8.5)	2,807 (1,046)	38.9 (38.2)
All	1,942 (699)	31.4 (28.5)	822 (144)	10.7 (4.9)	4,678 (1,323)	57.8 (57.8)

Table 2A. Matching statistics of the 103 matched pairs of clientcode-mgrnum

Mean, (median)

Iteration	MATCH1		MATCH2	
	MATCH1_1st	MATCH1_2nd	MATCH2_1st	MATCH2_2nd
1	27.71 (20.50)	0.06 (0.00)	39.92 (35.00)	0.51 (0.00)
2	7.04 (7.00)	0.04 (0.00)	22.83 (21.00)	0.58 (0.00)
3	N/A	N/A	N/A	N/A
4	5.59 (5.00)	0.06 (0.00)	12.76 (13.00)	0.82 (1.00)
All	19.24 (14.00)	0.06 (0.00)	31.46 (28.00)	0.58 (0.00)

For each pair of clientcode and mgrnum, MATCH1 is the percentage of a clientcode's firm-quarters whose quarterly institutional ownership (ΔIO) is exactly matched with an mgrnum's quarterly institutional ownership in the same firm-quarters. MATCH2 is defined similarly except that ΔIO in Spectrum is deemed identical to ΔIO in Abel/Noser as long as the absolute difference in ΔIO across the two databases is less than 10% of the absolute ΔIO from Abel/Noser.

Table 3A. Descriptive statistics from the pseudo matching of mgrnums within Spectrum

Variable	Mean	Min	10%	50%	90%	Max
MATCH1_1st	0.12	0.00	0.02	0.08	0.26	1.19
MATCH2_1st	0.76	0.00	0.00	0.69	1.38	4.74

See Table 2A for the definitions of MATCH1 and MATCH2. MATCH1_1st is the highest MATCH1 for a pair of mgrnums in Spectrum.