Pair Trading with an Ontology of SEC Financial Reports

Can Erten
Dept. of Computer Science
University of York
York, UK
https://orcid.org/0000-0003-3163-5220

Neel Chotai
Dept. of Computer Science
University of York
York, UK
https://orcid.org/0000-0001-9671-0936

Dimitar Kazakov
Dept. of Computer Science
University of York
York, UK
https://orcid.org/0000-0002-0637-8106

***Abstract*—Pair trading is a market-neutral strategy which is based on the use of standard, well-known statistical tests applied to time series of price to identify suitable pairs of stock. This article studies the potential benefits of using additional qualitative information for this type of trade. Here we use an ontology to represent and structure information extracted from financial SEC reports in a way that is optimised for search. These mandatory reports are originally published as XML using a schema that has varied over the years. The XML format itself is not easy to query, e.g. projections to fields or their composition are hard to find even when using an XML store. Our ontology-based approach provides uniformity of representation, which is further enhanced by the strife to use common vocabulary wherever possible. The ontology is then used to identify links between companies, by finding common senior employees or major shareholders. This is also potentially useful information to identify suitable pairs of stock. We show that the ontology increases the probability of selecting cointegrated pairs of stock from the data, with no negative effect on the survival time of such pairs when compared to random ones.**

***Index Terms*—Pair trading, cointegration, ontologies, SEC financial reports**

I. Introduction

In its simplest form, pair trading is based on the notion of a simple portfolio of two stocks whose weighted average produces a stationary time series [9]. Maintaining the prescribed percentage of each stock in the portfolio is expected to preserve this property, and make profit through a mean reversion strategy, i.e. where one sells the stock that goes up in relative terms, and buys the other one. The tests used to identify such pairs are purely numerical, and therefore constitute a so-called technical criterion, one that does not use deeper knowledge of either company. Nevertheless, a look at the pairs that pass the test often shows they have something in common, e.g. they represent companies from the same sector or even two types of share of the same company.

While initially the concept of pair trading was a highly profitable, closely guarded secret [5], it is common knowledge now, and the tests on which it can be based are readily available in standard statistical packages. This can reduce the potential for profit as an increasing number of traders adopt the same strategy and crowd the same stock market positions. It is therefore tempting to consider the use of additional information in the process of selecting the pairs in order to focus on those that are more likely to (1) meet the necessary condition, and (2) retain that property over time.

We have already remarked that different types of stock of the same company can be well suited to pair trading. Noticing that both stocks in such a pair are managed by the same CEO and board of directors, it is not unreasonable to consider whether two companies sharing a certain number of top decision makers have an above-average chance to form a suitable pair. The question becomes even more relevant as one realises that the necessary data is readily available, e.g. in the U.S. it forms part of public, mandatory filings known as periodic SEC (U.S Securities and Exchange Commission) reports. In this article, we show how this information can be extracted, represented as an ontology, and used to test the following two hypotheses:

1) Restricting the search for trading pairs to companies that share a certain number of prominent persons will affect the proportion of suitable pairs in a statistically significant way;
2) The above restriction will affect significantly the proportion of pairs that remain suited to pair trading over time.

II. Background

A. Ontologies and Their Benefits

An ontology [11], [12] is a structured database representing objects and their relations, known as *individuals* and *roles* in the field's parlance. An ontology allows for the grouping of individuals into classes (aka *concepts*). Both concepts and roles can form hierarchies, depending on the exact formalism used. The use of standard dictionaries of concepts and relevant roles not only assists the representation of domain knowledge, but also the easy integration of multiple ontologies. From an abstract point of view, the content of an ontology is equivalent to a set of Description Logic (DL) axioms, which provides the semantics for query languages (such as SPARQL [8]) and reasoners [4]. Machine learning algorithms specialised in the use of DL to represent data and models also exist [6], [2]. In this article, the information extracted from financial reports is

978-1-7281-2547-3/20/$31.00 ©2020 IEEE

converted to an ontology, and hand-written queries are used to extract the data sets needed to test the two research hypotheses.

The storage unit is a triplet with fields known as subject, predicate and object. These combine to a create a directed graph with entities represented by Unique Resource Identifiers (URI).

