Strategically Comparable Firm-Finding Algorithm (CFFA):
A Strategic Management Approach to Valuation

Hyoung-Goo Kang
Hanyang University Business School
222 Wangsimni-ro, Seongdong-gu
Seoul, Korea, 047633
E-mail: hyoungkang@hanyang.ac.kr

Abstract
This paper proposes an intuitive valuation framework based on the ideas of strategic management, regression, and portfolio theory. I call the framework the "Comparable Firm Finding Algorithm" (CFFA). The CFFA includes, as a special case, the Comparable Company Analysis (CCA), one of the most popular valuation approaches. Although the CFFA generalizes the CCA, it is equally simple, as well as being flexible enough to permit wide applications and variations. Furthermore, since the CFFA is based on regression analysis, researchers can easily leverage their data-analysis techniques to incorporate instrument variables, machine learning, and big data. This paper demonstrates the merits of the CFFA by addressing a key challenge in finance, accounting, and strategic management: how to compute the value of intangible strategic resources. The CFFA is also used to identify ESG contributions, to assess intangible (price-to-intangible ratios), to evaluate combinative capabilities, to price entrepreneurial activities, and to develop investment strategies and scenario planning.

Keywords: valuation, resources, intangibles, regression, venture capital, machine learning.

INTRODUCTION

Standard valuation and investment models do not work well in the era of intangible, more generally unobservable strategic resources and capabilities (Gu & Lev, 2017). The rise of the digital economy and intangibles-based business models, such as GAFA in the US and BATX in China, have been changing the characteristics of capitalism (Haskhel & Westlake, 2017). Nevertheless, while intangibles, such as data and algorithms, have become an important source of value creation and competitive advantage, it is very hard to evaluate them.[1] This paper addresses that problem with a simple idea.

The question of how to evaluate intangibles lies at the heart of the management literature. *Hard-to-objectively-assess* firm-specific characteristics cause and sustain heterogeneous firm performances. As strategic-management scholars have argued, a firm's strategic resources determine its values by shaping its business model and strategies; ultimately, they determine its sustainable competitive advantage (Barney, 2001; Barney et al., 2001; Barney & Arikan, 2001; Mahoney & Pandian, 1992; Peteraf, 1993; Wernerfelt, 1984, 1995). Dierickx and Cool (1989) characterize resources as time-compression diseconomies, asset mass efficiencies, inter-connectedness, asset erosion, and causal ambiguity. Then, these characteristics also describe data, some of the most important intangible assets in the digital economy (Brynjolfsson & Kahin, 2002). Specifically, data beget data; it is very difficult to replicate a successful social network service by reaching its critical *mass efficiency* or designing a business model that automatically collects data. Data become very valued when they are *interconnected*, but their value can *decay* quickly (e.g., behavioral data). Data-driven business models tend to generate network effects that *blur causal relationships*.

[1] Why Book Value Has Lost Its Meaning (*The Economist*, 2019).

This paper proposes a new idea to overcome the limitations of existing valuation models when assessing intangibles, such as data and strategic resources. The idea is based on the intuition of strategic-management approaches, such as the resource-based view, and combinative capabilities in the knowledge-based view (Kogut & Zander, 1992).[2] The idea generalizes the comparable company analysis (CCA), a widely popular approach in both academia and industry. Indeed, identifying a comparable firm is one of the most important and practical ways of evaluating a firm or a project. The most popular analytic tool, the CCA, tends to focus on one or two financial ratios; this paper extends that approach in a novel way.

For example, the CCA averages the EBITDA/EV of comparable firms and then divides the target firm's EBITDA (Earnings Before Interest, Taxes, Depreciation, and Amortization) by the average to determine its EV (Enterprise Value). The question then arises: How can the difference between other financial ratios (e.g., the price-to-book and price-to-sales ratios) be incorporated? How can this approach be implemented when big data, including qualitative and alternative data, become available for both target and comparable firms? Must one simply ignore such information?

Despite the popularity of the CCA and the comparable-firm approach, analysts are unaware of how to incorporate diverse information about a firm's resources and capabilities—or how to compute the weighted average of peer-company characteristics. Furthermore, they cannot separate out the value of data, strategic resources, or other intangibles by using existing comparable-firm approaches. This is a major limitation, given that firm-related big data are available, and intangibles are increasingly important. This paper therefore fills an important gap in the literature and practice by proposing an innovative method, the **CFFA** (**Comparable Firm-Finding Algorithm**).

[2] "Intuition" is explained in the Applications and Variations section below.

