

BIG DATA-DRIVEN STOCHASTIC BUSINESS PLANNING AND CORPORATE VALUATION

Roberto Moro Visconti ^{*}, Giuseppe Montesi ^{**}, Giovanni Papiro ^{**}

^{*} Corresponding author Department of Business Administration, Università Cattolica del Sacro Cuore, Milan, Italy

Contact details: Università Cattolica del Sacro Cuore – Largo A. Gemelli, 1 – 20123 Milan, Italy

^{**} School of Economics and Management, Università degli Studi di Siena, Italy



Abstract

How to cite this paper: Moro Visconti, R., Montesi, G., & Papiro, G. (2018). Big data-driven stochastic business planning and corporate valuation. *Corporate Ownership & Control*, 15(3-1), 189-204. <http://doi.org/10.22495/cocv15i3c1p4>

Copyright © 2018 The Authors

This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0). <http://creativecommons.org/licenses/by-nc/4.0/>

ISSN Online: 1810-3057

ISSN Print: 1727-9232

Received: 11.03.2018

Accepted: 18.04.2018

JEL Classification: C15, C53, C22, G31, O22

DOI: 10.22495/cocv15i3c1p4

The research question of this paper is concerned with the investigation of the links between Internet of Things and related big data as input parameters for stochastic estimates in business planning and corporate evaluation analytics. Financial forecasts and company appraisals represent a core corporate ownership and control issue, impacting on stakeholder remuneration, information asymmetries, and other aspects. Optimal business planning and related corporate evaluations derive from an equilibrated mix of top-down and bottom-up approaches. While the former follows a traditional dirigistic methodology where companies set up their strategic goals, the latter are grass-rooted with big data-driven timely evidence. Real options can be embedded in big data-driven forecasting to make expected cash flows more flexible and resilient, improving Value for Money of the investment and reducing its risk profile. More accurate and timely big data-driven predictions reduce uncertainties and information asymmetries, making risk management easier and decreasing the cost of capital. Whereas stochastic modeling is traditionally used for budgeting and business planning, this probabilistic process is seldom nurtured by big data that can refresh forecasts in real time, improving their predictive ability. Combination of big data and stochastic estimates for corporate appraisal and governance issues represents a methodological innovation that goes beyond the traditional literature and practice.

Keywords: Revenue Model, Forecasting, Free Cash Flow, Real Options, Valuation Metrics, Stochastic Simulation, Sales

Acknowledgement: This publication has been financed with research funds from the Università Cattolica del Sacro Cuore of Milan, Italy (Line D.3.1. /2018).

1. INTRODUCTION

Business planning is strongly related to the evaluation of corporations. Both metrics are often flawed by imprecise forecasts that bring to huge differences between expectations and real outcomes. The research question of this study is focused on (optimal) forecasting, concerned with predictive information and analysis of trends transferred over time, which can be improved with a combination of big data and stochastic projections. While big data raise the quality, quantity, and timeliness of input information, stochastic forecasts upgrade the evaluation mechanism in a condition of uncertainty.

Interactions of big data and stochastic modeling may so bring to better estimates and deeper analysis.

Risk-adjusted business planning makes a probabilistic evaluation of different strategies characterized by different probability distributions of stochastic variables or following alternative big data-driven patterns). These alternative strategies may be confronted to select the one that is most likely to achieve the company's goals.

Corporate evaluation is a key parameter for corporate governance metrics. Stakeholders are naturally interested in the appraisal of their company, for remuneration purposes and other complementary reasons. Even conflicts of interest and information

asymmetries, which represent core corporate governance concerns, are strongly influenced by the measurement of the company's perspective value and are reduced when predictive quality improves.

This research question is particularly relevant in a world that lives on expectations that shape actions and strategies, magnetizing resources that look for proper rewarding. Whenever the difference between expectations and real outcomes narrows, the risk is minimized, and efficiency/efficacy in the use of (limited) resources is improved.

Big data is the term for any gathering of large-volume information sets from multiple sources and is so expensive, fast-changing, and complex that can become hard to process. The difficulties incorporate investigation, catch, duration, inquiry, sharing, stockpiling, exchange, perception, protection infringement and quantification of financial value. The explosive growth of data in almost every industry and business area is driven by the rapid development of the web, Internet of Things (IoT), and cloud computing (Jin et al., 2015).