B. Data Source

U.S Securities and Exchange Commission (SEC) reports are formally specified documents that every company in the United States has to provide. It is an official government requirement, with the standards defined by SEC organisation. Here we use Form 3 SEC reports, containing a company's statement of beneficial ownership of securities by company insiders (directors or other officers) or major shareholders. Although SEC does not provide an API, it publishes an index file to the files that are published for all the company data based on company name, period, or report type. We have built a downloader and parser for all reports data using Edgar online index file and parsing the XML with their XBRL format and we generated an ontology from data in Forms 3, 4 and 5.

The SEC reports are in an annotated XML format containing there are different sections split between several XML files, such as header and body, each with a formal specification (Table I).

We parse this XML data to create the ontology for our subsequent experiments. A sample of it is shown in Figure 1.

We are now able to construct powerful SPARQL queries to extract information from this representation, and use rich visualisation tools to see and understand the data. A sample of such data visualisation is shown on Figure 2.

C. Pair Trading

The suitability of a pair of stock to pair trading is commonly tested via a statistical test applied to a pair of time series representing historical prices. Both the Engle-Granger and Johansen tests can be used for this purpose. These tests are far from perfect: research by Do et al. compares Engle and Granger's 2-step approach with Johansen's, finding the former to be influenced by the ordering of the variables and, on occasion, returning spurious estimators [7]. Gonzalo and Lee show that, for most situations, Engle-Granger is more robust than Johansen's likelihood ratio test [10]. These authors recommend using both Engle-Granger and Johansen tests in order to detect the possibility of false positive results, and subsequently increase the chance of avoiding that trap.

III. Methodology

The data used to test the first of the two hypotheses formulated in the introduction was based on Form 3 SEC reports for 2017 Q2[1] and close-of-business stock prices for 01/04/2017 – 30/06/2017.

The reports covered stock for up to 4767 companies. These were paired up at random until a sample of size 50,000 was generated. A second set of pairs was generated by selecting all pairs of companies that shared at least one link, i.e. a prominent officer or significant shareholder, as listed in the Form 3 reports. We refer to these as linked pairs in the rest of the paper. The time series of stock prices are then used to run a cointegration test on all pairs in both sets, and the percentage of cointegrated pairs in each set is calculated, and the results compared, with their significance tested using the χ^2 test.

Both the Engle-Granger and Johansen tests were used in all cases with the level of significance set to $p < 0.05$ for the former and the trace statistic used to reject the null hypothesis at a 95% confidence level for the latter. Only pairs that passed both tests were considered cointegrated. The same procedure was repeated to compare the ratio of cointegrated pairs in the random sample with the ratio for the set of pairs sharing an increasingly greater minimum number of links: 2, then 3.

[1]https://www.sec.gov/Archives/edgar/full-index/2017/QTR2/form.idx

TABLE I
SAMPLE SEC DOCUMENT SOURCE

```
<SEC-DOCUMENT>0001021432-17-000060.txt : 20170320
<SEC-HEADER>0001021432-17-000060.hdr.sgml : 20170320
<ACCEPTANCE-DATETIME>20170320184653
ACCESSION NUMBER:                0001021432-17-000060
CONFORMED SUBMISSION TYPE:       3
PUBLIC DOCUMENT COUNT:           1
CONFORMED PERIOD OF REPORT:      20170320
FILED AS OF DATE:                20170320
DATE AS OF CHANGE:               20170320
</SEC-HEADER>
<DOCUMENT>
<TYPE>3
<SEQUENCE>1
<FILENAME>primary_doc.xml
<DESCRIPTION>PRIMARY DOCUMENT
<TEXT>
<XML>
<?xml version="1.0"?>
<ownershipDocument>
    <schemaVersion>X0206</schemaVersion>
    <documentType>3</documentType>
    <periodOfReport>2017-03-20</periodOfReport>
    <noSecuritiesOwned>0</noSecuritiesOwned>
    <issuer>
        <issuerCik>0001693689</issuerCik>
    </issuer>
    <reportingOwner>
        <reportingOwnerId>
            <rptOwnerCik>0001450604</rptOwnerCik>
        </reportingOwnerId>
        <reportingOwnerRelationship>
            <isDirector>1</isDirector>
            <isOfficer>1</isOfficer>
            <isTenPercentOwner>1</isTenPercentOwner>
            <isOther>0</isOther>
            <officerTitle>vice president</
                officerTitle>
        </reportingOwnerRelationship>
    </reportingOwner>
</ownershipDocument>
</XML>
</TEXT>
</DOCUMENT>
</SEC-DOCUMENT>
```

