Hierarchical and Segmented Approaches to Startup Valuation: What They Are. Why They Work

Max Berre

1 Introduction

Why do startups in some cities attract higher valuations than those elsewhere in the country, can be observed in California vis-à-vis New York or Boston? In Europe, similar things are true both within and among European markets. Not only do valuation differentials exist between Paris and Lyon, for example, but also between Paris and London. This is the case even when the startups in question are based in similarly-sized economies, share the same industries and many of the same investors?

Although classical economic theory describes that valuations are based on revenues, growth-rates, and risk-adjusted discount rates, the valuation of startups often proves the exception to the rule. Fundamentally, due to their short histories, difficult-to estimate intangible assets, and opaque details, startups are notoriously difficult to value, a phenomenon described in detail by Damodaran (2009). Over the past 30 years, scholars have been attempting to formalize both valuation and valuation-drivers within startup markets.

Whereas overall published knowledge is both sparse and dispersed across several academic fields, Bellavitis et al. (2017) and Budhwar et al. (2022) agree on the importance of focusing on startup valuations as a key avenue of research, with potential to tie together, financial, entrepreneurial, and macroeconomic microeconomic theoretical perspectives, thereby forging dynamic and innovative insights.

Traditionally, startup valuation has relied on classical data-driven approaches. These include the discounted cashflow (DCF) valuation approach, which takes

This article has been sponsored by the Partners of the "Finance for innovation" Chair at Audencia Business School. Special thanks is accorded to Benjamin Le Pendeven.

M. Berre (✉)
Université de Lyon, Lyon, France
e-mail: maxens.berre@univ-lyon3.fr; mberre@audencia.com

S. Derindere Köseoğlu (ed.), *A Practical Guide for Startup Valuation*, Contributions to Finance and Accounting, https://doi.org/10.1007/978-3-031-35291-1_5

several forms, which discount different measures of firm income, revenue, or cashflow, using the appropriate risk-adjusted discount-rate. Approaches include the free-cashflow-to-firm (FCFF) approach, which discounts sales revenues, discounting them with the weighted average cost of capital, as well as the free-cashflow-to-equity (FCFE) approach, which discounts net income by the cost of equity as per the capital asset pricing model (CAPM), and the Gordon Growth Model, which mitigates discount-rates in accordance with revenue growth-rates. A prominent alternate valuation approach in startup markets is the relative-valuation approach, which is described in detail by Damodaran (2010). Mechanically, this approach relies on estimating valuation as a multiple of the firm's balance sheet components, income statement components, or cashflow statement components. Valuation multiples, in turn, are driven by averages drawn from comparable firms, which typically share the same contextual market conditions, such as industry, business model, and economic geography. Empirically speaking, regression-based approaches supplement these models by controlling for contextual market-conditions (Berre & Le Pendeven, 2022). These range from macroeconomic conditions to industry-level market-dynamics, to relevant entrepreneur characteristics. Lastly, more niche-case approaches such as the real options valuation approach depend on the accurate and detailed estimation of risk and volatility.

In contrast to the classical valuation approaches which dominate peer-review empirical finance literature, industry practitioners often use summation-based segmented models to estimate valuation in a piecemeal fashion (Ernst & Young., 2020). While this approach is widespread due to its straightforward architectural simplicity, its use is often confined to specific industry-sectors or economic geography, rather than being applied in a more general fashion. Mechanically, this approach relies on attaching estimate values to a wide range of valuation factors, which are ultimately aggregated to produce the final valuation estimate. Overall, economic theory holds that these diverse approaches to valuation are equivalent to one another (Fama, 1970; 1991; Damodaran, 2002).

Fundamentally, this is intended to constitute a detailed, in-depth how-to guide describing methods and approaches for the application of segmented hierarchical startup-valuation, as well as how they can be applied using existing data and regression-models. The rest of the study proceeds as follows: The subsequent section describes segmented startup-valuation models, describing their emergence and use in both practitioner-focused grey-literature, as well as in peer-review literature. Following this, section three describes how segmented models can be made hierarchical, as well as explaining how this modelling-approach can be used for microtargeting-based valuation approaches. Lastly, a discussion and conclusion section describes why segmented, hierarchical, and microtargeting valuation approaches are used by industry practitioners, by describing their added-value vis-à-vis more traditional valuation-approaches.

2 Segmented Models: How Contextual Factors Play a Role

A useful theoretical approach used by a minority of scholars is that of the scorecard-based approach. A critical advantage of scorecard valuation-approaches is the ability to incorporate qualitative, geographic, sectoral or other types of categorical valuation-factors in several ways ranging from the non-financial and deal-characteristics prevalent in a given municipal or industry-specific sectoral ecosystem, to the role of national-level or macroeconomic and macrofinancial market-conditions. This approach is capable of shedding light into valuation even as detailed related economic and financial information is missing, scare, or unevenly available.

Segmented valuation-methods are modular and relatively straightforward valuation-approaches based on summation of key valuation-determinants, firm-characteristics, market-conditions, and deal-conditions developed mainly by industry practitioners. One principle advantage of this type of approach is that valuation can be modelled, captured, and contextualized via the inclusion of specific categorical information, which could be general, highly-specific, and/or be organized as joint, combined, or hierarchical segmentation.

2.1 Practitioners: Segmented Models in Markets

In industry, scorecard approaches are typically employed by business angels. Industry-emergent techniques for scorecard valuation include Berkus (2016) and Payne (2011). Perhaps the most well-established segmented startup valuation model is the Scorecard Model, outlined by Payne (2011). Outlined in Table 1, the scorecard model segments the impact of valuation factors into management team, target market, competitive environment, and need for further funding. Valuation is established via summation of the model's component factors:

Alternatively, another well-known alternative to Payne's Scorecard model can be found in the Berkus Model (EY, 2020). Outlined in Fig. 1, the Berkus model segments valuation into component risks. Valuation is established via summation.

2.2 Segmented Models in Peer Review

While the segmented valuation-model approach has made considerable traction among industry practitioners, within published economic literature, this same concept appears in the form of summation-based valuation models, such as the models outlined in Hand (2005) and Sievers et al. (2013). Concretely, Eq. (1) for example, outlines the Hand (2005) startup-valuation model, describing the model's component deterministic valuation-factors as being segmented into financial-statement data

Table 1 Abbreviated Payne Scorecard Model

Weighting		Impact on Startup Valuation
0–30%	*Impact*	*Merits of the entrepreneur and management team*
	+	Several years of overall business experience
	++	Experience in the industry in question
	+++	Experience as a CEO
	++	Experience as a COO, CFO, CTO
	+	Experience as a product manager
	–	Experience in technology or sales
	– –	No business notable experience
		Size of the opportunity
0–25%	*Impact*	*Scale of target market (measurable in total sales)*
	– –	<$50 million
	+	$100 million
	++	>$100 million impact
		5-year potential for revenues of target company
	– –	<$20 million
	++	$20 to $50 million – >$100 million (will require substantial additional funds)
0–15%	**Impact**	**Strength of the product and intellectual property**
	– – –	Not well defined, still seeking or developing a prototype
	0	Well defined, prototype looks interesting
	++	Good feedback from potential customers
	+++	Orders or early sales from customers
0–10%	**Impact**	**Competitive environment**
		Strength of competitors in this marketplace
	– –	Dominated by a single large player
	–	Dominated by several players
	++	Fractured, many small players
	Impact	**Strength of competing products landscape**
	– –	Competing products are excellent
	++	Competing products are weak
0–10%	**Impact**	**Marketing/sales/partners**
		Impact sales channels, sales and marketing partners
	– – –	Haven't even discussed sales channels
	++	Key beta testers have been identified and contacted
	+++	Channels secure, customers placed trial orders
	– –	Firm has not identified partners
	++	Key partners in place
0–5%		**Need for additional rounds of funding**
	+++	None
	0	Another angel round needed.
	– –	Need venture capital