NUMERICAL EXAMPLE

Consider three firms: *A*, *B*, and *C*. As A and B are publicly traded, their share prices are observable. C is a private firm (possibly a startup), and consequently its share price is not publicly observable. Assume the goal of a venture capitalist (VC) is to evaluate C's "fair" share price. The VC regards {A, B} as comparable to C and aims to use information about {A, B} to evaluate C.

Now assume that the VC can observe and verify two kinds of information {growth, profitability} for {A, B, C}. Growth information can include asset growth, sales growth, and earnings growth; profitability information can include EV/EBITDA, return on asset (ROA), and return on equity (ROE). Let us assume that the growth variables for {A, B, C} are {1%, 2%, 2%} and the profitability variables are {2%, 6%, 5%}. Table 1 presents a summary.

******* Insert Table 1 about here *******

Next, we combine {A, B} to synthesize C by matching the attributes, i.e., growth and profitability. To denote the positions (weights) of {A, B} as $\{a, b\}$, the synthetic-C can be constructed by solving the following system of equations:

$$\textit{Growth equation}: 1\% \cdot a + 2\% \cdot b = 2\%.$$

$$\textit{Profitability equation}: 2\% \cdot a + 6\% \cdot b = 5\%.$$

With two unknowns and two independent equations, we can solve the system as follows: $\{a, b\}$ = {1, 0.5}. Thus, the synthetic-C that matches the attributes (growth and profitability) of C is the {1, 0.5} *weighted average* of {A, B}. To rephrase, the synthetic-C is the portfolio of {A, B} with the weights {1, 0.5}. It is therefore straightforward to estimate C's market value. If {A, B}'s P/S ratio (price/sales ratio) is {2, 6}, synthetic-C's P/S is 2 x 1 + 6 x 0.5 = 5. If C's sales are \$10m, then C's estimated market value is \$10m x 5 = \$50m. For a compact discussion, this logic can be expressed using matrix algebra, as follows:

$$\textit{Attributes of comparable firms}: X = \begin{bmatrix} 1\% & 2\% \\ 2\% & 6\% \end{bmatrix}.$$

$$\textit{Attributes of a target firm}: y = [2\%, 5\%]^T.$$

$$\textit{Weights of comparable firms}: w = [a, b]^T.$$

The system of equations and its solutions are:

$$Xw = y \iff w = X^{-1}y.$$

This example assumes an "exact" identification because the number of comparable firms and attributes is the same (2, in this case). What happens if the number of attributes is larger than the number of comparable firms? For example, what happens if we have data, in addition to growth and profitability (e.g. big data)? How should an analyst proceed if she is interested in a nonlinear construction of synthetic-C?

When the number of attributes is larger than the number of comparable firms, the system becomes "over-identified." Assume that the number of comparable firms and the number of attributes are T and N, respectively ($T < N$). Then, X is an N x T matrix; y is an N-dimensional vector; and w is a T-dimensional vector. To find w (the weights of comparable firms), we add N-dimensional measurement error terms and express the system of equations as:

$$y = Xw + \epsilon.$$

It thus becomes straightforward to "estimate" w. This can be done using regression approaches (e.g. ordinary least squares, weighted least squares, generalized method of moments). Practically, weighted least squares (WLS) are useful and simple because an analyst can add her own conviction or confidence about the value of information contained in an attribute. If the analyst believes that all attributes are equally important, she can construct a synthetic-C using an ordinary least squares (OLS) regression "without a constant," as follows:

$$\textit{Weights for synthetic target firm}: w = (X'X)^{-1}X'y.$$

******* Insert Table 2 about here *******

This problem ($N > T$) is illustrated using the numerical example in Table 2. Suppose the VC adds a new attribute, growth*profitability, to account for nonlinear matching. The values of the new attribute, growth*profitability, are then {2bp, 12bp, 10bp} for A, B and C, respectively.

The regression approach produces the weights of comparable firms as {0.996, 0.502}, which happen to be close to {1, .5}, the previous solution. If an "intercept term" is included in the regression, the coefficients are:

Intercept: 2pb,

Coefficients: [96%, 51%].

There are many ways to interpret the intercept. For instance, it can be the value of intangible strategic resources that a firm owns (the source of a sustainable competitive advantage) because it cannot be replicated with a combination of comparable firms (Barney, 1986). However, analysts with other views may interpret it as a measure of overvaluation. Whether the intercept term signals outperformance or overvaluation is ultimately an empirical question. We discuss this interesting question in a later section.