It should be noted here that selecting pairs sharing links from the entire ontology is not particularly demanding in terms of time complexity. The result of this query leaves a small proportion of all possible pairs of stock (of 10^4 order of magnitude or lower), and running the cointegration test on them is also a viable task without the need for special computational resources. On the other hand, testing approx. 10^7 possible pairs for cointegration becomes difficult, which is why only a random sample (of size 50,000) of all pairs was selected to estimate the proportion of cointegrated pairs when no restriction on their pairing is imposed.

The second experiment compares the rates at which cointegrated pairs lose that property over a given period of time, depending on whether they share a certain minimum number of links, or are chosen completely at random. The period used here is 12 months, from 2017 Q2 to 2018 Q2. We start with the results from Experiment 1 on the 2017 Q2 data: all possible pairs are tested against the 'share-at-least-X-links' criterion (for $X = 1$, 2, and 3) to produce 3 data sets of cointegrated pairs at time 2017 Q2, one for each value of X. All cointegrated pairs from the sample of 50,000 random pairs constitute the fourth data set of cointegrated pairs. We then use the time series for the period 1 Apr 2018 — 30 June 2018 to find the number of pairs in each of the 4 data sets that are still cointegrated during that period. Again, the results for each of the first three sets are compared with the fourth data set or random pairs, and the statistical significance of any difference tested using the χ^2 test. Note that if a company has folded in the 12-months period, its (no longer existing) stock is assumed not cointegrated with any other stock. This experiment is then repeated using a sliding window of three months to test the retention of the cointegration relationship between quarters, and study how linked pairs and random pairs compare.

IV. Implementation

In order to create usable sets, every pair being added must be validated and any erroneous entries discarded. The criteria used for this filtering was as follows:

- Any reflexive pairs (e.g. (TSLA, TSLA)) are removed.
- If there is an occurrence of a pair and reversed pair (e.g. (INTC, AMD) and (AMD, INTC)) in a set, the first pair takes precedence.
- Any stock with missing entries for any trading day in the time period tested is removed.

To create the set of randomly selected companies, all companies in the ontology for the time period 2017 Q2 are selected using the following query:

```
PREFIX my: <http://york.ac.uk/>
SELECT ?t
WHERE { {
    ?company my:tradingsymbol ?t . } }
```

Given this list of companies, a set of all possible combinations of pairs is created. Pairs are randomly selected using the choice method from the random Python module [1], filtered and appended to the final set until it contains 50,000 pairs.

Creating sets of pairs sharing some number of links is a little more involved, pairs are first selected from the ontology using the following query:

```
PREFIX my: <http://york.ac.uk/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?person ?p ?t1 ?t2
WHERE { {
    ?person my:worksat ?company .
    ?person my:worksat ?othercompany .
    ?person foaf:name p .
    ?company my:tradingsymbol ?t1 .
    ?othercompany my:tradingsymbol ?t2 .
    FILTER(?t1 != ?t2)     } }
```

This query returns the relevant SEC report, the name of the person, and the tickers for both companies the person is affiliated with. This information gets fed into a dictionary where the key is the pair of companies and the value is a set containing names of people affiliated with the company. Once the dictionary is fully populated, the values for each key are replaced with the number of unique names linking each respective pair (i.e. the size of that set). With this dictionary, separate subsets of pairs are created for a gradually increasing minimum number of links between the two companies in each pair (see Table II).

For each company, matching tickers are found and their closing price for the aforementioned time period retrieved. The historical stock data is provided by the `yfinance` Python module (through Yahoo Finance) [3] and comprises of information for the stock using the date as an index. For example, a query on NVDA provides the information displayed in Table III.