Source: Ernst and Young (2020)

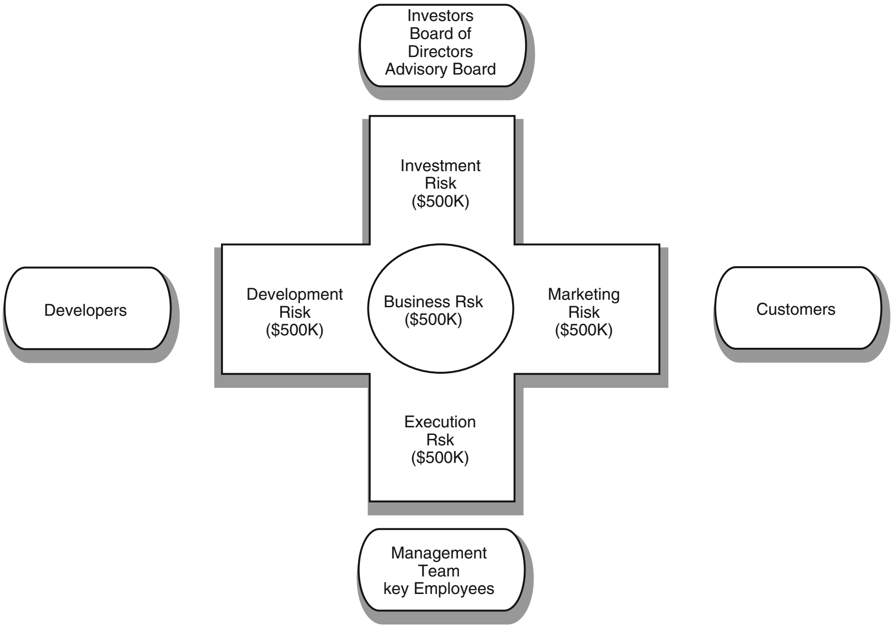


Fig. 1 Berkus model for startup valuation. Source: Berkus (2016)

such as Net Income, Cashflows, and Assets on one hand, and operational and industry-related data on the other.

Equation 1: Hand (2005) Summation-based Segmented Valuation Model

$$\begin{aligned} &\text{HAND (2005)} \;\; Ln(\text{Pre} - \text{Money Valuation}) \\ &= \sum \Theta_b Ln(\text{Financial Statement Data}_{bik}) \\ &+ \sum \Upsilon_c Ln(\text{NonFinancial Statement information}_{cik}) + \varepsilon_{ik} \end{aligned} \tag{1}$$

Meanwhile, Eq. (2), outlines another prominent segmented startup-valuation model developed by Sievers et al. (2013) as a summation-based valuation model, assigning valuation based on summation of financial, and non-financial firm-attributes, as well as deal-characteristics along with their relevant valuation-coefficients. Essentially, whereas Hand (2005) segments valuation-factors into accounting and non-accounting data, with each segment contributing to valuation with its own coefficient, Sievers et al. (2013)'s model uses similar model-architecture to segment valuation-factors into financial factors such as risks and revenues drawn from a firm's income statement and balance sheet, and assets and capital-invested drawn from balance-sheet data, as well as, non-financial factors such as industry-level data and firm-level operational data, and deal characteristics such as investor-syndication, and shareholder-agreement clauses such as tag-along, redemption, and ratchet clauses in the venture capital investment deal.

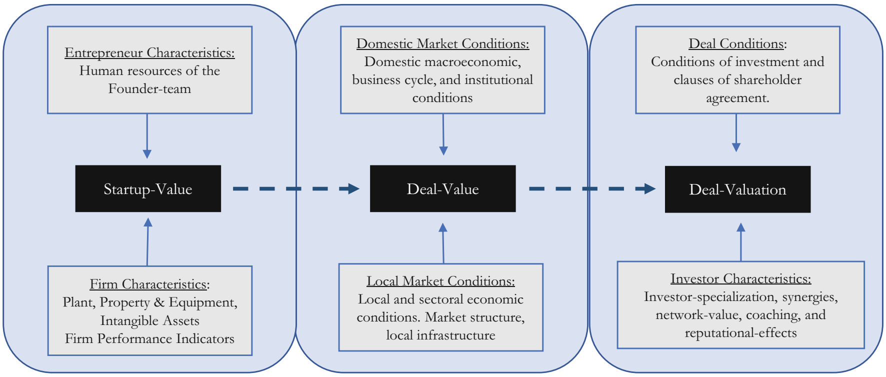
Equation 2: Sievers et al. (2013) Summation-based Segmented Valuation Model

$$\log(\text{Valuation}_{it}) = \sum \Phi\text{Non} - \text{financial}_{it} + \sum \Delta\text{Financial}_{it} + \sum \Psi\text{Deal Characteristics}_{it} \quad (2)$$

As with the Payne and Berkus valuation models, the startup's valuation is established via summation of the established segments. What these model-models have a tendency to overlook, however, are interactions and hierarchies among the identified valuation-determinants. Architecturally speaking, a closely-related alternate functional form for segmented valuation-models can be elaborated as multistage valuation approaches, such as the Startup-Valuation Meta-Model described by Berre and Le Pendeven (2022) outlined in Eq. (3). This would account for phases, hierarchies and interactions among valuation-determinants. Formally this can be expressed as:

Equation 3: Berre and Le Pendeven (2022) Startup Valuation Meta-Model[1]

$$\text{Pre} - \text{Money Valuation} = f\left(\left(\left(\sum \text{Start} - \text{Up Value}\right) \sum \text{Deal Value}\right) \sum \text{Deal Valuation}\right) \quad (3)$$



[1] Source: Berre and Le Pendeven (2022).

3 From Segmentation to Microtargeting: The Hierarchical Modelling Approach

The recent emergence of ever-developing machine learning techniques has led to increasing methodological sophistication of scorecard approaches, as predictive techniques incorporating to categorical, geo-spatial, and qualitative data become widespread.

Mechanically, microtargeting by means of data mining is described by Murray and Scime (2010), as the process of inductively analyzing data to find patterns, fault-lines and relationships among the data, on the basis of trends related to both descriptive and numerical characteristics, such as average age, number of family members, and geographic area, via construction of decision trees. Essentially, this is an analytical technique which is both explanatory and predictive, and is useful for both variable predictions, as well as to provide key insight regarding structure, segmentation, and interrelationships among data.

This approach provides insight into how specifically the outcome variable's value is dependent on the model's deterministic factors, with each identifiable fault-line constituting a segment of individual observations. Data-mining-driven microtargeting, for instance, allows scorecard-based valuation-approaches to incorporate categorical and qualitative data to a potentially-extreme degree of detail, given the added explanatory power of variable-hierarchy for accurately modelling relationships among explanatory variables.