Characteristics as volume, velocity, variety, and veracity make big data particularly interesting for sophisticated economic and financial planning, where several variables stored in interoperable databases need to be simultaneously considered. Big data are driving better decision making and can help to detect growth drivers.

Stochastic modeling is a form of financial modeling that includes random variables to estimate how probable outcomes are within a forecast to predict different states of the world.

The literature concerning either big data or, respectively, stochastic processes and business planning or corporate valuations is constantly growing and has a long tradition, particularly if it concerns the latter topics, since big data are relatively recent. Little if any research has however been dedicated to these joint topics as it will be pointed out in the discussion.

While data have always existed, "big" data represent a novelty, with many undiscovered applications.

The structure of the paper is the following: a short literature review will show in paragraph 2 that the research question is innovative. The methodology illustrated in paragraph 3 explains the process innovation that jointly considers big data and stochastic estimates.

Traditional business planning and corporate evaluation metrics will be briefly described in paragraph 4, and financial information will be integrated with big data sources and then stochastic modeling.

Figure 2 in paragraph 5 will describe in intuitive terms how big data interact with cross section data and time series that follow a stochastic pattern.

Paragraph 6 will summarize the stochastic approach to business planning, with a multi-period model that expresses forecasting variables in probabilistic terms.

Big data forecasting will be critically examined in paragraph 7, showing how they affect the evolutionary pattern of sales, which represent a key variable in the budgeting process.

The stochastic approach is then combined with big data applications in paragraph 8.

Paragraph 9 will examine how real options,

nurtured by Internet of Things and big data, can support budgeting and planning with bottom up evidence from the real world.

The discussion provides a critical analysis of the main findings and the corporate governance implications, together with tips for new research avenues. The conclusion summarizes the results and the findings, consistently with the research question.

2. LITERATURE REVIEW

This topic is innovative and so it cannot be properly referred to a specific literature strand. There are however complementary issues that are discussed in the current literature and represent an inspiration source.

Currently, corporate governance research follows two major routes: classical empirical corporate governance research and multidisciplinary research aimed at finding non-conventional methods to solve existing problems (Kostyuk et al., 2018). This study is multidisciplinary and deals with risk and firm performance issues that are strongly related to corporate valuation. There are however implications for other governance matters as family ownership, regulation, financial markets, accounting and audit concerns.

Corporate valuation remains a hotly debated issue (Bancel & Mittoo, 2014; Vernimmen et al., 2014), especially when it concerns uneasy forecasts of key value drivers.

Corporate governance affects firm value and effective governance reduces the information asymmetry through better financial reporting quality (Latif et al., 2017).

Big data help enterprises to make well-informed decisions (Al-Barznji & Atanassov, 2017), and can so be usefully incorporated in business planning and predictive analytics (Hazen et al., 2014; Moro Visconti et al., 2017). Big data analytics are also used in marketing (Zhenning et al., 2016), which is linked to sales forecasting, a crucial issue of business planning, as it will be show in paragraph 8.

Cognitive big data represent a powerful, albeit under-exploited, source of information for descriptive, prescriptive, and predictive analytics (Franks, 2014; Jumi et al., 2016; Tanner, 2014; Jin et al., 2015) supporting decisions especially in data-rich industries. Big data mining supports business analytics (Duan & Xiong, 2015).

Recent studies have empirically shown how big data could significantly improve the accuracy of forecasting. See Moro Visconti et al., (2017), Archak et al. (2011), Lau et al. (2018). More specifically, social media and online communication can be a relevant source of information in predicting sales dynamic (Xiaohui et al., 2012; Sonnier et al. 2011).

Big data are also used in accounting (Vasarhelyi et al., 2015), which provides the basic input data for valuation. Big data mining can also reduce enterprise risk (Olson & Wu, 2017), with a positive impact on corporate valuation and governance concerns.

There is also a link between big data and corporate governance (Tallon, 2013).

Stochastic simulations have been increasingly employed as a powerful technique to effectively address in an innovative way typical corporate finance analysis issues, such as equity valuation or credit risk assessment (Montesi & Papiro, 2014).

Big data are also combined with stochastic estimates and statistics (Daas et al., 2015; Karim et al., 2017; Chen et al., 2018) and with data-driven simulations that can optimize complex stochastic systems (Xu et al., 2015).