V. Results

The results of the two experiments are listed in Tables IV–VIII. First, Table IV lists the results of Experiment 1 on 2017 Q2 data, comparing the number of cointegrated pairs in the random pair set with one of the sets of linked pairs (with a minimum number of links equal to 1, 2, and 3). The table shows that in all 3 cases the proportion of cointegrated pairs in the set of linked pairs is higher, and the results are statistically significant at the 95% level. The number of linked pairs rapidly drops as the required minimum number of links is increased. This affects negatively the significance levels (p-value) calculated by the χ^2 test, which is why this number was capped at 3 and results for higher values are not reported here.

TABLE II
Number of linked pairs for each quarter

	Number of links				
	1+	2+	3+	4+	5+
2017Q2	5143	395	247	216	104
2017Q3	4207	2484	2351	850	829
2017Q4	1885	226	90	71	63
2018Q1	9026	3152	169	94	83
2018Q2	4192	314	80	59	46

TABLE III
INFORMATION PROVIDED BY YFINANCE ON NVDA (NVIDIA CORP).

Date	Open	High	Low	Close	Adj Close	Volume
2017-03-31	109.010002	109.889999	108.400002	108.930000	107.806709	11020200
2017-04-03	108.949997	109.650002	107.419998	108.379997	107.262360	11130800
2017-04-04	103.400002	104.419998	100.339996	100.779999	99.740738	31782000
2017-04-05	100.019997	102.370003	99.500000	100.029999	98.998482	18676200
2017-04-06	100.239998	101.250000	98.410004	100.760002	99.720970	15878000
...	...	...	...	...	...	...
2017-06-23	158.679993	159.320007	153.220001	153.830002	152.404037	27214700
2017-06-26	155.160004	156.600006	148.330002	152.149994	150.739578	26599000
2017-06-27	151.440002	151.789993	146.350006	146.580002	145.221252	24987300
2017-06-28	149.320007	151.940002	145.750000	151.750000	150.343323	24873700
2017-06-29	150.600006	150.720001	144.080002	146.679993	145.320282	26610600

TABLE IV
EXPERIMENT 1: LINKED VS RANDOM PAIRS (2017 Q2)

	Coint.	Non-coint.	Total	Coint/Total
Random	3,406	46,594	50,000	6.81%
Links$\geq$ 1	399	4,744	5,143	7.76%
	χ^2 test	p = 0.0108		

	Coint.	Non-coint.	Total	Coint/Total
Random	3,406	46,594	50,000	6.81%
Links$\geq$ 2	41	354	395	10.38%
	χ^2 test	p = 0.0051		

	Coint.	Non-coint.	Total	Coint/Total
Random	3,406	46,594	50,000	6.81%
Links$\geq$ 3	25	222	247	10.12%
	χ^2 test	p = 0.0397		