Functionally speaking, a hierarchically-structured valuation-model that would result from a microtargeting approach can be expressed via a staged valuation approach, such as the Startup-Valuation Meta-Model described in Eq. (2). Figure 2

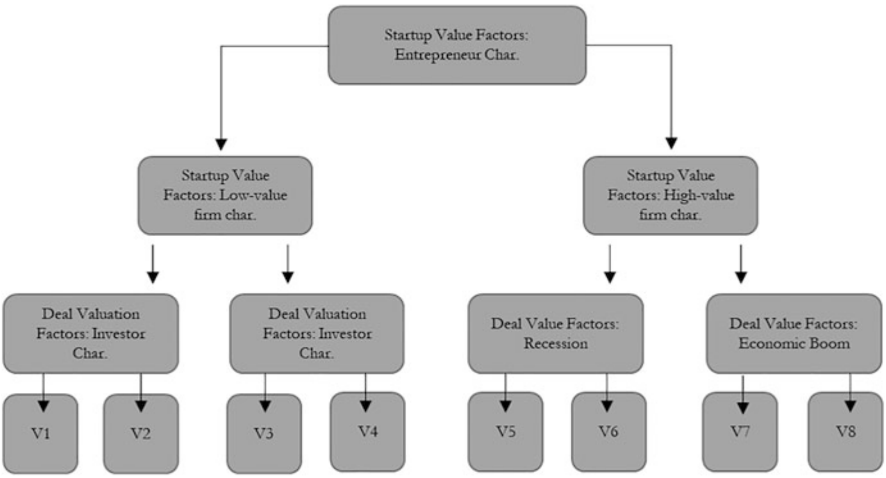


Fig. 2 Decision tree based on the Berre-Le Pendeven Meta-Model. Source: Author's own creation

displays the form that the Berre-Le Pendeven Meta-Model would adopt as a hierarchical decision tree.

3.1 Why Hierarchical Approaches Work: Regression Trees and Random Forests

In principle, CART-based microtargeting using regression-trees and random forests, which aggregate multiple regression-trees can reorganize valuation-determinant data such that several key insights emerge, which regression-model approaches might otherwise overlook, as well as by more rudimentary valuation-models. First, data consisting of qualitative and categorical valuation-factors such as sectoral geographic, and business-model details, which carry the potential information-density are taken into account. Second, CART trees demonstrate areas and subsections of the data where given valuation-determinants might be more or less-influential, granting very precise insight into how valuation emerges. Third, fundamental fault-lines are displayed as values along which branches diverge.

For the purposes of startup valuation, the informational content of descriptive and categorical characteristics such as geographic location, industry-sector, and business model are often overlooked, despite the general possibility that these characteristics might possess explanatory power equivalent to multiple associated numerical variables. Meanwhile, incorporation of descriptive categorical-characteristics into econometric models via use of fixed-effects suffers losses in explanatory-power as the number of descriptive-characteristics increases, whereas decision-tree-based microtargeting approaches see improved accuracy as the number of categorical and qualitative characteristics increases. Consequently, a principal advantage of this approach is that it is viable to microtarget valuation by including ever small-scale and highly-specific categorical information.

3.2 Functional Form of Segmented Valuation Models

Krzywinski and Altman (2017) explain that CART approaches do not develop a prediction equation per se. Instead, data are partitioned along the predictor axes into subsets with homogeneous values of the dependent variable. This process is represented by a decision tree that can be used to make predictions from new observations. Accordingly, several functional-form options exist mathematically, which can be used both in markets and in research settings. Furthermore, the combination and/or selective use of these can be a useful way to investigate valuation in detail, as this may serve to maximize nuances.

3.2.1 Log Transformation

Given that log transformation makes multiplication and summation interchangeable, the log transformation of regression-variables can architecturally simplify regression models and mathematical relationships for purposes of empirical specification (Neter, 1990; Wooldridge, 2010; Benoit, 2011), while also lending themselves to model-flexibility. Since log transformation brings the product property of logarithms to bear, it is possible to represent any model in its entirety in the form of a summation-model for intermediate-stage purposes, given the interchangeability of logarithm summation and multiplication (Miller et al., 2010). Strictly-speaking, this means that intermediate-stage functional forms can be functionally reoriented both in terms of variable-order and in terms of interaction-effects.

Furthermore, log transformation serves to "flatten" relationships, by restraining outlier impact on dataset means and medians. Given that regression trees and partitioning methods outputs in general can be sensitive to the influence of dependent-variable outliers (Khan et al., 2013), outlier-flattening has potential to add substantial explanatory power to regression-tree models, as log-transformation reduces estimation-issues associated with percentage changes from baseline (Keene, 1995), while maximizing data-scale flattening (Ribeiro-Oliveira et al., 2018). Additionally, variables showing skewed distribution can be made symmetric using log transformation (Keene, 1995).

On the other hand, given that log transformation also impacts multiplicative models and functional-forms (Benoit, 2011), the specific architectural shape of the valuation-function becomes unclear, as multiplication, summation, ratios, and other functional form elements inherent in the valuation-function might also become unclear.

In order to reach a viable comprehensive valuation-outlook, it is necessary to examine the model's log-transformed expression alongside its original version, whose functional-form would capture both variable-order and possible interaction terms in detail. In order to establish a decision-tree model however, both variable-order and relative variable-importance need to be established. Fundamentally, regression variable-interaction terms can convey how specifically a model's explanatory variables interact with one another. This serves to indicate variable-position within the model's decision tree, granting more complete and wholistic insights on the details of relationship's causality structure.

3.2.2 Regression-Model Equations

Functionally speaking, regression-model equations consist of a summation of key-variables, modified by factor-coefficients, alongside constant and error-terms. Structurally, this model-architecture lends itself to near-direct transposition of segmented valuation approaches, as well as the approximation of most classically-established firm-valuation models, ranging from the discounted cashflow valuation (DCF) approach, to a multiples-valuation approach.

Since regression-model equations are structured as summation functions, with each term consisting of a variable and a coefficient, valuations can ultimately be expressed as a summation of variables, coefficients, constants, the error-term. For instance, a discounted-revenue-based valuation model approach, incorporating similar information to a discounted cashflow valuation (DCF), can approximate an FCFE approach by regressing startup-valuation on historic and current Net Income figures, thereby capturing both the free cashflow and its growth rate, as well as risk factors which would drive the discount rate, which can be expressed as a combination of the risk-free-rate and the applicable equity risk-premium described in the CAPM model. This is captured in Eq. (4).

Equation 4: Regression model simulating free cashflow to equity

$$\begin{aligned} \text{Valuation}_{it} = \ & \alpha_i + \beta_1(\text{Net Income}_{it}) + \beta_2\ (\text{Net Income}_{it-n}) \\ & + \beta_3(\text{Risk} - \text{Free rate}_t) + \beta_4(\text{Risk} - \text{Premium}_{it}) + u_{it} \end{aligned} \tag{4}$$

On the other hand, multiples-valuation approaches, whose widespread popularity flows, in part, from its simplicity and ease with which these models communicate valuation, as well as its ability to convey the market's current mood (Damodaran, 2002), might seek to estimate valuation from as few as one valuation-determinant factor drawn from a firm's income-statement, balance-sheet, or cashflow-statement. This however, comes at the cost of sample-selection, as choosing the sample of relative firms and assets against which to compare, can lead to standardization (or assumption of standardization) of variables outside the valuation-model. According to Damodaran (2002), the most widespread multiples-valuation model is the price/sales ratio, describing valuation as a function of a firm's sales revenues, as demonstrated in Eq. (5):

Equation 5: Price-to-Sales Ratio

$$\text{Price} - \text{to} - \text{Sales} = \frac{\text{Firm's Total Market Share} - \text{Price}}{\text{Sales Revnue}} \tag{5}$$

Equation (6) expresses the valuation-impact of the Price-to-Sales ratio as an Ordinary-Least-Squares (OLS) regression-model, given by the parameter Sales Revenue, while β estimates the Price-to-Sales ratio. Outside factors ranging from quantitative valuation-factors such as total assets, borrowing costs, or CAPEX, to qualitative valuation factors such as factors driven by sector, industry-specialization, or economic geography can be sample-selected to be constant, or assumed to be constant across the sample.