While interaction terms and other nonlinear transformations (e.g., log[*growth*]) can be included, data availability in this area of big data is not a source of concern. In the above example, the CFFA uses a firm's cross-sectional information. Since cross-sectional "alternative data" have become widely available (e.g., text, audio, visual and other alternative data), various data-collection methods (e.g., web scraping) can be used to fill entries in the CFFA tables. This suggests that the CFFA can overcome the limitations of standard methods, which use only a small number of variables. Hence, the CFFA approach to valuing a firm is better adapted to the age of big data and machine learning. A later section discusses how the CFFA specifically enriches traditional methods, such as comparable company analysis (CCA) in this regard.

GENERALIZATION

This section generalizes the numerical examples and develops the CFFA in detail. The CFFA is a regression-based approach, which constructs a synthetic version of a target firm in order to find the unobservable “fair” value of the target (e.g. stock price, bond price, risk, and volatility). If the value of the target is observable, the CFFA can compare the “fair” synthetic with observed values to develop recommendations for the target firm’s managers, investors, and stakeholders. The section below explains how the CFFA generally constructs a synthetic firm for a target firm. The following notations are employed:

N: # of attributes.

T: # of firms (*T* < *N).*

$\vec{y}$: *N dimension column vector of the attributes of a target firm.*

$\vec{x}_i$: *N dimension column vector of the attributes of a comparable firm i.*

x: *N* x *T dimension matrix of the attributes of comparable firms.*

$x \equiv (\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T)$.

N is the number of attributes per firm. *T* is the number of comparable firms. The attributes can include textual information and sentiment, as well as financial factors, such as firm size, price-to-book ratio, and ROA. Since *N* > *T* is to be ensured to run the regression, the attributes can be modified/combined or a machine-learning algorithm can be run to manufacture additional attributes, as follows:

e.g. $f_1, f_2, \ldots, f_1 f_2, f_1 f_2 f_3, log(f_1), tanh(f_3), \ldots$

In this way, one can increase N to infinity. To illustrate a machine-learning algorithm, one can use an autoencoder (Hinton & Salakhutdinov, 2006) to generate attributes. However, the manner in which attributes are generated and selected is more of an art than a science. An analyst's intuition may be the most valuable tool for generating attributes. If the use of intuition is impractical or needs validation, one can use an autoencoder, a method of unsupervised learning or denoising datasets that uses artificial neural networks.

Let x be an N x T sized matrix by binding the attributes of comparable firms by columns. Then, the goal becomes:

How to combine $x = (\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T)$ *in order to match* $\vec{y}$*?*

The problem is expressed using a regression approach as follows:

$$\vec{y} = x\vec{\beta} + \vec{\epsilon}.$$

$$\vec{\beta} : \textit{ (portfolio) weight to construct a synthetic comparable firm.}$$

$$x\vec{\beta} : \textit{ constructed attributes of the synthetic firm.}$$

$$\vec{\epsilon} : \textit{ errors between the attributes of the target and the synthetic firm.}$$

To find a synthetic version of a target firm, one can use OLS, GLS, GMM, lasso regression, and similar approaches. For instance, a GLS-solution for the portfolio weights of comparable firms to match the attributes of the target firm is:

$$\beta_{GLS} = (x'\Omega^{-1}x)^{-1}x'\Omega^{-1}\vec{y}.$$

Next, assuming estimating the stock-return volatility or *pe* (price-to-earnings) ratio of a target is the areas of interest; one can use:

$$\sigma^2_{target} = (\sigma^2_1, \sigma^2_2, \cdots, \sigma^2_T) \cdot \beta_{GLS}.$$

$$pe_{target} = (pe_1, pe_2, \cdots, pe_T) \cdot \beta_{GLS}.$$

An analyst who is only interested in a weighted average can impose further restrictions, such as:

$$Neither\ positive\ or\ negative\ leverage : \vec{1}^T \cdot \vec{\beta} = 1.$$

$$No\ short\text{-}selling : \vec{\beta} \geq 0.$$

These equations demonstrate that the CFFA can estimate unobservable financial information associated with a target and generalize the CCA, the most popular relative-valuation framework, which the next section will discuss in more detail.

GENERALIZED COMPARABLE COMPANY ANALYSIS

The CFFA approach can be seen as a generalization of the popular comparable company analysis (CCA) approach for practitioners, a subset of relative valuation models. The process of implementing CCA is as follows (Damodaran, 2007, 2012). The first step is to identify companies that are similar to the company in question (the "target company"). The identified similar companies are called *the peer group*. The second step is to collect financial data about the peers and choose *one* financial ratio for all; this ratio is usually a standardized price. The third step is to calculate the average or median of the peer group's financial ratios. The fourth step is to apply the average or median to the target company in order to compute the target's standardized value, and therefore a fair price (de-standardizing).

