The Elias Distance

By: Gabe Nardi-Huffman and Mark Fesenmeyer

Abstract:

The Elias Distance provides a framework for objectively comparing the quantitative characteristics of one or more companies. In order to compare two companies, we first define a set of quantitative characteristics that we would like to compare them on (we suggest either Income-statement, balance sheet, cash-flow statement or overall financial statement comparisons). Then, we scale the financial statements by allowing them to be represented as percentages of certain line items (for example, gross profit as a percentage of sales). This inherently eliminates the characteristic of size of the companies, which may or may not be crucial to the individual analyst's analysis. Next, we write a vector composed of the company's financial statements. This allows us to think of each company as a point in *n* space, defined by a flexible set of quantitative characteristics.

The natural next step is to calculate the distance between two companies using the traditional Euclidean distance formula. The resulting number, although not helpful in isolation, gives us an objective data point for which to say: Company X is 5.78 EUs (Elias Units) away from Company Y, for example. We can then employ a variety of analytical exercises based off of this information such as computing ED stats throughout a portfolio, industry or entire stock universe. Thus we can build an environment where this simple data point allows us to compare two companies at a point in time, two portfolios of companies, or one or more companies across time. A simple but powerful concept, the Elias Distance provides a doorway into useful analyses such as the path dependency of two companies throughout time (which, if using the OED (overall Elias Distance), would give the tendency of their financial characteristics to move in tandem), and also allows us to approach screening from the bottom up - starting with a company whose

characteristics we like and quickly and efficiently constructing a list of necessarily similar companies. We provide several of our ideas for the uses of this type of analysis, although the main scope of this work is to provide the process of constructing Elias Distances and leaves the analytical possibilities up to the user. We offer the following definition:

> Definition: *The Elias Distance between companies X and Y is the Euclidean "distance" between X and Y, with the coordinates of each company defined by a chosen set of quantitative characteristics that apply to both X and Y.*

Introduction:

This project began as an answer to the following question: Given 2 distinct public companies, how quantitatively similar are they? When I think about this question, I think the normal response is to start rattling off relevant statistics: "Well, they both have similar revenue growth rates, their net income margins are sort of close, their debt profiles are different, etc." There are several drawbacks to this kind of thinking. First, it becomes exceedingly inefficient to describe a large number of financial characteristics. Second, how can we effectively compare a group of companies quantitatively? Would we go through and say that Company A shares characteristics with Company B, but some characteristics are different, and that Companies C and D are mostly similar but Companies B and C are not because of these characteristics that are different? I need not go any further with this analogy - it is clear how messy this could get.

Now, one way to get a list of companies with similar characteristics would be to screen for those characteristics. But this in essence provides a top-down way of comparing companies (i.e. these companies are similar because we screened for certain characteristics). The Elias Distance, on the other hand, allows its user to take a bottom-up approach to comparing companies (these are the companies we have, now how similar are they quantitatively?). So, the Elias Distance in part is a way of simplifying and speeding up this process of comparing 2 or more companies, or 1 or more companies across time. Although in isolation the Elias Distance between two companies isn't of much use, by computing Elias Distances between companies in an industry, or for an entire index or universe of stocks or for one or more companies across time, we can start to build an environment within which we can compare Elias Distances between companies to, say, the broader industry or universe average. This allows us to make statements such as: “with an OED (overall Elias Distance) of 6.7897, these two companies rank in the top decile of financial statement similarity out of the S&P 500.” And, if we introduce a path dependency element, we can compare the tendency of companies’ financial statements to move together over time. Does Company A exhibit margin erosion while Company B does not even though both of their balance sheets remain similarly configured? The Elias Distance allows us to approach this type of question with objectivity.

In summary, we hope to offer a framework for answering the following:

- Given a set of n companies, which two are the most similar?
- Which groups of companies are more similar than other groups?
- How has the similarity between the companies in our set changed over time?
- What does it even mean to compare two companies?

The quick and dirty of this paper is that there doesn’t appear to be a quick and dirty answer to these questions (at least not an efficient answer that doesn't require working through line item by line item). We hope to provide a framework for comparing companies in an efficient and thorough way.

In this paper we will introduce the basic framework, along with some interesting applications that we think become possible and easier when you approach company and portfolio comparisons through an Elias Distance framework.