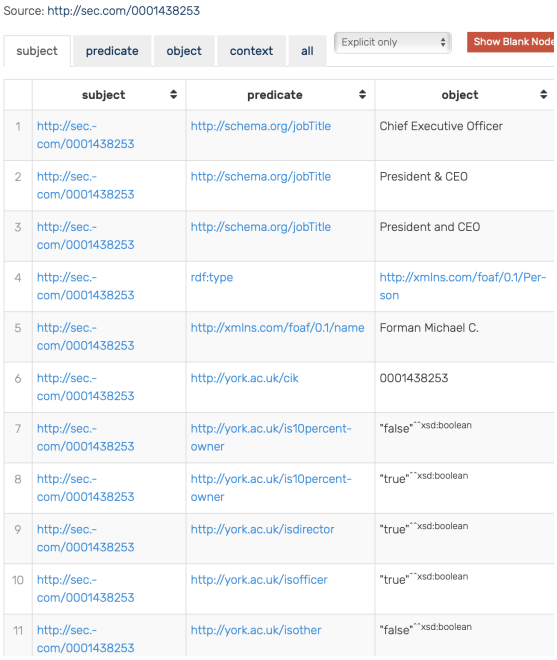

Fig. 1. Ontology sample

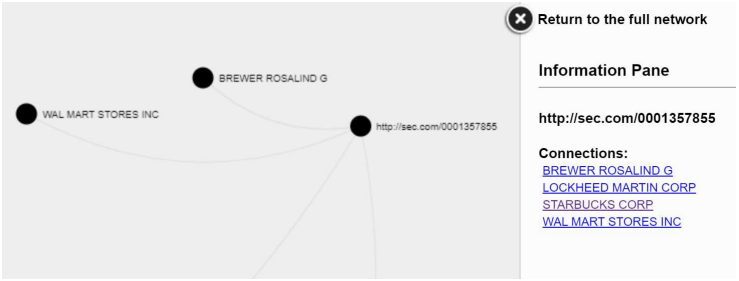

Fig. 2. Ontology visualisation with a person represented by an URI

TABLE V
EXPERIMENT 1: COUNT/TOTAL RATIO ($p < 0.05$ WINNERS IN BOLD)

	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2
Random	6.81%	5.20%	5.29%	6.59%	**5.87%**
Links$\geq$ 1	**7.76%**	**8.72%**	5.15%	**9.92%**	4.53%
p-value	0.0108	6E-22	0.78	1E-29	0.0004

	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2
Random	6.81%	5.20%	5.29%	6.59%	5.87%
Links$\geq$ 2	**10.38%**	**10.02%**	**10.62%**	**8.98%**	4.46%
p-value	0.0051	4E-25	0.0004	2E-7	0.2895

	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2
Random	6.81%	5.20%	5.29%	6.59%	5.87%
Links$\geq$ 3	**10.12%**	**10.12%**	**15.16** %	7.10%	10.00%
p-value	0.0397	8E-25	1E-5	0.7897	0.1162

TABLE VI
EXPERIMENT 2: ATTRITION AMONG COINTEGRATED PAIRS

	2017Q2	$\rightarrow$ 2018Q2	2018Q2 / 2017Q2
Random	3,406	272	**7.99%**
Links$\geq$ 1	399	18	4.51%
	χ^2 test	p = 0.0133	

	2017Q2	$\rightarrow$ 2018Q2	2018Q2 / 2017Q2
Random	3,406	272	7.99 %
Links$\geq$ 2	41	6	14.63%
	χ^2 test	p = 0.1202	

	2017Q2	$\rightarrow$ 2018Q2	2018Q2 / 2017Q2
Random	3,406	272	7.99 %
Links$\geq$ 3	25	2	8.00%
	χ^2 test	p = 0.9979	

TABLE VII
RESULTS USED TO CALCULATE THE SURVIVAL RATE OF COINTEGRATED PAIRS

t: start of period $t + \Delta t$: end of period	2017Q2– 2017Q3	2017Q3– 2017Q4	2017Q4– 2018Q1	2018Q1– 2018Q2
All pairs in DB with at least one link	5143	4207	1885	9026
Cointegrated linked pairs at time t	399	394	91	983
Remaining coint. pairs at time $t + \Delta t$	48	31	3	107

t: start of period $t + \Delta t$: end of period	2017Q2– 2017Q3	2017Q3– 2017Q4	2017Q4– 2018Q1	2018Q1– 2018Q2
All pairs in DB with at least 2 links	395	2484	226	3152
Cointegrated linked pairs at time t	41	263	25	308
Remaining coint. pairs at time $t + \Delta t$	5	22	2	38

t: start of period $t + \Delta t$: end of period	2017Q2– 2017Q3	2017Q3– 2017Q4	2017Q4– 2018Q1	2018Q1– 2018Q2
All pairs in DB with at least 3 links	247	2351	90	169
Cointegrated linked pairs at time t	25	254	14	12
Remaining coint. pairs at time $t + \Delta t$	3	20	2	4

t: start of period $t + \Delta t$: end of period	2017Q2– 2017Q3	2017Q3– 2017Q4	2017Q4– 2018Q1	2018Q1– 2018Q2
Random sample of pairs	50000	50000	50000	50000
Cointegrated pairs at time t	3406	2604	2894	3110
Remaining coint. pairs at time $t + \Delta t$	255	235	301	308