This study combines the abovementioned literature trends proposing a product and process innovation and putting big data-driven data at the center of valuation issues in a stochastic scenario.

3. METHODOLOGY

The methodology of this study follows a conceptual definition of the fundamental issues of the research question, which are progressively illustrated and finally jointly considered.

In this study we propose a process innovation methodology that considers the interaction of the following methods / parameters:

1. Cross-section data integrated with time series analysis;
2. Stochastic modeling of time series forecasts;
3. Big data inputs for both cross-section data and time series analysis;
4. Real options to feed big data and stochastic forecasts.

Even if these four methodologies are already widely known and used in forecasting analysis and valuation, their joint consideration represents a process innovation. Big data are the cornerstone of this process since they feed both cross-section data and their evolving patterns through stochastic variables modeling.

Traditional business planning techniques combine expected balance sheet and income statements, to get forecast cash flows. This standard accounting metrics produces key parameters that are used in corporate evaluations, like discounted cash flows or EBITDA forecasts. An overview of these well-known techniques is summarized in par. 3.

The methodology of this study is innovative since it combines in an organic framework these apparently unrelated topics. Other complementary methods of research could be used, showing the quantitative relationships among big data and stochastic estimates and incorporating them in algorithms. A further implementation may use artificial intelligence and machine learning as a tool for optimal forecasting, increasing the number of variables instantaneously captured in the model for deeper and wider scenarios.

4. TRADITIONAL BUSINESS PLANNING AND CORPORATE EVALUATION METRICS

Traditional economic and business planning is based on an interaction of accounting data coming from the forecast income statement (profit & loss account), matched with a pro-forma balance sheet to derive a forecast cash flow statement. The estimate of future revenues, costs and cash flows represents the cornerstone of corporate evaluation methodologies. Company appraisal is based on complementary methodologies that bring to an estimate of either the enterprise or the equity value of the corporation. While the former represents the overall value, comprehensive of debt, the latter is related to the residual shareholder value.

The two most common methodologies for

estimating the enterprise value respectively refer to discounted cash flows (DCF) or sustainable EBITDA (Earnings Before Interests, Taxes, Depreciation, and Amortization) applied to a market multiplier of comparable entities.

DCF metrics is based on operating cash flows discounted at the weighted average cost of capital (WACC), according to the following formula:

$$WACC = \frac{E}{EV} \cdot Re + \frac{D}{EV} \cdot Rd \cdot (1 - Tc) \quad (1)$$

Where:

Re = cost of equity

Rd = cost of debt

E = market value of the firm's equity

D = market value of the firm's debt

$EV = E + D$ = total market value of the firm's financing (equity and debt)

E/EV = percentage of financing with equity

D/EV = percentage of financing with debt

Tc = corporate tax rate

Since operating cash flows are unlevered, i.e., expressed before debt servicing, they are discounted using WACC, a consistent parameter that weights the cost of capital collected from debtholders and equity-holders. Enterprise Value is the result of such an estimate.

Another estimate can be conducted with a complementary method based on a forecast average EBITDA, the only parameter that simultaneously expresses economic and financial operating marginality. Whenever EBITDA is multiplied by the average Enterprise Value/EBITDA of comparable companies, a market value estimate of the company's Enterprise Value comes out.

Whereas business planning produces an estimate of the accounting set, corporate evaluation represents one of its most valuable outcomes.

A further evaluation parameter is represented by excess returns, i.e., by returns that exceed industry and risk-adjusted benchmarks. Excess returns represent a key component of goodwill (Nijeholt & Grift, 2007) and discontinuously accrue over time, accumulating value that must be shared among stakeholders (debtholders and residually shareholders). Forecasting of excess returns is a key part of business planning and corporate evaluation models.

Excess return is correlated to Economic Value Added that estimates the firm's economic profit, or the value created more than the required return of the company's shareholders. EVA is the net profit less the equity cost of the firm's capital. Value is created when the return on the firm's economic capital employed exceeds the cost of that capital. EVA can bring to a complementary estimate of the Enterprise Value.

5. INTEGRATING FINANCIAL INFORMATION WITH BIG DATA

The classic evaluation model illustrated in paragraph 4. can be usefully integrated with big data analytics. Each information can improve and refresh the predictive power of key estimation variables, previously indicated.