<u>The Process (Our Approach):</u>

<u>Methodology:</u>

Here we discuss the methodology we have created for comparing companies. There are many complicating factors we could take into account, so this is a simplified version.

The way we look at company comparison is that we would like to be able to standardize the process and incorporate a multitude of quantitative characteristics into our analysis in an efficient manner. Instead of going one by one and examining the attributes of a particular company set, we want to be able to compare many different attributes at once, thereby marking an efficient and more thorough process.

We think about each company as being defined by its quantitative characteristics (although many qualitative characteristics like quality of management and firm culture are quite important, for the purposes of this analysis we assume such qualitative characteristics are captured in the financial snapshot of the company (i.e. attributes like better managements and favorable industry classification will manifest themselves in higher margins and revenue growth rates).

By thinking of each company as a vector made up of its quantitative characteristics, we can use the distance formula to create a standard measure of comparison (we call this the Elias Distance). Jesus I'm explaining this so badly already.

First, consider a certain set of quantitative characteristics that we think aptly describe a company:

$$Company = \{\text{Characteristic1}, \text{Characteristic2}, \ldots, \text{CharacteristicN}\}$$

We can include any number of financial characteristics, but in order for the analysis to work we have to use a standard set across all companies. We think it makes sense to try and capture the three financial statements (Income Statement, Balance Sheet, and Cash Flow Statement). In order to remove the effect of size, we can construct common size financial statements, where we take each financial statement line item as a percentage of a chosen top line item. For example, looking at the income statement, we can gather line items as a percentage of revenue and compute: Gross margin (gross profit/revenue), SG&A Margin (SG&A as a percentage of revenue), R&D as % of revenue, taxes, etc. Look to Appendix X at the end of this report for a full description of our chosen set of characteristics. (spoiler alert, Appendix X does not exist yet)

Since every public company has to report financial statements, we should be able to gather a relatively stable dataset with which we can start to compare companies.

After creating our set of characteristics, we compile data and compute the Elias Distance between each pair of companies, where the Elias Distance is found simply by applying the distance formula to each pair (described in much more detail below).

The resulting data set is a matrix listing the Elias Distances between each pair. This is the core result of our analysis, as we now have a standard quantitative framework with which to compare companies. A number of interesting applications follow as a result, and we cover some of our favorites later on.

Step 1: Define characteristic universe.

The first step in the process is to choose the quantitative attributes that we want to use to compare our companies. In this step, we define our ecosystem, and this is the backbone of the analysis because it determines the types of relationships that will surface among the companies.

The method we use to construct our universe is to use the three financial statements available for every public company: The income statement; the balance sheet; and the cash flow statement. To remove the effects of size, we convert the three financial statements into what are called common size financial statements. What this means is that we take each line item as a percentage of some top line item. For example, looking at the income statement, we can take each line item as a percentage of revenue. Similarly, looking at the balance sheet we could compute each line item as a percentage of total assets, and for the cash flow statement we could take each item as a percentage of Operating Cash flow (or we could also take the income statement items as a percentage of total assets).

The point of removing the effect of size is to get rid of the weighting problem that would be present if we didn't adjust (bigger companies would always look more similar than smaller companies). Note: you could definitely make the case that size is an important quantitative characteristic with which to compare companies, but for the sake of this paper we remove it.

It is important to think critically about what characteristics you want to include because the results that come out of any Elias Distance analysis will be determined by the inputs we select.

Step 2: Gather data.

This step is self-explanatory. Once we figure out what data points we want to use in our analysis, we have to put together our data set.

Let C(i,j) = The value of characteristic j for company i. Then we want to compile:

Company1 = {C(1,1), C(1,2),..., C(1,N)}

Company2 = {C(2,1), C(2,2),..., C(2,N)}

.

.

.

CompanyM = {C(M,1), C(M,2),..., C(M,N)}

Step 3: Create an Elias Distance matrix.