TABLE VIII
SURVIVAL RATE OF COINTEGRATED PAIRS AT THE END OF EACH 3-MONTH PERIOD

t: start of period $t + \Delta t$: end of period	2017Q2– 2017Q3	2017Q3– 2017Q4	2017Q4– 2018Q1	2018Q1– 2018Q2
Pairs with at least one link	**12.03**%	7.87%	3.30%	10.89%
Random sample pairs	7.49%	9.02%	**10.40**%	9.90%
χ^2 test p-value	.0040	.4895	.0399	.4232

t: start of period $t + \Delta t$: end of period	2017Q2– 2017Q3	2017Q3– 2017Q4	2017Q4– 2018Q1	2018Q1– 2018Q2
Pairs with at least two links	12.20%	8.37%	8.00%	12.34%
Random sample pairs	7.49%	9.02%	10.40%	9.90%
χ^2 test p-value	.3028	0.7437	.7211	.2264

t: start of period $t + \Delta t$: end of period	2017Q2– 2017Q3	2017Q3– 2017Q4	2017Q4– 2018Q1	2018Q1– 2018Q2
Pairs with at least three links	12.00%	7.87%	14.29%	**33.33**%
Random sample pairs	7.49%	9.02%	10.40%	9.90%
χ^2 test p-value	.4384	.5728	.6743	.0264

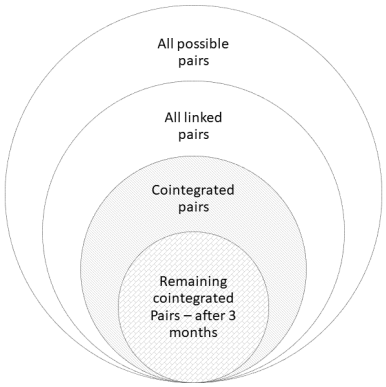

Fig. 3. Selecting linked pairs for Experiment 2

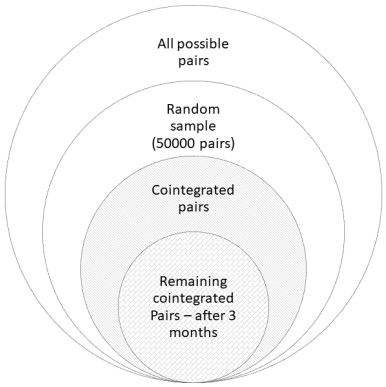

Fig. 4. Selecting random pairs for Experiment 2

The experiment was repeated with data from further 4 consecutive quarters in order to study how stable our first findings are over time. Table V lists only the number of cointegrated pairs as a fraction of the set of pairs in question. With one exception, all results show that selecting only linked pairs is significantly better than using random pairs with respect to this criterion or there is no clear winner. The best results are achieved for Links ≥ 2 with 4 'wins' and one 'draw' in the studied period.

While the first experiment tested – and confirmed – the hypothesis that information about links between companies can be used to increase the likelihood of selecting pairs that are cointegrated, the second experiment focuses on the attrition levels among cointegrated pairs over time, that is, measuring the percentage of pairs that remain cointegrated at the end of the studied period. Again, the attrition in a set with a given minimum number of links between each pair is compared to that of a set of random pairs.

We test this property on windows of two different sizes. The survival rates of cointegrated pairs over a single period of 12 months is reported in Table VI. Here the only statistically significant difference is observed between the set with Links ≥ 1, and the sample of random pairs, with the latter showing a lower attrition rate. The same experiment was then repeated for a number of shorter time periods of 3 months each. Figures 3–4 show the way all relevant counts were produced. The raw counts of linked pairs, cointegrated pairs, and pairs that survived the whole period are reported in Table VII. The corresponding survival rates and the statistical significance of their comparison are shown in Table VIII. It can be seen from the table that with a couple of exceptions, using only linked pairs does not affect the survival rate of the cointegrated pairs found.

VI. Discussion

This article shows that converting information from the SEC reports into an ontology provides a way to answer queries of quadratic complexity, which would be unfeasible without such automation. The (rather restricted) part of the reports extracted here is already showing its potential relevance to traders. We proposed two hypotheses based on a combination of common expectations about the relevance of additional information (which is commonly used in fundamental trading), and generalisations from previous insights about trading pairs of stock of the same company.

The results on the first hypothesis have statistical significance for certain common values of its only parameter. The findings here have the potential to inform a more efficient search for companies suited to pair trading. At the same time, it is also evident that while our initial intuition was good, the choice of parameter, namely, the minimum number of links between a pair of companies, was of importance.