Equation 6: Price-to-Sales Ratio as an OLS regression model

$$\text{Valuation}_i = \ \alpha_c + \beta_c(\text{Sales Revnue}_i) + u_i \tag{6}$$

Beyond the use of regression-model functional-form to convey or approximate classical firm-valuation models such as DCF or relative-valuation, the OLS regression-model's functional form can also be used for summation-based segmented valuation-models, such as those outlined in Eqs. (1) and (2). Moreover, this is even the case for models using hierarchical approaches, such as Mahmoud et al. (2022) express random forest regressions using OLS-style regression-model equations, simulating the summation-based segmented functional form used by OLS models.

3.2.3 Decision Tree Functional-Forms

Architecturally speaking, there is flexibility regarding the function forms that decision tree model could adopt considering the possible contexts in which they can be deployed, the factors enumerated, and both their relative and hierarchical explanatory power. While Krzywinski and Altman (2017) describe that the CART approach does not express a prediction equation (i.e., that this approach is backwards-looking), CART regression tree results can be used to extend and modify segmented models. Fundamentally, the regression tree model's outputs make possible two architecturally-viable segmentation approaches.

For example, Mahmoud et al. (2022) express random forest regression models using OLS-style regression-model equations, simulating the summation-based functional-form of an OLS model. This modelling-approach has the advantage of capturing the overall directionality of the causal relationship to be tested empirically, without specifically precluding existence of complex model functional-forms.

Comparative Model Explanatory Power and Goodness-of-Fit

In general, the accuracy of regression tree models can be compared to those of equivalently-constructed regression models on the basis of their goodness-of-fit indicators. Whereas explanatory power of OLS and panel-data regression models are evaluated on the basis of R^2, Sandeep (2014) and Firmin (2021) outline that regression trees are to be evaluated on the basis of $1 - R^2$ root-mean-squared-error.

Weighted Summation Segmentation

First, a rudimentary "back-of-the-envelope" segmentation-approach can essentially be considered a modification of Payne's Scorecard Model, which includes model-weighting to its segmentation approach. In order to obtain regression-tree model-weights from the CART approach, it would suffice to examine variable-importance. While CART variable-importance outputs can aggregate to a maximum of 100%, as demonstrated by Table 2, aggregate variable importance model-outputs might also add to less than 100%. While for CART models whose aggregate variable

Table 2 OLS Model Using DCF Valuation-Factors and Business Model

OLS Coefficients:	Estimate	Std. Error	T-Value	P-Value
(intercept)	5.10E+08	6.12E+07	8.326	5.02E-16***
Revenue	4.27E-01	6.20E-02	6.892	1.31E-11***
Country-risk-premium	−4.19E+09	3.18E+09	−1.317	0.188
Sectoral-Beta	−4.61E+08	6.02E+07	−7.664	6.67E-14***
B2B & C	3.06E+08	6.13E+07	4.995	7.59E-07***
B2B	9.80E+07	6.25E+07	1.569	0.117
B2C	6.37E+08	6.28E+07	10.138	<2.00E-16***
*** p < 0.01, ** p < 0.05, *p < 0.1				
Residual standard error:	546,900,000 on 644 degrees of freedom			
Multiple R-squared: 0.279		Adjusted R-squared: 0.273		
F-statistic: 41.6 on 6 and 644 DF, p-value: <2.20E-16				

Source: Berre (2022)

importance adds to 100%, it suffices to assign the model's variable-importance figures as valuation-model factor-coefficients, for instances in which variable-importance outputs aggregate to less than 100%, factor-importance proportionalities would need to be calculated as an initial step, as outlined in Eq. (7):

Equation 7: CART Variable-Importance Proportionality

$$\text{Factor} - \text{Coefficient}_i = \sigma_i(X)_i = \frac{\text{Variable Importance}_i}{\sum_n^i \text{Variable Importance}_i} \tag{7}$$

Fundamentally, this approach is highly useful as a generally applicable model-approach, giving rise to a Payne-style scorecard valuation model, which can be applied in a general fashion to startup markets as a whole. For example, a Payne-style scorecard valuation-model, involving valuation-weights, which could be constructed on the basis of firm characteristics and market characteristics, can take the form outlined in Eq. (8), combining the FCFE valuation-factors with Payne valuation-factors outlined in Table 1:

Equation 8: Weighted Summation Segmentation Regression-Tree Valuation Model Simulating FCFE Valuation Model

$$\begin{aligned} \text{Valuation}_i = \; & \sigma_1\beta_1(\text{Net Income}_i) + \sigma_2\beta_2(\text{Risk} - \text{Free rate}_i) \\ & + \sigma_3\beta_3(\text{Risk} - \text{Prem.}_i) + \sigma_4\beta_4(\text{Size of Opportunity}_i) \\ & + \sigma_5\beta_5(\text{Competitive Environ.}_i) + \sigma_6\beta_6(IP_i) \end{aligned} \tag{8}$$

Where:

$\sum_{i=1}^{n} \sigma_i = 1$ but where $\sum_{i=1}^{n} \widehat{\sigma_i} \leq 1$.

Here, σ expresses the weighting-coefficient n of startup i (e.g., the scale of Net Income's impact on startup i's valuation), driven by the factor's variable-importance drawn from the CART output, while β expresses the impact-coefficient n of startup i (e.g., country-level sovereign risk-premium is a valuation-determinant known to be a constituent of DCF-model discount-rates (Damodaran, 2009), and as such, can be expected to have negative valuation-impact and therefore a negative β-coefficient).

Mechanically, this functional-form approach can work for either continuous valuation-determinants drawn from firm-level financial statements (i.e., Net Income, Fixed Assets, etc.) and from market indicators (i.e., business-cycle and macroeconomic indicators), or for binary factors such as intellectual-property or entrepreneur-characteristics. Moreover, because CART regressions segregate data into dichotomous subsets along the predictor axes, categorical variables (i.e., classifications such as sectoral-industry classifications and business-model classifications, as well as variables linked to economic geography such as cities, counties, inclusions in regional-clusters) which are treated as binary-variables.

Hierarchical Ordinal Segmentation

A second approach could be called the hierarchical ordinal segmentation approach. Given that the data are partitioned along predictor axes into subsets with homogeneous dependent-variable values, a more complex hierarchical modelling-approach is also possible. The basis of this model-approach begins with adoption of terminal-node average-values as ω-coefficients. These can be multiplied by the regression-tree's branch-thresholds and branch-conditions, as follows:

$$\omega_i(X)_j \left(\begin{cases} =1 & \text{if } X \text{ is true} \\ =0 & \text{if } X \text{ is false} \end{cases}\right.$$

Or

$$\omega(X)_j \left(\begin{cases} =1 & \text{if } X \text{ is above threshold} \\ =0 & \text{if } X \text{ is below threshold} \end{cases}\right.$$

Thereafter regression-tree models can be elaborated for specific given startups, following any given startup's position within the regression tree. Eq. (9) describes this model functional-form.

Equation 9: Valuation Regression-Tree Model Using Hierarchical Ordinal Segmentation

$$\text{Valuation}_i = \omega_i \left(\prod_{i1}^{in} \text{Branch Threshold}_{ii} \right) + \ldots$$
$$+ \omega_n \left(\prod_{n1}^{nn} \text{Branch Threshold}_{nn} \right) \tag{9}$$

As a specific example building on Eq. (9), establishing a specific startup valuation-model, Eq. (10) applies the hierarchical ordinal segmentation approach to Eq. (8)'s combined FCFE-market-conditions valuation-model, while ranking the nodes in hierarchical-order following their order in Eq. (8). Note that this causes the factor-order described in the equation to change somewhat to reflect the conditionality-relationship.