The CFFA generalizes all four steps in the CCA. First, the CFFA quantifies the similarity between companies. If a company is dissimilar to the target, it is assigned little weight during the construction of a synthetic target (i.e., regression coefficient). Therefore, the CFFA not only quantifies how to compare a company to the target, but also clarifies how to assemble a strong peer group.

Second, any financial and nonfinancial big data can be collected and used in the valuation —beyond standardized prices in the CCA. I believe that this innovation represents a major contribution to the management literature, and one that will assist practitioners. A standard CCA collects and generally uses only one financial ratio. In the CFFA, by contrast, the more data there are, the more statistical power because the CFFA uses regression analyses. One can even extend the CFFA to use machine-learning analyses beyond regression. This is possible because the CFFA can incorporate large amounts of data. In particular, the CFFA can explicitly and easily consider strategic resources and business-model characteristics as long as they can be encoded in the dependent and independent variables. This offers a solution to a problem that has challenged practical analysts for some time: *how to evaluate intangible assets, given the rise of asset-light intangible-intensive business models* ("Intangible Assets Are Changing Investment," The Economist, 2017).

Third, while the CCA uses the median or a simple average of peer group metrics, the CFFA uses regression coefficients and potentially nonlinear functions. Table 3 presents the details of a nonlinear transformation.

******* Insert Table 3 about here *******

From the perspective of a regression, while Table 2 illustrates how to increase the number of observations (*N*) to generalize the CCA, Table 3 explains how to increase the number of variables (*K*), comparable to more traditional interaction terms (e.g., the moderation effect). Suppose the regression coefficients of {A, B, A*B} are {0.3, 0.5, 0.1} and the price-to-book (*pb*) ratios of {A, B} are {2, 3}. Then, the target's estimated *pb* ratio is:

$$0.3 \cdot pb_A + 0.5 \cdot pb_B + 0.1 \cdot pb_A \cdot pb_B$$

$$= 0.3 * 2 + 0.5 * 3 + 0.1 * 6 = 2.7.$$

If the target's book value is \$100M, its estimated market value is \$270M. This can be expressed as an equation:

$$P_{target} = B_{target}[0.3 \cdot pb_A + (0.5 + 0.1 pb_A) \cdot pb_B].$$

$$P_{target} : \textit{the target's market value},$$

$$B_{target} : \textit{the target's book value}.$$

Hence, the weight to A can be regarded as fixed to 0.3 while the weight to B is dynamically varying as $(0.5 + 0.1A)$. In other words, Firm A *moderates the valuation relationship* between Firm B and the target. A more complex function can be expressed as $Target = f(A,B)$. As far as a function, f, is analytic, it admits Taylor series expansion and allows the same interpretation.

Fourth, while the CCA assumes an efficient market (i.e., a market that values peers accurately), the CFFA has less need for this assumption for the following reasons: (1) The CFFA can average out more peers and their characteristics in the regression approach. It is therefore less prone to peer-related single measurement errors (e.g., mispricing). (2) The CFFA can modify the regression approach to directly accommodate mispricing in the market. In fact, an important insight in the resource-based view is the incomplete strategic-factor market, which logically implies mispricing and different opinions (Barney, 1986). This problem can be solved by including a *constant term* in the regression approach. The next section details this solution. (3) The CFFA can modify "input variables" when some peer characteristics are unreliable because it uses multidimensional information about the peer instead of the CCA's univariate approach. The next subsection discusses this idea in more detail.

APPLICATIONS AND VARIATIONS

The CFFA is flexible enough to admit wide applications. The next subsections illustrate how to use the CFFA for strategic management, intangible valuations, and investment strategies.

Regression with a constant to quantify the value of strategic resources

By definition, it is difficult to price a firm's strategic resources due to "different expectations about the future value of a strategic resource" (Barney, 1986). However, this does not mean that scholars cannot price the value of strategic resources. For instance, the fair price of a stock is unobservable, but practical stock analysts work hard to assess whether a stock is under- or over-priced. Different opinions about a fair price should be averaged out; the average tends to be equal to the observed market price. A similar, albeit more difficult, problem is to quantify the value of strategic resources. In particular, I argue that stock analysts in financial institutions have ignored the significance of pricing strategic resources and that the managers interested in formulating strategies need to make more concerted efforts to price the resources and dynamic capabilities of their own firms and those of their competitors. The CFFA assists the challenges. In this subsection, I modify the CFFA's regression-based approach and propose a new method to quantify the value of strategic resources.