Although the basic template for business planning and corporate valuations remains unchanged, big data may so impact on the main

economic and financial parameters and assumptions.

This information needs to be collected, processed, and made available to users through digital platforms where stakeholders meet and cooperate to co-create value. Business models, intrinsically data-centric and data-driven, are primarily connected with digital information provided by big data. Predictive models exploit patterns of historical and transactional (big) data to forecast market trends, with their risks and opportunities.

Information and knowledge extracted from big data are relevant for business planning in many complementary ways: velocity (examined in table 1) allows a mark-to-market update and refresh of forecasts that flexibly adapt to changing market conditions, whereas data volume makes estimates more precise and less volatile.

Big data improve the features of the key planning and evaluation parameters:

- Sales (operating revenues) are the major driver that brings to EBITDA, after having considered fixed and variable operating (and monetary) costs. While fixed costs are irrespective of the volume of sales, variable costs represent a percentage of the sales that can be estimated through stochastic models, fueled by timely big data inputs.

- Supply chain management (smart logistics). Big data can foster inventory/logistic optimization and supplier coordination, shortening the supply chain and making it more resilient to external shocks. This has a positive impact on operating leverage and on economic and financial margins as EBITDA.

- EBITDA can be interpreted as a synthesis of operating revenues and costs, expressed in monetary

terms. Big data can shed light both on the revenue model and on its supporting costs;

- Operating cash flows (Cfo) derive combining EBITDA with net working capital and capital expenditure (capex), made available from forecast balance sheets. Big data can improve the analysis and prediction of the combination of income statement margins as EBITDA with their backing assets and liabilities (accounts receivables and payables, inventory, fixed assets ...);

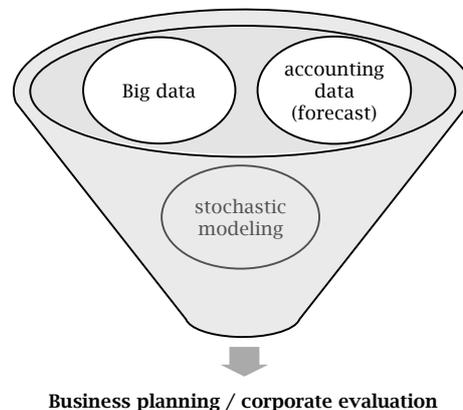
- WACC is another parameter that considers the weighted cost of collected capital. Big data can help to detect and fine-tune its outlook, estimating its evolution with always updated inputs and jointly considering internal parameters (the stock of debt and equity) with expected interest rates or cost of equity;

- The Net Financial Position (NFP) is the algebraic difference between financial debts and liquidity. NFP and debt servicing costs are algebraically added to Enterprise Value to pass to Equity Value (in economic terms, from operating to net income and in financial terms from operating to net cash flows). Big data can ease the prediction of estimated Net Financial Positions, linking it to the evolution of the Capex and the Net Working Capital that depend on input drivers (as sales).

The market multiplier (m) of comparable corporations represents similar assets that should sell at similar prices, determining the value of a company based on the value of another.

Figure 1 shows how valuable information can be extracted from big data and accounting data.

Figure 1. Extraction of informative value from big data and accounting data



While cross-section data concern the observation of a variable's determinants at the same point of time, time series data observe its dynamics.

Economic datasets come in a variety of forms. The cross-sectional, time series, and panel data are the most commonly used kinds of datasets. A cross-sectional dataset consists of micro- or macroeconomic units taken at a given point in time. A time series dataset contains information on a variable or a set of variables over time. Panel (longitudinal) data combine both dimensions (time-series and cross-sectional) at the same time.

The use of time series data for forecasting purposes is an integral component of management,

planning and decision making. Following the big data trend, substantial amounts of time series data are available from many heterogeneous data sources in more and more applications domains (Hartmann et al., 2015).

The interaction between cross-section data and time series analysis, nurtured by big data in a stochastic evolutionary pattern, is illustrated in Figure 2.

Economic and financial information can be linked to big data considering their dimension and their impact on traditional value and supply chains. Characteristics of big data can be represented in Table 1.