Equation 10: Valuation Regression-Tree Using Hierarchical Ordinal Segmentation Model Approach

$$\text{Valuation}_i = \omega_i \left(\prod_{i}^{I} \text{Net Income}_{ii} \right) + \omega_j \left(\prod_{j}^{J} Risk - \text{Free rate}_{jj} \right)$$
$$+ \omega_k \left(\prod_{k}^{K} \text{Risk} - \text{Premium}_{kk} \right) + \omega_l \left(\prod_{l}^{L} \text{Size of Opportunity}_{ll} \right)$$
$$+ \omega_m \left(\prod_{m}^{M} \text{Competitive Env.}_{mm} \right) + \omega_n \left(\prod_{n}^{N} IP_{nn} \right) \tag{10}$$

A fundamental difference between the hierarchical ordinal approach and a weighted-summation approach is that the hierarchical ordinal model-approach is specific to the individual startup's position within the decision tree. Essentially, this means that the segmentation's functional-form differs from that of weighted-summation approach, since a startup's regression-tree branch-placement may indicate functional form featuring either an omission or a repetition of some of the regression model's valuation-determinants, a feature which may be functionally-indicative of either conditional valuation-impacts or variable interaction-effects.

Another core difference between the two model approaches, is that while the weighted-summation approach can grant a holistic view of σ-weights across the dataset as a whole, the ordinal-model approach can directly provide a valuation-estimate by placing the firm along regression-tree's terminal-nodes (i.e., the regression-tree's leaf-nodes).

3.2.4 Two-Tiered Approach

Given that the inclusion of categorical variables is able to grant key insights on valuable information, of both qualitative and quantitative nature, and holds the explanatory-power potential to be as information-dense as the joint-inclusion of multiple numerical variables, their use for research purposes remains a very valuable tool. This is in particular the case with fixed-effects regressions, given that they can meaningfully incorporate categorical indicators such as geographical or industry-level designations. In the face of multiple information-dense categorical variables however, this approach is subject to a hard-limit, taking into consideration that the explanatory power of joint-fixed-effects can be limited as the number of categorical variables grows.

What this means therefore is that either OLS or fixed-effects regressions can be deployed in order to capture the general causal-overview among the valuation-drivers and in order to detect information-density and explanatory-power of relevant categorical labels. In order to elaborate on any OLS or fixed-effects findings, CART (or possibly-other cluster-driven approach) can be utilized.

With this in mind, combined empirical approaches are possible, with the potential to outperform single-method analysis in terms of detailed insights in two important ways. First, this approach can outperform an OLS-based summation model in terms of model-accuracy, model sophistication, and explanatory power, because it can grant insights on the roles, relative-position, and hierarchy of near-significant explanatory-factors. Second, the two-tiered approach can provide detailed insight vis-à-vis scale and sign of factor-impacts (i.e., β-coefficients), thereby improving upon pure CART-based weighted-summations.

4 Example of CART-Based Microtargeting with One Categorical Variable

Tables 2 and 3 demonstrate both OLS and CART approaches to examine valuation-regression-models, which include revenues, and discount-factor components consisting of country-risk-premium (conveying country-level risk-free-rate), and sector-level CAPM-beta (conveying sector-level risk-premium) as discounted-cashflow valuation-factors alongside business model.

In principle, one can expect firm revenues to have positive β-coefficients, given their positive valuation-impact, while the DCF-discount-factor components (country-risk-premium and sector-level CAPM-beta) can both be expected to have negative coefficients. Meanwhile, business model is a categorical variable, which may take the value "business-to-business" (B2B), "business-to-customer" (B2C), business-to-business-and-customers" (B2B & C), or business-to-government" (B2G).

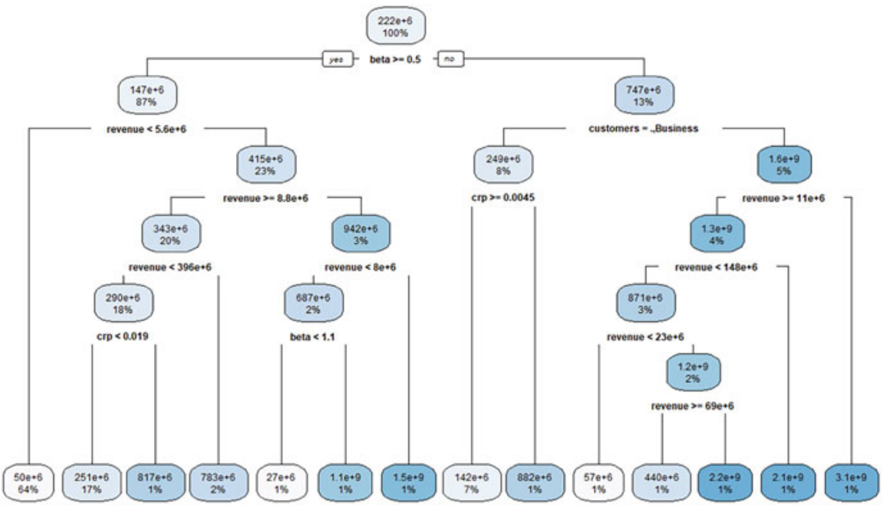
First, Table 2 uses an OLS model to examine the relationship between DCF-factors, business model, and startup-valuations, splitting business-model into dummy-variables, finding that the valuation-impact of revenue is DCF-consistent, while the discount-factor appears to be driven by sector-level CAPM-beta, and the valuation-impact of B2B is outweighed by both B2C and B2B & C.

Table 3 DCF Valuation-Factors and Business Model CART

Observations:	1048			
End nodes:	15			
Complexity parameter	No. of Split	RMSE	Cross-validation error	Cross-validation St. dev.
0.1280	0	1.0000	1.0024	0.1634
0.0623	2	0.7441	0.8255	0.1484
0.0574	3	0.6817	0.8100	0.1479
0.0376	4	0.6243	0.7245	0.1398
0.0285	5	0.5867	0.7133	0.1397
0.0241	7	0.5296	0.7016	0.1385
0.0148	8	0.5055	0.6458	0.1366
0.0132	9	0.4906	0.6200	0.1317
0.0132	11	0.4643	0.6219	0.1318
0.0102	12	0.4512	0.6242	0.1318
0.0100	13	0.4409	0.6154	0.1318
Variable importance				
Revenue	Business model	Beta	Country-risk premium	
35	24	23	18	

Source: Berre (2022)

Meanwhile, Table 3 outlines a decision-tree-based CART valuation which includes revenue, country-risk-premium (capturing country-level risk-free-rate), and sector-level CAPM-beta (capturing sector-level risk-premium) as discounted-cashflow valuation-factors alongside business model, and describes premoney startup-valuations ranging from €27 Million to €3.1 Billion, and are partitioned hierarchically.



Given the architectural shape of Table 3's regression-tree, the weighted-summation approach and the hierarchical-ordinal approach would lead to somewhat-different functional-forms. Eq. (11) demonstrates a weighted-summation functional-form expression of the valuation-model resulting from the regression-tree outlined in Table 3, taking the resulting variable-importance indicators as σ-coefficients.

Equation 11: Valuation Regression Tree Model Using Weighted-Summation Segmentation

$$\begin{aligned}\text{Valuation}_i = \;& 0.35\beta_1(\text{Revenue}_i) + 0.24\beta_2(\text{Business Model}_i)\\ &+ 0.23\beta_3(\text{Sectoral} - \text{Risk Beta}_i) + 0.18\beta_4(\text{Country} - \text{Risk Premium}_i)\end{aligned} \tag{11}$$

As per this approach, the highest-valuation tranche would first and foremost be startups with substantial revenue figures. This is followed by firms which have business models focusing on B2C, B2B & C, or B2G commerce, and whose revenues are discounted by low sector-level CAPM-betas, as well as by low country-risk premiums. Essentially, this means that the highest-valuation EU startups are firms combining substantial revenue figures with a B2C, a B2G, or a B2B & C, business model, and which are located in a low-volatility industry, and based in a AAA-rated home-market such as Denmark, Germany, or Switzerland (Damodaran, 2021), whereas lowest-valuation EU startups are more likely to be based in higher-risk EU markets (for example in the CEE or Euro-Med region), and are characterized by low-revenues, high-risk industry-sectors, and a B2B business model. Table 3 presents the regression-tree results outlined in Table 2, as a Payne-Style valuation-scorecard.