Previously, this paper has proposed using the regression approach without a constant term. The restriction that *a constant term equals zero* makes sense because the aim is to construct a synthetic version of the target firm that permits the calculation of its unobservable "fair" value. The synthetic version is essentially a portfolio of firms with characteristics comparable to those of the target firm. The constant term is dropped because it is unrelated to a position, weight, or the target's peers. However, what happens when the zero-constant restriction in the regression approach is relaxed?

If a constant term is significant in constructing a synthetic target using the CFFA, it captures factors that competitive financial markets do not capture. This intuition corresponds to the incompleteness noted in the strategy literature (Barney, 1986) as a characteristic of strategic-factor markets. If the constant term is positive, the firm in question will enjoy a positive premium that observable factors cannot explain away. If the term is negative, the market will underprice factors associated with the firm. It is difficult to generalize why a constant term can become positive or negative, but a positive value must relate to valuable, rare, inimitable, and non-substitutable (VRIN) firm-related factors. If a factor is not valuable, the constant term cannot be positive; if it is not rare, the constant term must be insignificant; otherwise, other value-maximizing firms would acquire the factor and so peer characteristics would absorb the factor. If the factor is not inimitable or non-substitutable, other peers will develop it quickly, absorbing away its statistical significance.

Note that, by construction, a constant term is the component of a firm's value that the combination of comparable firms cannot replicate. Thus, the constant in the regression approach is a "stock" variable, not a "flow" variable. At this point, the argument becomes consistent with Dierickx and Cool (1989), who argue that a firm chooses the paths of flow variables to develop its stock of strategic assets; the extent to which that the stock is substitutable or imitable determines its value, as well as eventually the firm's competitive advantage.

Although this paper highlights intuition in discussing the constant term, similar ideas can be used to explain the sum of coefficients greater than one, negative coefficients, or R-squared.

Omitted-value approach to assessing the value of intangibles

Standard investment models have limitations in the age of intangibles, including unobservable strategic resources and capabilities (Gu & Lev, 2017). This subsection illustrates the process of calculating the contribution that intangibles make to a firm's value. Table 4 illustrates this idea, which I refer to as the "omitted-value approach."

******* Insert Table 4 about here *******

Panel A contains hypothetical initial data. Panel B omits Target C's intangible information from Panel A by setting it at zero. Next, solving systems of equations or using the regression approach for both Panel A and Panel B, we produce the following solutions to generate the synthetics of Target C *with or without intangibles*.

$$\text{Case A} : C_{intangibles} = 0.1429A + 0.7857B.$$

$$\text{Case B} : C_{no-intangibles} = -0.7143A + 1.0714B.$$

Suppose that {A, B}'s P/S ratio (price/sales ratio) is {2, 3}; C's P/S ratio is unknown, while C's sales are $10m. Then, the estimated P/S ratios and the value (P) of Target C are as follows:

$$C_{intangibles} : [P/S, P] = [2.6428, \$26.428].$$

$$C_{no-intangibles} : [P/S, P] = [1.7857, \$17.857].$$

These solutions imply that intangibles increase C's value by 0.8571x in terms of the P/S ratio or by $8.571m in dollars. In addition, one can define the "price-to-intangible ratio" as 26.428/8.571 = 3.1x, the definition of which is:

$$\textit{Price-to-intangible ratio}$$

$$= \frac{Firm\ value\ computed\ with\ the\ original\ input\ data}{Intangibles'\ value\ on\ the\ omitted\text{-}value\ approach}.$$

The idea is simple. To compute the contribution of specific resources, one must simply construct modified data by changing the resources' input values to some other values (e.g., zero or epsilon changes). Next, one must run the regression approaches for the original and modified data and compare the outputs between the data.

This "omitted-value approach" can be generalized to compute the value of resource combination or combinative capabilities (Kogut & Zander, 1992). For example, a target firm's input vector can be changed as follows:

$$Original\ input : [1.1, 6.9, 9.8, 3.9, 5.1]$$
$$\longrightarrow New\ input : [1.1, 0, 9.8, 0, 5.1].$$

This example sets {6.9, 3.9} to {0,0} in order to find the value of the combinative capabilities of the second and fourth elements. Instead of zeros, the variables can be changed into other numbers. Furthermore, the regression-based approach with a constant term makes it possible to determine how the value of strategic resources is likely to change by comparing the constant terms before and after omitting or modifying input values.