Figure 2. Modeling sales growth patterns with big data and stochastic processes

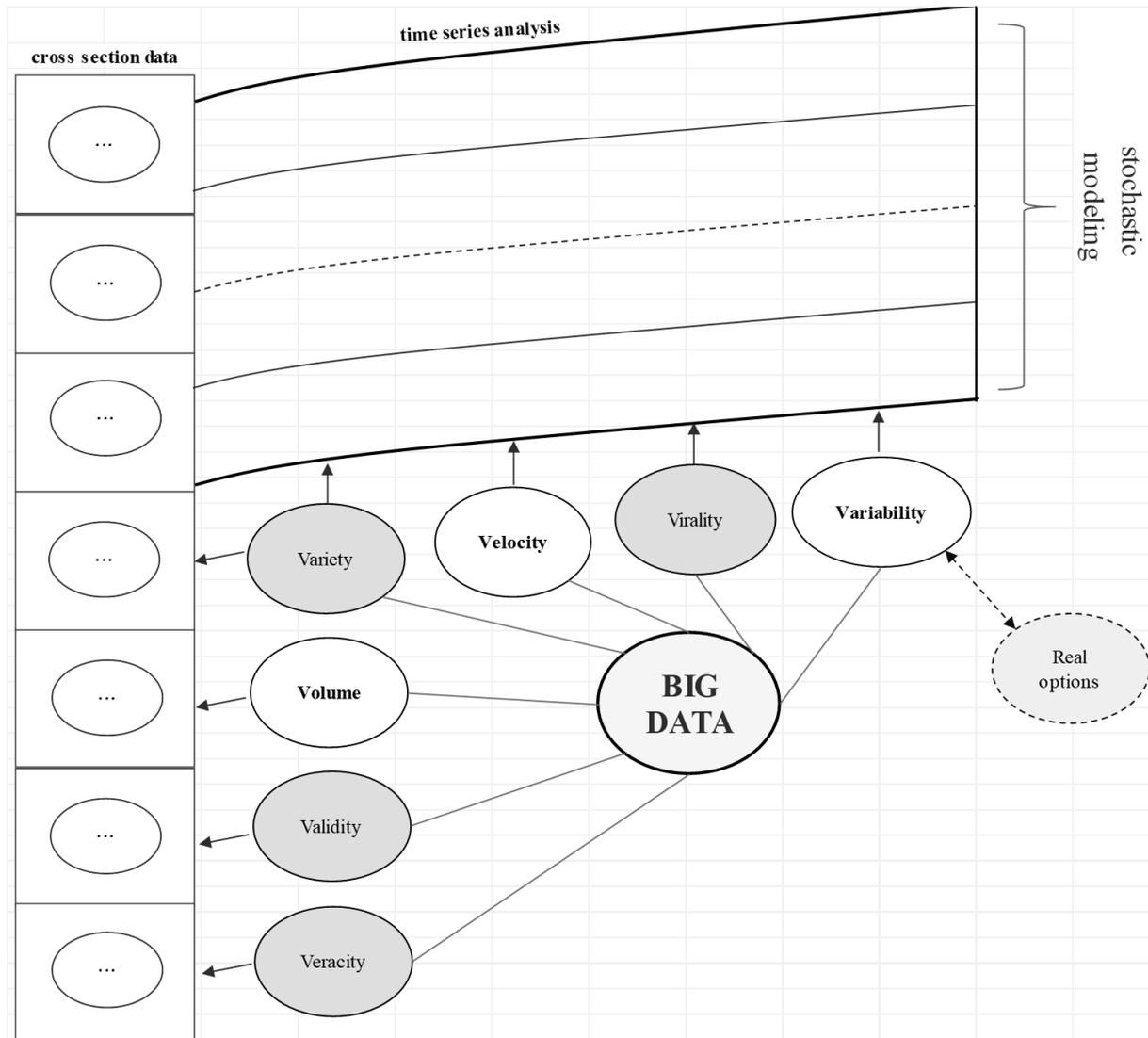


Table 1. Impact of big data dimensions on standard forecasting

<i>Big data dimensions</i>	<i>Impact on standard forecasting</i>
Volume	Big volumes of data dramatically increase the width, scale, and quality of available information. Its use for budgeting and reporting purposes would allow for more accurate estimates. Differences between forecast and actual data decrease, reducing risk (inversely proportional to Value for Money) and making the supply chain more resilient. Relationships among stakeholders are likely to improve.
Velocity	Data are accumulated in real-time and at a rapid pace. The velocity of proliferating data increases when the system (represented for instance by interoperable databases) improves, due to machine learning and artificial intelligence. Velocity, like volume, may minimize the gap between forecast and actual data.
Variety	Evidence-based samples combine and analyse with big data fusion a variety of structured, semi-structured, and unstructured data, to match forecasts with outcomes, predict risk patterns, and provide deeper and more effective analyses. Variety increases the comprehension of the stakeholders' needs.
Veracity	Key parameter in investments, corresponding to data reliability. Increased variety and high velocity hinder the ability to cleanse data before analysing it and making decisions, magnifying the issue of data "trust". Veracity can contribute to minimizing opportunistic behaviours and conflicts of interest.
Validity	Data integrity is defined as the validity, accuracy, reliability, timeliness, and consistency of the data. It is linked to veracity.
Variability	Variability of data increases their information value and should be linked with other parameters as velocity or variety. When variability is considered and promptly incorporated in updated business models, the risk is reduced.
Virality	Measures the spread rate of data (sharing speed) across the network. It increases the involvement of different stakeholders.
Visualization	The connection between information visualization and visual analytics with IT systems through representation technologies would help users to understand data. The synthesis produced by data visualization tools is a key element to transform the information revealed by big data processing, understood only by specialists, into accessible knowledge.
Viscosity	Characterizes the resistance to navigate in the dataset or complexity of data processing. It is a common feature of complex data in many industries.
Value	Monetized value (Walker, 2015) is the synthesis of big data V-dimensions, considering data as an asset to exploit to produce innovation and new information-sensitive products and services.

6. STOCHASTIC APPROACH TO BUSINESS PLANNING: A SYNTHESIS

Business planning is a process focused on developing assumptions and forecast scenarios over a multi-annual time, aimed at supporting management in taking optimal investment decisions and strategies. Since the future is uncertain and we do not know what may occur in the medium and long-term, business planning generally considers more than one scenario to check what may happen under different forecast assumptions.