Once we have all of the related data, we have effectively defined each company as a vector of the characteristics that we chose. Now we can go ahead and create the Elias Distance Matrix. This is the core of our analysis, and is done as follows:

Take the vectors for each set of companies and use the distance formula to compute the Elias Distance between each company. So if we have:

CompanyA = < C(A,1), C(A,2),..., C(A,N) >

CompanyB = < C(B,1), C(B,2),..., C(B,N) >,

Then we can compute:

Sqrt[(C(A,1)-C(B,1))^2 + (C((A,2)-C(B,2))^2 + ... + (C(A,N) – C(B,N))^2]

This is just the classic distance formula calculation. We can label this ED(A,B), which means the Elias Distance between Company A and Company B. The resulting matrix will look like the following (keeping in mind that ED(A,A) = 0):

Example of a portfolio with 10 companies:

ED Matrix 1: Portfolio **Note: ED(XY) = ED(YX)

	Com1	Com2	Com3	Com4	Com5	Com6	Com7	Com8	Com9	Com10
Com1	0	ED(1,2)	ED(1,3)	ED(1,4)	ED(1,5)	ED(1,6)	ED(1,7)	ED(1,8)	ED(1,9)	ED(1,10)
Com2	ED(2,1)	0	ED(2,3)	ED(2,4)	ED(2,5)	ED(2,6)	ED(2,7)	ED(2,8)	ED(2,9)	ED(2,10)
Com3	ED(3,1)	ED(3,2)	0	ED(3,4)	ED(3,5)	ED(3,6)	ED(3,7)	ED(3,8)	ED(3,9)	ED(3,10)
Com4	ED(4,1)	ED(4,2)	ED(4,3)	0	ED(4,5)	ED(4,6)	ED(4,7)	ED(4,8)	ED(4,9)	ED(4,10)
Com5	ED(5,1)	ED(5,2)	ED(5,3)	ED(5,4)	0	ED(5,6)	ED(5,7)	ED(5,8)	ED(5,9)	ED(5,10)
Com6	ED(6,1)	ED(6,2)	ED(6,3)	ED(6,4)	ED(6,5)	0	ED(6,7)	ED(6,8)	ED(6,9)	ED(6,10)
Com7	ED(7,1)	ED(7,2)	ED(7,3)	ED(7,4)	ED(7,5)	ED(7,6)	0	ED(7,8)	ED(7,9)	ED(7,10)
Com8	ED(8,1)	ED(8,2)	ED(8,3)	ED(8,4)	ED(8,5)	ED(8,6)	ED(8,7)	0	ED(8,9)	ED(8,10)
Com9	ED(9,1)	ED(9,2)	ED(9,3)	ED(9,4)	ED(9,5)	ED(9,6)	ED(9,7)	ED(9,8)	0	ED(9,10)
Com10	ED(10,1)	ED(10,2)	ED(10,3)	ED(10,4)	ED(10,5)	ED(10,6)	ED(10,7)	ED(10,8)	ED(10,9)	0

Step 4: Analyze data.

Now that we've calculated all of our Elias Distances, we have successfully created a standardized framework for comparing companies within our universe. While alone the ED(a,b) doesn't mean much, once we have computed all of the Distances within our set, we can use them relationally to compare our companies. We can now say that, based on our inputs, and given the set of calculated Elias Distances EDSET = {ED(1,2), ED(1,3),..., ED(i,j),...ED(M,1)}, then the two "most similar" companies in terms of aggregate financial statements are companies a,b with

ED(a,b) = min{EDSET}. That is the core takeaway from this framework. We now have a standard way of efficiently comparing the companies in our universe, and can run some really interesting (at least I think it's interesting) analysis based on our results. The applications we can create using the Elias Distance Matrix are discussed in the following section.

(optional part – not sure if I want to take this out)

Simple Example:

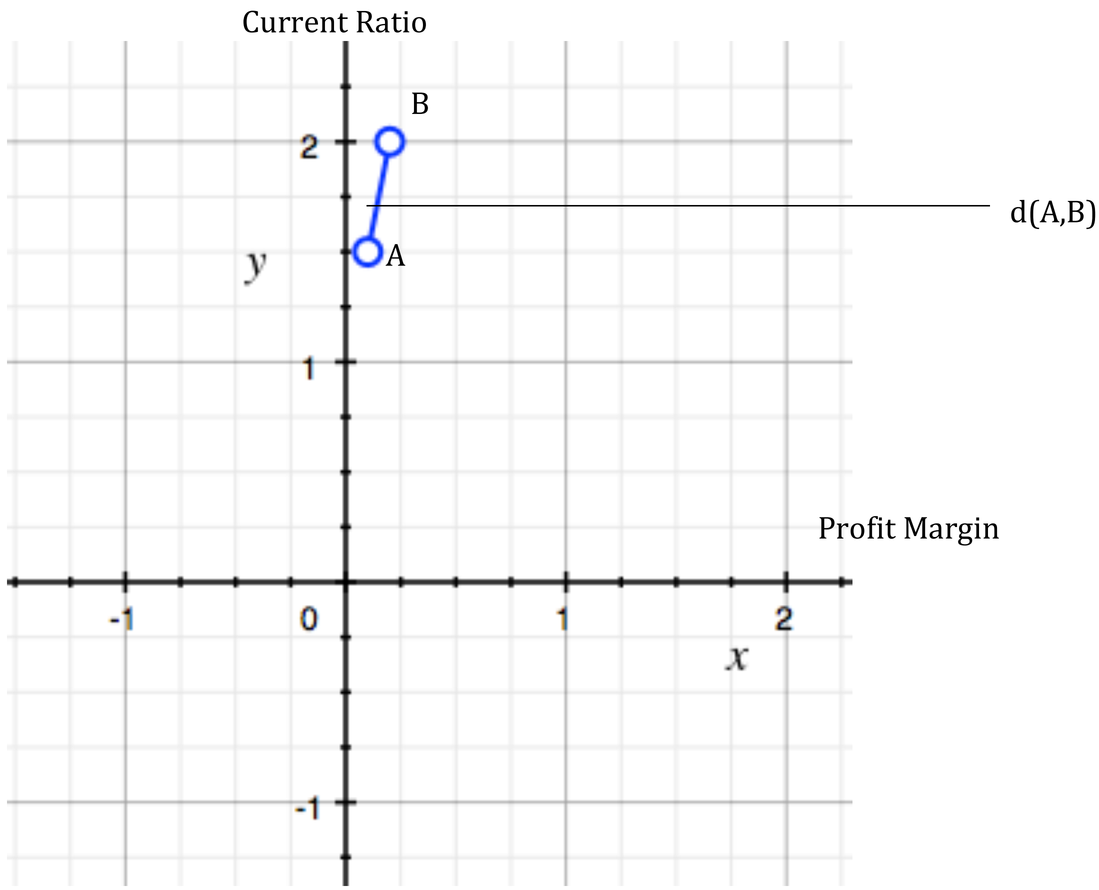
First, look at what happens if we apply just one financial metric, namely, profit margin to each company. Suppose the profit margin of A is 10% and the profit margin of B is 20%. Then we can represent companies A and B as points on a line where the axis represents profit margin:

———————A(10%)——————B(20%)————————————(Profit Margin)

Then the distance between A and B is d(A,B) = [(20-10)^2]^(1/2) = 20 - 10 = **10** . Notice that this distance represents how close the profit margin of A is to that of B. The smaller the distance between the points on this line, the closer the profit margins were together. Hence if d(A,B) = 0 then Profit Margin (A) = Profit Margin (B). Although this seems obvious at this stage, the notion of distance implying similarity is the core principle here and will become very important later.

Continuing on with this notion of defining each axis as a financial metric, and each company as a point in the space, we can extend our model to 2 dimensions.

Now, let {Cn} = {Profit Margin, Current Ratio}. Take A = (.1, 1.5) and B = (.2, 2), meaning that Company A has a profit margin of 10% and a current ratio of 1.5 and Company B has a profit margin of 20% and a current ratio of 2.



In this case, the X-axis represents Profit Margin and the Y axis represents Current Ratio. Again, the distance between A and B given by the standard Euclidean distance formula shows how similar the coordinates of A and those of B are, which is the same as saying that the financial metrics we are using to evaluate A and B are similar. The closer A and B become, the greater degree their quantitative characteristics converge.

With the same thoughts in mind, we can generalize and let

A = < C(A,1), C(A,2),…., C(A,N) > and
B = < C(B,1), C(B,2),…., C(B,N) >

for any number of quantitative characteristics used to describe a company. Then each "characteristic" represents its own axis in n-space. Since the distance formula applies to all of n-space, we can attain the “distance” between any 2 companies as long as we are using the same set of quantitative descriptors for each company. The distance between the 2 points/companies would then represent the degree to which their quantitative characteristics were similar.

d(A,B) = sqrt[(C(A,1) – C(B,2))^2 + (C(A,2) – C(B,2))^2 + …….. + (C(A,1) – C(B,2))^2]
(Standard Euclidean distance formula)

Thus, in this manner we could theoretically incorporate all of the quantitative information available into each coordinate point A and B, and by calculating the distance between the points we get a number that represents the degree to which the full list of quantitative characteristics of company A is comparable to that of company B. Put in other words, the closer A and B get in n-space, the more comparable the companies are by a strictly quantitative standpoint. Hence this distance is as close an approximation as possible for how similar the companies are (quantitatively).