By also drawing on the OLS findings outlined in Table 2 as a source of β-coefficients, a two-tiered approach is possible. Here, Eq. (12) and Table 5 capture the revisions possible by inclusion of β-coefficients drawn from Table 2. Because Business Model has been re-transcribed as its constituent (statistically-significant) dummy variables, B2C and B2B & C, the valuation-model's functional-form includes terms and coefficients for both of these business-models, but excluding B2B and B2G.

Equation 12: Valuation Regression Tree Model Using Weighted-Summation Segmentation

$$\begin{aligned}\text{Valuation}_i = \;& (0.35 * 0.4273)(\text{Revenue}_i) + (0.24 * 637{,}000{,}000_{B2C})\\ &\times (\text{Business Model}_i) + (0.24 * 305{,}900{,}000_{B2B\&C})(\text{Business Model}_i)\\ &+ 0.23\beta_3(-460{,}900{,}00_i) + 0.18\beta_4(._i)\end{aligned} \tag{12}$$

Building on this revision, Table 5 represents a revision of the Payne-style summation scorecard outlined in Table 4, featuring the incorporation of β-coefficients drawn from use of a two-tiered valuation-approach.

Alternatively, hierarchical ordinal segmentation, a second segmentation modelling-approach, gives rise to a significantly more extensive valuation-model

Table 4 CART-based Valuation as Weighted-Summation Segmentation Results Presented in Payne-Style Scorecard

Weighting	Sign of β Coef.	Impact on startup valuation
35%	**Impact**	*Revenue*
	+	Valuation is positively impacted by revenues
		Business model
24%	**Impact**	*Client focus of the business*
	–	Business-to-business (B2B)
	+	Business-to-customer (B2C)
	+	Business-to-business and customer (B2B & C)
	+	Business-to-government (B2G)
		Discount factor
23%	**Impact**	*Sector-level CAPM-beta*
	–	Valuation negatively impacted by sectoral risk
18%	**Impact**	*Country-risk premium*
	–	Valuation is negatively by country-risk-premium
Total		
100%		

Source: Author's own creation

functional-form, as each of the regression-tree's branch and terminal-nodes can be represented in the model. Equation (13) demonstrates an example of this second valuation-segmentation approach, outlined in Eq. (8). Because the CART results include 14 terminal-nodes, as well as numerous branch-nodes, the complexity and size of the entire long-form valuation equation is substantial.

Equation 13: Valuation Regression Tree Hierarchical Ordinal Segmentation Model Approach

$$\text{Valuation}_i =$$

$$50{,}000{,}000(Sectoral - Beta \geq 0.5) * (Revenue_i < 5{,}600{,}000)$$

$$+251{,}000{,}000(Sectoral - Beta \geq 0.5) * (Revenue_i \geq 5{,}600{,}000) * (Revenue_i \geq 8{,}800{,}000) * (Revenue_i < 369{,}000{,}000) * (Country - Risk - Premium_i < .019)$$

$$+817{,}000{,}000(Sectoral - Beta \geq 0.5) * (Revenue_i \geq 5{,}600{,}000) * (Revenue_i \geq 8{,}800{,}000) * (Revenue_i < 369{,}000{,}000) * (Country - Risk - Premium_i \leq .019)$$

$$+783{,}000{,}000(Sectoral - Beta \geq 0.5) * (Revenue_i \geq 5{,}600{,}000) * (Revenue_i \geq 8{,}800{,}000) * (Revenue_i \geq 369{,}000{,}000)$$

Table 5 Two-Tiered Revised-Valuation as Weighted-Summation Results Expressed as a Payne-Style Scorecard

Weighting	Sign of β Coef.	Impact on Startup Valuation
35%	**Impact**	*Revenue*
	0.427	Valuation is positively impacted by revenue. Per EUR of revenue.
		Business model
24%	**Impact**	*Client focus of the business*
	.	Business-to-business (B2B) —— (not significant)
	637,000,000	Business-to-customer (B2C)
	305,900,000	Business-to-business and customer (B2B & C)
	.	Business-to-government (B2G) —— (not significant)
		Discount factor
23%	**Impact**	*Sector-level CAPM-beta*
	– 460,900,000	Valuation is negatively impacted by sectoral risk. Per 1.00 of CAPM-Beta
18%	**Impact**	*Country-risk premium*
	–	Valuation is negatively impacted by country-risk-premium. But not statistically significant using a European EU/EEA dataset. Near-significance of coefficient indicates that CRP is likely to be significant in more diverse datasets.
Total		
100%		

$$+27{,}000{,}000(Sectoral-Beta \geq 0.5) * (Revenue_i \geq 5{,}600{,}000) * (Revenue_i < 8{,}800{,}000) * (Revenue_i < 8{,}000{,}000) * (Sectoral-Beta < 1.1)$$

$$+1{,}100{,}000{,}000(Sectoral-Beta \geq 0.5) * (Revenue_i \geq 5{,}600{,}000) * (Revenue_i < 8{,}800{,}000) * (Revenue_i < 8{,}000{,}000) * (Sectoral-Beta \geq 1.1)$$

$$+1{,}500{,}000{,}000(Sectoral-Beta \geq 0.5) * (Revenue_i \geq 5{,}600{,}000) * (Revenue_i < 8{,}800{,}000) * (Revenue_i \geq 8{,}000{,}000)$$

$$+142{,}000{,}000(Sectoral-Beta < 0.5) * (Business\ Model_i = B2B) * (Country-Risk-Premium_i \geq .0045)$$

$$+882{,}000{,}000(Sectoral-Beta < 0.5) * (Business\ Model_i = B2B) * (Country-Risk-Premium_i < .0045)$$

$$+57{,}000{,}000(Sectoral-Beta < 0.5) * (Business\ Model_i = B2C\ or\ B2B\&C\ or\ B2G) * (Revenue_i \geq 11{,}000{,}000) * (Revenue_i < 148{,}000{,}000) * (Revenue_i < 23{,}000{,}000)$$

$$+440{,}000{,}000(Sectoral - Beta < 0.5) \\ * (Business\ Model_i = B2C\ or\ B2B\&C\ or\ B2G) * (Revenue_i \geq 11{,}000{,}000) \\ * (Revenue_i < 148{,}000{,}000) * (Revenue_i \geq 23{,}000{,}000) \\ * (Revenue_i \geq 69{,}000{,}000)$$

$$+2{,}200{,}000{,}000(Sectoral - Beta < 0.5) \\ * (Business\ Model_i = B2C\ or\ B2B\&C\ or\ B2G) * (Revenue_i \geq 11{,}000{,}000) \\ * (Revenue_i < 148{,}000{,}000) * (Revenue_i \geq 23{,}000{,}000) \\ * (Revenue_i < 69{,}000{,}000)$$

$$+2{,}100{,}000{,}000(Sectoral - Beta < 0.5) \\ * (Business\ Model_i = B2C\ or\ B2B\&C\ or\ B2G) * (Revenue_i \geq 11{,}000{,}000) \\ * (Revenue_i \geq 148{,}000{,}000)$$

$$+3{,}100{,}000{,}000(Sectoral - Beta < 0.5) \\ * (Business\ Model_i = B2C\ or\ B2B\&C\ or\ B2G) \\ * (Revenue_i < 11{,}000{,}000) \tag{13}$$

An interesting detail about the regression-tree described in Table 3 is that several of the nodes indicate unicorn valuation. Stated otherwise, this decision tree appears to describe the recipe for the establishment of unicorn-valuations. Furthermore, we see that revenue drives the majority of the lower and intermediate branches, corroborating revenue's dominant-position in terms of variable-importance.