A stochastic simulation approach to business planning (Wong, 2007; Xu & Birge, 2006) goes further in this direction, allowing the business planners to handle and develop an extremely high number of different scenarios, going well beyond the traditional base/best/worst case scenario analysis, exploring virtually all the potential future patterns of the company, thus notably deepening the quantity and quality of information available for analysis and strategic decisions. The main advantage of this technique is that it enables to represent results in probabilistic terms that is the best method to take decisions under the condition of uncertainty. It allows to develop innovative solutions to specific problems that could not be obtained with traditional deterministic models. For example, we can obtain an ex-ante estimate of the probability that a given event – such as reaching a profitability target or the company default – will occur.

In this way, it is possible to compare different strategic plans to choose the one that presents the higher probability to reach the company's targets (profitability, dividends, equity fair value, etc.) or the one that minimizes the risk, showing the lowest probability of default of the company or triggering a critical threshold (for example, for a bank breaching the minimum capital ratio regulatory requirements).

In a nutshell, the approach is based on a stochastic simulation process (generated using the Monte Carlo method, see Rubinstein & Kroese, 2017), applied to a forecasting model, which generates thousands of different multi-period random scenarios, in each of which coherent projections of the company's income statement and balance sheet are determined. The random forecast scenarios are generated by modeling all the main value drivers (sales, operating costs, investments, etc.) as stochastic variables. The simulation results consist of distribution functions of all the output variables of interest: net income and margins, financial ratios, economic value added, etc. More specifically, the framework is based on the following features:

Step 1: multi-period stochastic forecasting model

A forecasting model to develop multiple scenario projections for income statement and balance sheet, capable of managing all the relevant value drivers to consistently ensure:

- The balancing of total assets and total liabilities in a multi-period context, so that the financial surplus/deficit generated in each period is always properly matched to a corresponding (liquidity/debt) balance sheet item.
- The setting of rules and constraints to ensure a satisfactory level of intrinsic consistency and correctly manage potential conditions of non-linearity. The most important requirement of a

stochastic model lies in preventing the generation of inconsistent scenarios. In traditional deterministic forecasting models, consistency of results can be controlled by observing the entire simulation development and set of output. However, in stochastic simulation, which is characterized by the automatic generation of a very large number of random scenarios, this kind of consistency check cannot be performed, and we must necessarily prevent in-consistencies ex-ante within the model itself, rather than correcting them ex-post. In practical terms, this entails introducing into the model rules, mechanisms and constraints that ensure consistency even in extreme tail scenarios.