Applications:

Main buckets:

- Basic Screening:
    - We can use the Elias distance matrix as a screening tool. Instead of picking quantitative characteristics a la carte, we can now pick a company we like and an Elias distance radius, and use that as a tool for idea generation tool. For example, find all companies that are “geosimilar” to company A. Find all the stocks that are geosimilar to my portfolio (so companies that are like what I have in my portfolio but aren't currently in it).
- Historical Analysis (“Historian Search”):
    - I think it would be useful to track how companies have trended over time. This is the highest-compute application because you would need a way to compare a

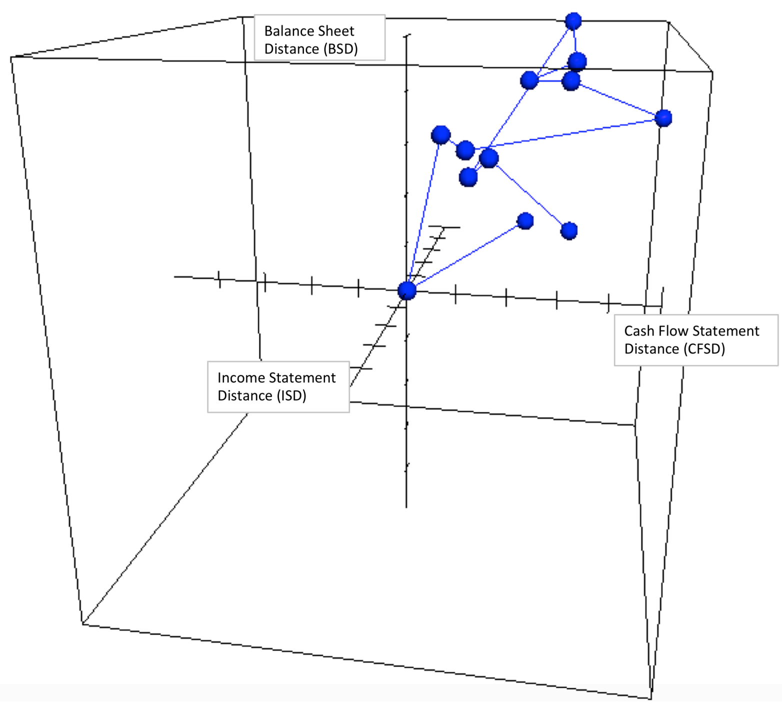
company to all other companies in the universe across different time-frames simultaneously. But the thing I think could be interesting would be to use a path dependency algorithm, so say basically who across history has had financial statements that were trending in this same direction? I guess in this case you would be comparing the same attributes but just indexing at time T. Is it as simple as that? Probably not. You would need the historical data from every company but that's essentially what capIQ is.

    - The main takeaway from this application is that you could, in a matter of seconds, answer the question: Which companies today look like Amazon did in the first few years of going public (from a financial statements perspective)? Find me the next Amazon!

- Portfolio comparisons:
    - You can apply the same company-specific application of the Elias Distance to portfolios. What other portfolios out there have financial characteristics similar to mine? The way to do it is to create a portfolio vector (asset-weighted vector using position size as the weight) that represents your portfolio and then apply the same analysis. I think it could be cool to see what other portfolios are similar to mine and how they have fared in comparison. Also a risk management tool, and you could maybe go so far as to look at ETF arbitrage possibilities:
        - ETF arbitrage: Find portfolios of exchange traded funds (ETFs) that have really similar fundamental characteristics but are trading at very different valuations. The "cheaper" portfolio should outperform from a price perspective if the fundamentals are very close.
- Risk management (Capital Cube):
    - I think it would be useful in risk management to **visualize your portfolio**. So the way to do this is to compute an ISD, BSD and CFSD (Income Statement

Distance, Balance Sheet Distance and Cash Flow Statement Distance) for each stock in the portfolio and them chart them. It will form a cube because all of the data points are non-negative (distance between a and b must be non-negative) What else could we do with risk management? It feels like there would be a lot here.

Capital Cube:



Conclusion?