Nevertheless, while the entire regression-tree valuation-function outlined in Eq. (13) is sizable and cumbersome, it is not necessary to estimate the function as a whole. Rather, because segments of the function where the criteria are not met are zero, it suffices to estimate the branches and terminal-node where the firm actually finds itself. For example, for a startup located in the rightmost terminal-node, whose sectoral beta would be larger than 0.5, and whose revenue is less than €50,000,000, Eq. (14) reduces to:

Equation 14: Valuation Regression Tree Model Reduced-form Ordinal Segmentation Model Approach

$$\text{Valuation}_i = \ 50{,}000{,}000(\text{Sectoral} - \text{Beta} \geq 0.5) * (\text{Revenue}_i < 5{,}600{,}000) \tag{14}$$

While this reduced-form is both compact and immediately-useful for practitioner purposes, substantial detail is lost in terms of other-path branches and terminal nodes, as well as their distributions and threshold-values.

5 Discussion, Conclusion, and Further Research

Overall, segmented valuation-models are historically underappreciated within empirical finance literature, with segmented models surfacing in but a small, obscure fraction of startup-valuation literature (Berre & Le Pendeven, 2022). Nevertheless, appearance of these models in practitioner and industry-sourced grey literature (e.g., Ewing Marion Kauffman Foundation (2007), Goldman (2008), Payne (2011), Berkus (2016), and Ernst and Young. (2020)), can be taken as indication that segmentation valuation approaches have established traction among industry practitioners ranging from venture capital investors and business angels to consultancy and auditing practitioners.

5.1 *Why Do Segmented Models Work?*

While these segmented valuation-models may be presently under-represented within the literature, the ongoing emergence and proliferation of machine learning techniques can be expected to increase the viability, diversity and popularity of segmented models within the literature, given that there are several empirical approaches drawn from both econometrics and machine-learning empirical approaches, to which segmented models can be adapted. In principle, the industry-popularity and usefulness in markets of segmented valuation-models can be attributed to several noteworthy positive qualities which characterize them.

First, segmented models are mechanically and mathematically straightforward, making them easy to intuit and understand, as well as easy to communicate to investors, clients, and stakeholders. This characteristic quality may partially explain widespread popularity of the Berkus and Payne methods among industry practitioners and among industry-sources. Indeed, Damodaran (2002) ascribes this quality to models using this approach.

Second, segmented models can be estimated quickly. Because of their mechanical simplicity, rough valuation-estimations can be executed quickly, in the field, and perhaps even with only partial information available. This detail contrasts more complex valuation approaches, which might require substantial access or estimation to key figures.

Third, segmented models are directly transposable to empirical modelling, making the investigation of their validity and accuracy relatively straightforward. Fundamentally, this is the case because both CART and OLS models can be expressed in segmented functional-form.

Fourth, segmented models have substantial flexibility. Because the segmented valuation-models' functional-form are readily-transposable for the purposes of empirical modelling, they are also highly-adaptable. This means that they can be altered by adding or modifying the impacts of valuation-determinant factors as the need arises, for example by adding segments to capture interaction terms or niche functional-form segments. Furthermore, they can be constructed by modifying other styles of valuation-models. For example, relative-valuation models can be combined into two-factor or three-factor segmented valuation-models.

The rise and proliferation of hierarchical empirical approaches, including not only CART-based regression-trees, but also related tree-based empirical approaches, such as the bottom-up Hierarchical Ascending Classification decision-trees, and Random Forest has yielded the proliferation of increasingly-accurate and flexible prediction-models, which can not only be used for valuation purposes, but also for speedy decision-making, as well as the construction of increasingly-flexible segmented valuation models. This indicates that the use of such approaches across business, market, and investment landscapes can only be expected to proliferate in the future.

5.2 Contributions and Further Research

Because this study focuses on the implementation of methodological approaches imported and drawn from industry practitioners, as well as from marketing and political science journals, within entrepreneurial finance literature, this study adds to the existing body of research in several ways by both addressing existing theory gaps, and by elaborating on currently-existing published empirical findings.

First, this study links practitioner-approaches with trends in peer-review literature. While practitioner-derived or industry-oriented sources such as Ewing Marion Kauffman Foundation (2007) or Ernst and Young (2020) point to segmented valuation-models such as valuation-approaches described by Payne (2011) and Berkus (2016), this approach, seen in studies such as Hand (2005) or Sievers et al. (2013) for valuation models and Siskos and Zopounidis (1987) for selection-models, has heretofore received relatively-little attention within peer-review literature. Principally, this is owed to overall need for model-sophistication in order to incorporate interaction-effects and variable-hierarchies within valuation models. This study provides an overview and synthesis of these approaches, which can be generally deployed by practitioners and valuation-experts across a wide variety of markets, while also providing context, as well as developmental-direction for the ongoing debate within peer-review literature concerning valuation-approaches for startup markets.

Second, by elaborating on already existing entrepreneurial finance research, this study gives rise to justification for a second-look at existing empirical findings, a research avenue which may indeed prove fertile. Existing studies which use segmented approaches devote little space to exploring model functional-form. Here again, the overall need for model-sophistication in order to meaningfully incorporate variable interaction-effects and variable-hierarchies within valuation-models is not only apparent, but also likely more relevant for startup markets than for more established (i.e., information-rich) markets.

Third, this study describes the use of newly-emergent empirical techniques and describes how to systematically make use of them in a consistent way. While hierarchical decision-tree-based microtargeting can take multiple forms in terms of machine learning algorithms (i.e., recursive-partitioning, agglomerative hierarchical clustering, random forest), the modelling functional-form that can be applied for startup valuation, startup-selection, or startup-survival intended to accompany such modelling-approaches has heretofore not yet appeared in the literature. This may be due to the overall novelty of such approaches within published entrepreneurial-financial literature up until now.

Given that machine learning approaches generally confront questions of model-selection and algorithm-selection relatively early-on, further research using the principles outlined in this paper should consider both model-complexity and shape of functional-form as a fundamental part of model-selection and algorithm-selection, as a combined model-outlook. Furthermore, this combined-outlook can and should be taken into consideration for all applications of machine learning approaches within economics, finance, or firm-strategy, or entrepreneurship research, as well and practice thereof in the professional marketplace.

Implications of this research are far-reaching. For markets and industry practitioners, elaboration on why and how segmented valuation models work, as well as how specifically they relate to emerging machine learning approaches can lead to the development of new and bespoke valuation-models going forward, as industry practitioners may increasingly adopt this style of valuation-approach. Meanwhile, the emergence of investors linked to the big data and machine learning industries (ranging from CVCs to specialized consultants and experts) may someday try to automate tree-based segmented-valuation approaches, in contexts where it may be appropriate to do so (for instance, implementation of trading-algorithms in a crowdfunding-platform or P2P-lending-platform setting). For investors, as well as for third-parties, implications are also far-reaching because these models can hypothetically deliver accurate valuation-estimations via microtargeting, which in its least numerical forms is able to bypass difficult-to-obtain or confidential firm-level accounting data, making accurate valuations considerably more widespread within startup markets.