Step 2: forecasting variables expressed in probabilistic terms

The variables that have the greatest impact on the model's results and those of which the future value is most uncertain are modeled as stochastic variables and defined through specific probability distribution functions to establish their future potential values, while interdependence relations among them (correlations) are also set. The non-stochastic variables can be functionally determined within the forecasting model, by being linked to the value of other variables (for example, regarding relationship to or percentage of a stochastic variable) or expressed regarding functions of a few key figures or simulation outputs. In this regard, it may be useful to express some variables as a function of a key independent variable; for example, variable costs and/or investments (capex) can be expressed as a function of sales. Therefore, to model the stochastic simulation, we must:

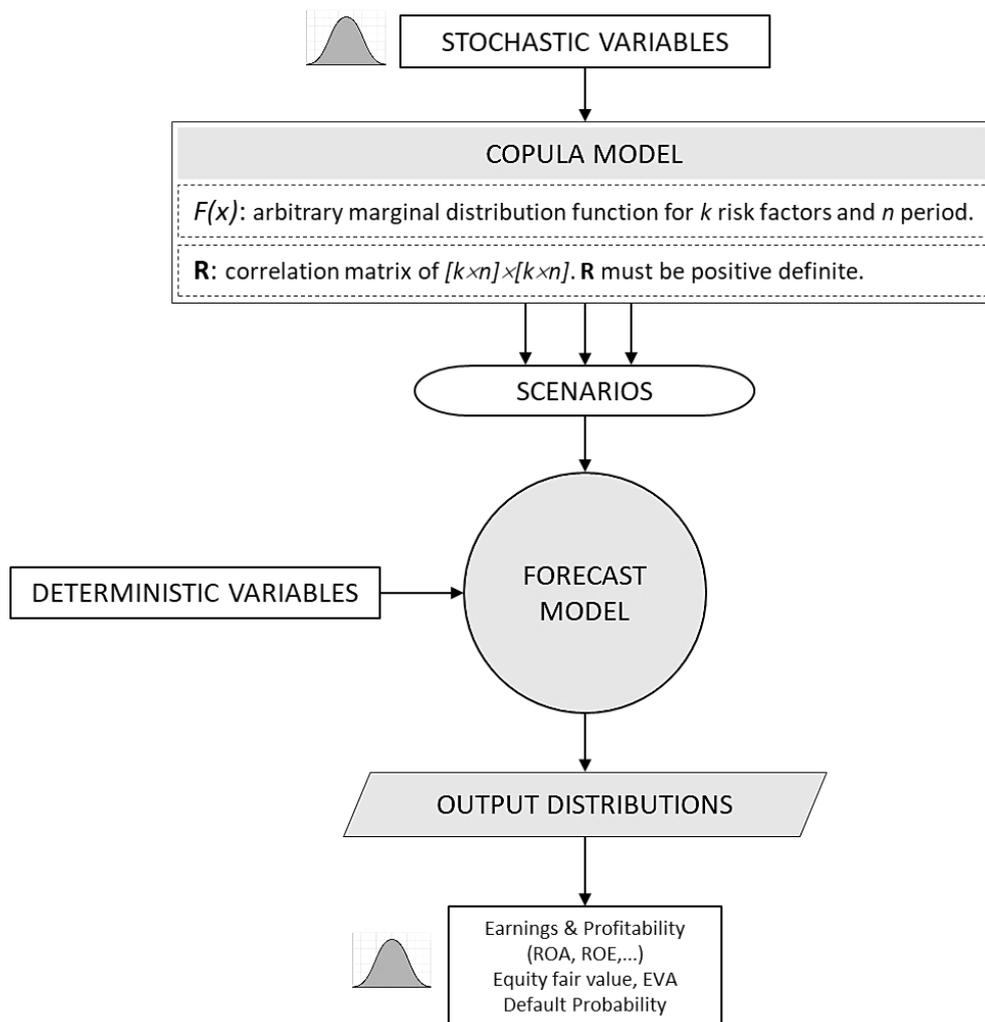
- Select the stochastic variable and set for each of them the kind of function that is best suited to describe the variable (i.e., Normal, Beta, Weibull, etc.), the two parameters that defines the distribution function (i.e., mean and standard deviation, minimum and maximum, etc.) and, depending of the kind of function chosen, the shape parameter that may set potential asymmetry and kurtosis features that need to be modeled;
- Set the correlation matrix for the stochastic variables, through the definition of cross-sectional and time-series dependence.