Consider data valuation as an example. On average, a firm's data increased 8.3x, from 1.45 petabytes in 2016 to 13.5 petabytes in 2019. Data have become a firm's most important intangible assets. Firms' data can be categorized into operational, manufacturing and business data (Dell

Technologies, 2020). Operational data include financial information, human resources, and other firm activities. Manufacturing data include work hours, outputs, and work progress. Business data include customer activities and purchased data. Business data are valuable for assessing the extent of a firm's data-driven business model. Manufacturing data can enhance cost efficiencies.

When the CFFA is applied to data valuation, company data can be quantified and categorized into operational, manufacturing, and business sets. Then, by omitting or epsilon-changing the categories one by one, one can ascertain the marginal contribution of a dataset to the firm's value (by running regression without a constant) or to its strategic factors (by running regression with a constant).

The omitted-value approach also makes it possible to simulate combinative capacities. Some combinations of input variables can be epsilon-increased to simulate new values of a firm and its resource (constant term). This clarifies the impact of the combinations and sets the stage for further scenario planning.

Finally, this idea can be modified to ascertain the ESG (environmental, social, and governance) value. By setting E-, S-, or G-related variables = 0 or not, one can compute the marginal contribution of these variables to a firm's value or valuation ratios, such as *price-to-ESG* ratios. This will clarify the extent to which ESG explains a firm's value.

Investment strategies

It is straightforward to apply the CFFA to design investment strategies. Although developing investment strategies are not the primary focus of this paper, the following ideas may inspire financial economists on the process of applying the ideas developed by strategy scholars in investment and trading. Some ideas are as follows.

First, the CFFA allows for a comparison of the value of a target firm with the value of its synthetic counterpart, which is in turn the weighted average of the target's comparable firms (i.e., its portfolio). If both the target and its synthetic versions are tradable, pair-trading strategies can be implemented. In other words, as the valuation gap between a target and its synthetic varies, traders can take long or short positions on the pair.

Second, the first idea can be extended to include portfolios, via the regression-based approach with a constant term. Firms with positive and negative constant terms are collected by running systems of regression equations. Then two portfolios are formed: one group with positive and the other with negative regression constants. The final step is to analyze the spread between the two portfolios to identify how macro-financial variables Granger-cause the spread. For example, when the economic environment is volatile enough to render a firm's dynamic capabilities more valuable, the spread between the portfolios will widen. If a sector bubble is serious, the spread between the portfolios formed within the sector will be narrow. Therefore, by combining the intuition of the resource-based view with the time-series pattern of the spread, one can trade the positive- and negative-constant portfolios to generate "alphas" (excess returns).

Third, the CFFA can be used to check the value of every firm using other firms. In fact, every firm can be synthesized with every other firm (a *complement* set). In other words, the entire COMPUSTAT database can be used to valuate a firm. This big-data approach suggests an algorithmic way to carry out a valuation and identify under- or over-valued stocks. This will facilitate the development of artificial-intelligence or robo-advisors to save portfolio-management fees while generating superior returns.

Other issues

Endogeneity:

The CFFA regression equation is endogenous. Since the CFFA is uninterested in causality, it can ignore inverse-causality issues. However, simultaneity may be a concern because the observable information of the target and its peers can be co-determined. Note that the independent variable is a target firm's characteristics, and the dependent variables are those of peer firms. In this context, codetermination is the reason why stock analysts use the CCA, a special case of the CFFA. The following example illustrates this simultaneity issue.

Consider Coupang, Inc. (CPNG), listed on the NYSE on 2021/03/11. Since Coupang is regarded as the Amazon of Korea, Amazon is an obvious peer. Indeed, many analysts use the CCA (e.g., price-to-sales ratio) because they believe that the valuation ratios of Coupang and Amazon co-move with correlated unobservable factors. Therefore, endogeneity is the reason that stock analysts use the CCA to evaluate Coupang on Amazon. Since the CFFA generalizes the CCA, they share the same issue. This paper suggests the following resolutions.

First, use econometric techniques. The CFFA has an advantage over the CCA because the former can, but the latter cannot control for endogeneity. For instance, find firm-specific variables (instruments) for each peer firm. The instruments are correlated with the peer firm's characteristics, but uncorrelated with the target's characteristics. Next, regress the characteristics of each peer firm on the instruments (first-stage regression). Following this, use the predicted characteristics as input variables when applying the CFFA (second-stage regression). If it is possible to obtain firm-specific variables, many other econometrics techniques can be deployed to resolve simultaneity. As the resource-based view assumes the presence of heterogeneous strategic resources, strategic resources are natural instruments. By definition, they affect peer values, but do not affect target values, due to the VRIN[3] characteristics of the resources.