For policy-maker circles meanwhile, the implications segmented model proliferation as machine learning approaches develop and evolve, are the rise of a more niche and targeting understanding of startup markets, a body of knowledge which may be very useful for the purposes of SME policy, as well as in targeting key sectors, asset-classes, regions, or municipalities going forward.

Fundamentally, future research will be able to build on this study by deploying modelling principles described here for empirical studies featuring hierarchical machine learning approaches for the development of segmented startup-valuation models. Since this approach is still in relatively-early phases of emergence, it may be feasible to "push the envelope" on what is empirically feasible. Doing so can be helped, for instance by development of taxonomy studies of entrepreneurial-finance-relevant configurations, clusters, and categorical variables, so that future microtargeting research can grow beyond reliance on industry-sector, business-model, and economic-geography variables (e.g., regions, cities, municipalities, or postal-codes).

Additionally, future research may build on this study by expanding the use of hierarchical empirical approaches to construct segmented models in other areas of entrepreneurial finance. Two topics adjacent to startup-valuation, which are also core to the entrepreneurial finance field are startup-selection (Berre & Le Pendeven, 2022), and startup-survivability. In principle, hierarchical empirical-approaches can be used to create segmented models to describe and predict these as areas as well. In particular, the approach can be useful for scholars interested in predicting startup-selection, as well, since qualitative factors play a more prominent role here than in startup-valuation (Berre & Le Pendeven, 2022), which may require a more sophisticated approach than OLS, capable of using both qualitative and quantitative data in order to estimate predictions. Startup-survivability on the other hand, would be most useful in an industry practitioner setting, where a way to accurately model any given startup's likelihood of survival or bankruptcy has the potential to substantially impact a VC's commercial outcomes.

Lastly, this research can be used as a roadmap for forthcoming studies intending to make use of hierarchical machine learning techniques within entrepreneurial finance, for industry practitioners interested in deploying machine learning techniques to establish bespoke segmented valuation models, or machine learning professionals interested in deploying their expertise for entrepreneurial finance (for example in a fintech setting).

References

Bellavitis, C., Filatotchev, I., Kamuriwo, D. S., & Vanacker, T. (2017). Entrepreneurial finance: New frontiers of research and practice: Editorial for the special issue embracing entrepreneurial funding innovations. *Venture Capital: An International Journal of Entrepreneurial Finance, 19*(1–2), 1–16.

Benoit, K. (2011). *Linear regression models with logarithmic transformations*. Methodology Institute London School of Economics. Working Paper.

Berkus, D. (2016). *The Berkus method - Valuing the early-stage investment*. Berkonomics.

Berre, M. (2022). Which factors matter most? Can startup valuation be micro-targeted? In *International conference on small business (ICSB) world conference, Washington DC. Working paper*.

Berre, M., & Le Pendeven, B. (2022). What do we know about start-up valuation drivers? A systematic literature review. *Venture Capital*, 1–45. https://doi.org/10.1080/13691066.2022.2086502

Budhwar, P., Cumming, D., & Wood, G. (2022). Entrepreneurial finance and the legacy of Mike Wright. *British Journal of Management, 33*, 3–8.

Damodaran, A. (2002). *Investment valuation: Tools and techniques for determining the value of any asset* (Wiley Finance Series). John Wiley & Sons.

Damodaran, A. (2009). Valuing young, start-up and growth companies: Estimation issues and valuation challenges. SSRN Scholarly Paper ID 1418687. : Social Science Research Network

Damodaran, A. (2010). *It is all relative... multiples, comparables and value!* NYU Stern. Retrieved December 2022, from http://pages.stern.nyu.edu/~adamodar/pdfiles/country/relvalAIMR.pdf

Damodaran, A. (2021). Equity Risk Premiums (ERP): Determinants, estimation, and implications – The 2021 edition. Available at SSRN: https://ssrn.com/abstract=3825823 or https://doi.org/10.2139/ssrn.3825823

Ernst & Young. (2020). Startup funding full eGuide - The Factory. EY - The Factory. Retrieved June 30, 2022, from https://thefactory.works/wp-content/uploads/2020/10/Eguide_Funding_A4.pdf?mc_cid=df85bfb95a&mc_eid=f80b29de26

Ewing Marion Kauffman Foundation. (2007). *Valuing pre-revenue companies*. Angel Capital Association. https://www.angelcapitalassociation.org/data/Documents/Resources/AngelCapitalEducation/ACEF_-_Valuing_Pre-revenue_Companies.pdf

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*, 383–417.

Fama, E. (1991). Efficient capital markets: II. *The Journal of Finance, 46*(5), 1575–1617.

Firmin, S. (2021). Understanding the outputs of the decision tree tool. *Alteryx Designer Knowledge Base*. Retrieved September 1, 2022, from https://community.alteryx.com/t5/Alteryx-Designer-Knowledge-Base/Understanding-the-Outputs-of-the-Decision-Tree-Tool/ta-p/144773

Goldman, M. (2008). Valuation of startup and early-stage companies. *The Value Examiner*. http://www.michaelgoldman.com/Publications/Goldman%20Valuation%20of%20Start-ups.pdf

Hand, J. R. M. (2005). The value relevance of financial statements in the venture capital market. *The Accounting Review, 80*(2), 613–648.

Keene, O. N. (1995). The log transformation is special. *Statistics in Medicine, 14*(8), 811–819.

Khan, I., Capozzoli, A., Corgnati, S. P., & Cerquitelli, T. (2013). Fault detection analysis of building energy consumption using data mining techniques. *Energy Procedia, 42*(2013), 557–566.

Krzywinski, M., & Altman, N. (2017). Classification and regression trees. *Nature Methods, 14*, 757–758. https://doi.org/10.1038/nmeth.4370

Mahmoud, F., Zahoor, A., Hussain, N., & Younes, B. Z. (2022). Working capital financing and firm performance: A machine learning approach. In *Financial economics meeting (FEM-2022). Paris*.

Miller, J., O'Neill, M., & Hyde, N. (2010). *Intermediate Algebra* (2nd ed.). McGraw-Hill Higher Education.

Murray, G. R., & Scime, A. (2010). Microtargeting and electorate segmentation: Data mining the American National election studies. *Journal of Political Marketing, 9*(3), 143–166. https://doi.org/10.1080/15377857.2010.497732

Neter, J., Wasserman, W., & Kutner, M. H. (1990). *Applied linear statistical models: Regression, analysis of variance, and experimental design* (3rd ed.). McGraw-Hill, Inc.

Payne, B. (2011) Scorecard valuation methodology: Establishing the valuation of pre-revenue, start-up companies. https://www.angelcapitalassociation.org/blog/scorecard-valuationmethodology-rev-2019-establishing-the-valuation-of-pre-revenue-start-up-companies/

Ribeiro-Oliveira, J. P., Garcia de Santana, D., Pereira, V. J., & Machado dos Santos, C. (2018). Data transformation: An underestimated tool by inappropriate use. *Acta Scientiarum Agronomy., 40*(1), e35300. https://doi.org/10.4025/actasciagron.v40i1.35300

Sandeep. (2014). (https://stats.stackexchange.com/users/45985/sandeep), Difference between rel error and xerror in rpart regression trees. URL (version: 2014-07-02): Retrieved September 1, 2022, from https://stats.stackexchange.com/q/105536

Sievers, S., Mokwa, C. F., & Keienburg, G. (2013). The relevance of financial versus non-financial information for the valuation of venture capital-backed firms. *European Accounting Review, 22*(3), 467–511.

Siskos, J., & Zopounidis, C. (1987). The evaluation criteria of the venture capital investment activity: An interactive assessment. *European Journal of Operational Research, 31*(3), 304–313.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). The MIT Press Cambridge.