The operational relevance of the second experiment is that the percentage of cointegrated pairs that remain cointegrated over three months is not affected negatively by the use of linked pairs only, which means that we can benefit from the advantages of this approach, which increases the probability of finding a cointegrated pair without any downside on the expected longevity of such pairs.

A few final considerations: it is also interesting to look at each of the six pairs of stock linked through $2+$ links that remain cointegrated after a year (Table IX and Appendix). Here one finds that with one exception, the two members of a pair have very similar names. Pairs of this type could potentially be selected on the basis of simple string similarity, at the risk of many false positives. Note that in our approach no information about name similarity was used. We had also thought of a simpler criterion for pre-selecting pairs, where both companies would have to be in the same sector. However, it is not clear what the relevance of such criteria would be in the case of trusts and funds with their diversified portfolios.

TABLE IX
SURVIVING COINTEGRATED PAIRS: LINKS$\geq$ 2, 2017 Q2$\rightarrow$ 2018 Q2

Pair 1:
MMT: MFS Multimarket Income Trust
MCR: MFS Charter Income Trust
Pair 2:
VCV: Invesco California Value Municipal Income Trust
VKI: Invesco Advantage Municipal Income Trust II
Pair 3:
PMM: Putnam Managed Municipal Income Trust
PMO: Putnam Municipal Opportunities Trust
Pair 4:
BFY: BlackRock New York Municipal Income Trust II
BQH: BlackRock New York Municipal Bond Trust
Pair 5:
BQH: BlackRock New York Municipal Bond Trust
MNE: BlackRock Muni New York Intermediate Duration Fund, Inc.
Pair 6:
DVA: DaVita Inc.
KND: Kindred Healthcare, Inc.

REFERENCES

[1] random library. https://docs.python.org/3.8/library/random.html. Accessed: 2020/08/05.

[2] E. Algahtani and D. Kazakov. Conner: A concurrent ILP learner in description logic. In *Proc. of the* 29^{th} *International Conf. on Inductive Logic Programming*, 2020.

[3] R. Aroussi. yfinance library. https://github.com/ranaroussi/yfinance. Accessed: 2020/05/09.

[4] F. Baader, I. Horrocks, C. Lutz, and U. Sattler. *An Introduction to Description Logic*. Cambridge University Press, 2017.

[5] R. Bookstaber. *A Demon Of Our Own Design*. Wiley, 2006.

[6] L. Bühmann, J. Lehmann, and P. Westphal. DL-Learner – a framework for inductive learning on the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 39:15–24, 2016.

[7] B. Do, R. Faff, and K. Hamza. A new approach to modeling and estimation for pairs trading. In *Proceedings of 2006 financial management association European conference*, pages 87–99. Citeseer, 2006.

[8] B. DuCharme. *Learning SPARQL*. O'Reilly Media, second edition, 2013.

[9] E. Gatev, W. N. Goetzmann, and K. G. Rouwenhorst. Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies*, 19(3):797–827, 2006.

[10] J. Gonzalo and T.-H. Lee. Pitfalls in testing for long run relationships. *Journal of Econometrics*, 86(1):129–154, 1998.

[11] J. Lehmann and J. Volker. *An Introduction on Ontology Learning*.

[12] H. Qu, M. Sardelich Nascimento, N. N. Qomariyah, and D. Kazakov. Integrating time series with social media data in an ontology for the modelling of extreme financial events. In *LREC 2016 Proceedings*, volume Joint Second Workshop on Language and Ontology & Terminology and Knowledge Structures, pages 57–63. European Language Resources Association (ELRA), 2016.