Second, disregard endogeneity. The CFFA is not about causality, but about interpolation. Suppose a target's synthetic is 50%*Firm A + 50%*Firm B on the CFFA. If {A, B}'s price-to sales ratios (P/S) are {3,7}, the target's P/S is likely to be around the interval between 3 and 7. This is an interpolation, not a causality. The CFFA interpolates the target's P/S as 5. In sum, endogeneity may not be a relevant issue because the CFFA uses the regression approach for interpolation only.

Historical analysis

Historical analysis is popular among stock analysts. For example, when analyzing a firm, a stock analyst generates a band of price-to-sales ratios, with the target firm's historical P/S data. Whether the target's current P/S is above, within, or below the band signals the extent of possible mispricing. The CFFA can incorporate such time-series or panel-data analyses easily. The example in Table 5 describes a CFFA-based historical analysis.

******* Insert Table 5 about here *******

---

[3] Valuable, rare, inimitable, and non-substitutable.

The estimated OLS equation is as follows:

$$C_{2021} = .9187 + .0014C_{2020} + .5843C_{2019} - .1391A_{2021} + e.$$

This result suggests that the target firm in 2021 enjoys a positive premium (positive strategic resource) in comparison to Peer A and its historical values in 2020 and 2019. If the P/S of the peer firms are all one, the fair P/S in 2021 is 2.5466. If C's sales in 2021 are $100M, C's fair-market value becomes $254.66M. I believe that this approach is much more analytic and intuitive than the popular practice of eyeballing a historical band of a valuation ratio.

CONCLUSION

Standard valuation models become less effective as intangible assets become more important. In fact, it is natural to face such challenges because existing methods have ignored the most important assets: intangible resources and capabilities that determine sustainable competitive advantage which have been highlighted by strategy scholars for long. However, while the limitations of the standard models are apparent, there has, to date, been no practical or scientific alternative. The present study therefore proposes a new valuation method, based on strategic-management intuition. I call this framework the Comparable Firm Finding Algorithm (CFFA).

The CFFA includes the Comparable Company Analysis (CCA) as a special case. Since the CCA is one of the most popular and practical valuation approaches, practitioners can easily adopt the CFFA as an extension. Furthermore, since the CFFA is based on regression analysis, researchers can easily extend their statistical skills to include machine learning or big data. This paper explains how to apply the CFFA by computing the value of strategic resources, identifying ESG contributions, assessing intangibles (price-to-intangible ratios), evaluating combinative capabilities, pricing entrepreneurial activities, and developing investment strategies and scenario planning.

Future research can extend the CFFA in many ways; although it is far more general than conventional models (e.g., CCA), the CFFA remains simple and flexible enough to permit wide applications and variations. First, researchers can combine accounting, financial and alternative data to measure strategic resources and intangibles. They can then identify macro- or micro-factors that influence changes in the value of resources. Second, organization researchers can use the CFFA to define performances and design incentive schemes across teams. Third, finance and accounting researchers can find under- or overpriced assets, and identify the causes of mispricing, while also generating trading strategies. Fourth, entrepreneurship researchers can complement existing entrepreneurial finance models by evaluating innovative asset-light projects on the CFFA.

The most significant limitation of the CFFA relates to that of the CCA. In comparison to the discounted cash flow (DCF) approach, the CCA is less interested in modeling a target's business models and more interested in the *mood* of the market. While the CFFA is more general than the CCA, uses bigger data, and does not need to assume that the market values a target's peers accurately, the approach does not directly model a target's business lines, as the DCF does. The DCF and NPV (net present value) are just different approaches. Future research can suggest ways to overcome the limitations of the CFFA and discover the best processes and routines in order to apply the CFFA to practical problems.

REFERENCES

Barney, J. B. (1986). Strategic factor markets: Expectations, luck, and business strategy. *Management Science, 32*(10), 1231–1241.

Barney, J. B. (2001). Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view. *Journal of Management, 27*(6), 643–650.

Barney, J. B., & Arikan, A. M. (2001). The resource-based view: Origins and implications. *The Blackwell Handbook of Strategic Management*, *5*, 124–188.

Barney, J. B., Wright, M., & Ketchen Jr, D. J. (2001). The resource-based view of the firm: Ten years after 1991. *Journal of Management*, *27*(6), 625–641.

Brynjolfsson, E., & Kahin, B. (2002). *Understanding the Digital Economy: Data, Tools, and Research*. MIT Press.

Damodaran, A. (2007). *Valuation approaches and metrics: A survey of the theory and evidence*. Now Publishers Inc.

Damodaran, A. (2012). *Investment Valuation: Tools and Techniques for Determining the Value of any Asset, University Edition*.

Dell Technologies. (2020, March). *Global Data Protection Index Survey*. https://www.delltechnologies.com/en-us/data-protection/gdpi/index.htm

Dierickx, I., & Cool, K. (1989). Asset Stock Accumulation and Sustainability of Competitive Advantage. *Management Science*, *35*(12), 1504–1511. JSTOR.

Gu, F., & Lev, B. (2017). Time to change your investment model. *Financial Analysts Journal*, *73*(4), 23–33.

Haskel, J., & Westlake, S. (2017). *Capitalism without Capital: The Rise of the Intangible Economy*.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507.

Intangible assets are changing investment. (2017, December 19). *The Economist*. https://www.economist.com/finance-and-economics/2017/12/19/intangible-assets-are-changing-investment

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, *3*(3), 383–397.

Mahoney, J. T., & Pandian, J. R. (1992). The resource-based view within the conversation of strategic management. *Strategic Management Journal*, *13*(5), 363–380.

Peteraf, M. A. (1993). The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal*, *14*(3), 179–191.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, *5*(2), 171–180.

Wernerfelt, B. (1995). The resource-based view of the firm: Ten years after. *Strategic Management Journal*, *16*(3), 171–174.

Why book value has lost its meaning. (2019, March 23). *The Economist*. https://www.economist.com/finance-and-economics/2019/03/23/why-book-value-has-lost-its-meaning

Table 1: Simple 2*2 numerical example for the CFFA
Let us assume that the growth variables for {A, B, C} are {1%, 2%, 2%} and the profitability variables are {2%, 6%, 5%}. {A, B}'s P/S ratio (price/sales ratio) is {2, 6}. C's P/S ratio is unknown while C's sales are $10m.

	A	B	C
Growth	1%	2%	2%
Profitability	2%	6%	5%

Table 2: How to increase the observations for firms in the CFFA

	A	B	C
growth	1%	2%	2%
profitability	2%	6%	5%
growth*profitability	2bp	12bp	10bp
...	...	...	...

Table 3: Nonlinear combination of comparable firms

	A	B	A*B	Target
Metric 01	1	2	2	3
Metric 02	2	5	10	7
Metric 03	3	12	36	13
...	...	...	...	...

Table 4: Valuation of intangibles using the CFFA
{A, B}'s P/S ratio (price/sales ratio) is {2, 3}. C's P/S ratio is unknown while C's sales are $10m.

Case A: Original data

	A	B	C
Intangibles	3%	2%	**2%**
Tangibles	2%	6%	5%

Case B: Data with the assumption that intangible = 0

	A	B	C
Intangibles	3%	2%	**0%**
Tangibles	2%	6%	5%

Table 5. Historical analysis (example)

The numbers in this table are simulated. The estimated OLS equation on the numbers is:

$$C_{2021} = .9187 + .0014C_{2020} + .5843C_{2019} - .1391A_{2021} + e.$$

{C_2021, C_2020, C_2019} are a firm C's characteristics in years {2021, 2020, 2019}. A_2021 is a Firm A's characteristics in year 2021.

ID	C_2021	C_2020	C_2019	A_2021
0	1.131227	0.382517	0.730461	0.794913
1	1.737502	0.802960	0.960281	0.357688
2	1.263956	0.001317	0.314716	0.109220
3	0.696964	0.927830	0.019635	0.750659
4	0.892390	0.457671	0.245259	0.560900
5	1.380565	0.614483	0.562517	0.845723
6	1.314547	0.341762	0.671936	0.415405
7	0.904892	0.398936	0.414260	0.631550
8	0.994300	0.461407	0.743484	0.998976
9	0.813707	0.041301	0.194462	0.487134
10	1.194239	0.398927	0.023123	0.724078
11	1.172419	0.987667	0.339988	0.249792
12	0.661431	0.920075	0.226551	0.448459
13	0.961606	0.086409	0.140788	0.865058
14	1.628945	0.617133	0.942514	0.726267
15	1.553247	0.342863	0.472014	0.641150
16	0.752172	0.609401	0.196094	0.385380
17	0.994184	0.373936	0.982315	0.501991
18	1.024734	0.020714	0.704673	0.671257
19	1.143079	0.348327	0.545741	0.924392