APPENDIX: APRIL–JUNE 2018 CLOSE-OF-DAY PRICE FOR SURVIVING COINTEGRATED PAIRS WITH LINKS≥ 2

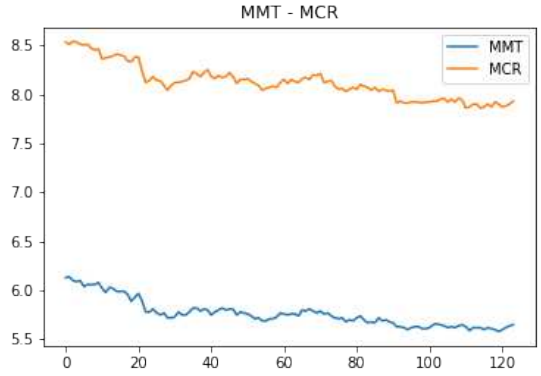

Fig. 5. MMT and MCR close-of-day price

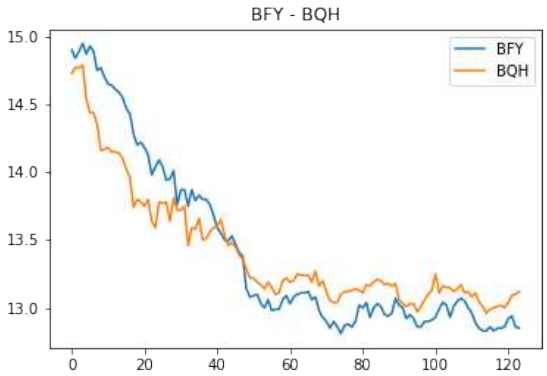

Fig. 8. BFY and BQH close-of-day price

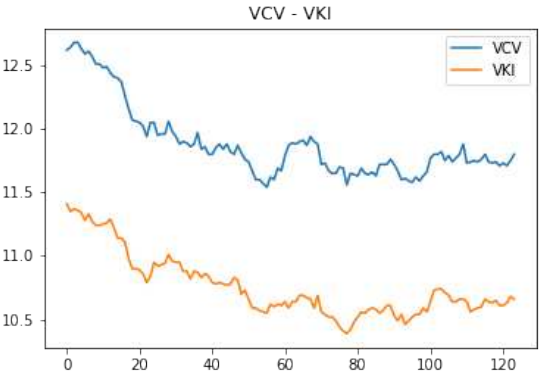

Fig. 6. VCV and VKI close-of-day price

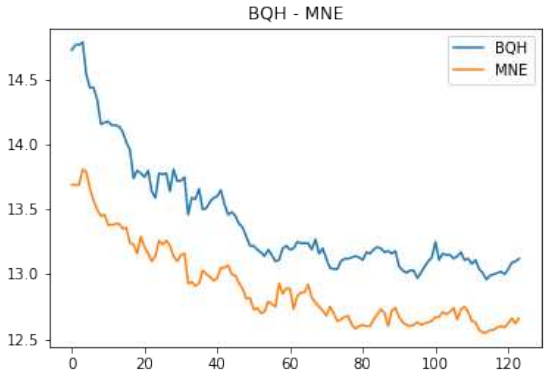

Fig. 9. BQH and MNE close-of-day price

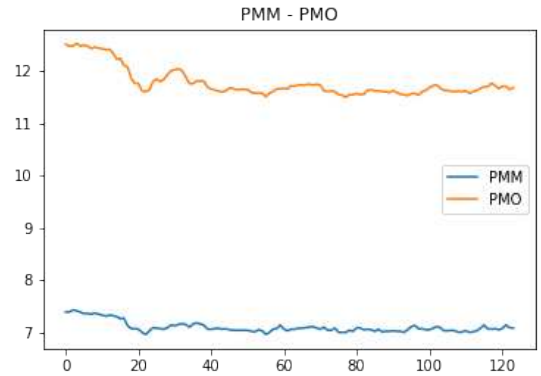

Fig. 7. PMM and PMO close-of-day price

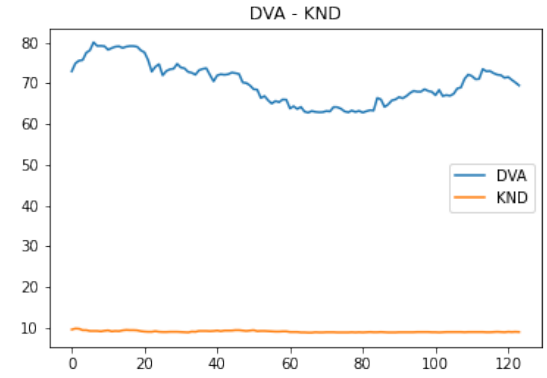

Fig. 10. DVA and KND close-of-day price