Step 3. Monte Carlo simulation

This technique allows to solve the stochastic forecast model in the simplest and most flexible way. The stochastic model can be constructed, for example, using a copula-based approach, with which it is possible to express the joint distribution of random variables as a function of the marginal distributions. For a description of the modeling systems of random vectors with arbitrary marginal distribution allowing for any feasible correlation matrix, see Rubinstein & Kroese (2017), Cario & Nelson (1997), Robert & Casella (2004), Nelsen (2006).

Analytical solutions – if it is possible to find them – would be too complicated and strictly bound to the functional relation of the model and of the probability distribution functions adopted, so that any change in the model and/or probability distribution would require a new analytical solution. The flexibility provided by the Monte Carlo simulation, however, allows to easily modify the stochastic variable probability functions.

Figure 3 illustrates the forecast modeling with deterministic and stochastic variables.

Figure 3. Forecast models with deterministic and stochastic variables

7. ENHANCEMENT AND LIMITS OF BIG DATA FORECASTING

Forecasting is increasingly difficult when the time horizon grows, and so (explicit and implicit) long-term forecasts need to be sound and grounded in theory. For example, high growth has necessarily to be nurtured by new investments, and strong competitive advantages sooner or later expire and converge to average growth. In this regard, the concept of competitive advantage period (CAP), which represents the number of years during which a company generates returns on incremental investment that exceed its cost of capital, matters.

CAP depends on several factors, such as a dominant market share, regulation, patents, trademark, management skills, technology, etc. In a competitive environment, sooner or later the CAP will progressively fade away, balancing the return on new investment to the cost of capital.

In a globalized economy, the CAP tends to shorten even more quickly. This simple economic principle should always be kept in mind when making long-term business planning where trendy variables are naturally converging towards a sustainable mean (see Mauboussin & Johnson, 1997; Mauboussin, 2006).

For example, by exploiting consumers'

comments on the social media, it is possible to "learn" the relationship between a forerunner key indicator of the sales dynamic in a specific product; in this way, by observing the evolution of the indicator it is possible to achieve accurate forecasts for a key variable such as sales. Thence, for instance, positive comments on social media about a movie, allow to forecast a certain increase in that movie sales (Xiaohui & Liu, 2012) in the next weeks; positive/negative reviews on shopping on line portal (such as Amazon) on a certain product, may predict the future dynamic of that product sales. Recurring to other sources of data, a similar approach might be considered to forecast other key variables for business planning, such as operating costs and working capital; for credit risk analysis information about the cash payments made and received by a company might be used to forecast its liquidity/solvency conditions, so assessing its risk of default.

This kind of big data use can provide extremely valuable information for company's key variables predictions, but in any case, these forecasts are still subject to a not negligible margin of error that needs to be considered. Also, they are limited to short-term forecasting, with a time horizon of weeks or at maximum months. In fact, considering the above-mentioned examples of movie sales forecast, we can

use comments and reviews on the latest movies to assess their forthcoming sales, but we will never be able to use that big data approach (comments) to forecast the sales of following years related to the movies that still must be broadcasted.

This kind of big data forecasting approach is therefore limited to short-term budgeting purposes (within one year), such as setting the level of sales for the current fiscal year or adjusting the budget targeting for the next quarter. Unfortunately, for company valuation purposes we mostly need medium-long-term forecasts, since generally most of the future cash flows will be generated by the company in the long run and not in the next quarters. For that aim, we must adopt a different forecasting process.

To make reasonable forecasts over the medium-long-term, we necessarily must recur to theoretical economic models, which somehow provide rational constraints to our forecasting process. Going beyond the short-term, the only “certain” benchmark for reasonable assumptions about the future evolution of the company’s performances are: a “starting point” based on the last historical available company’s records; and a long-term “arrival point” at which the company should tend to converge, based on theoretical economic conditions of equilibrium and sustainability. For example, considering a company with high profitability margins (well above the long-term average returns) due to its competitive advantage position, it is unlikely that those conditions will persist indefinitely; it is more realistic to assume that its profitability should converge towards sustainable long-term returns.

In this context, forecasts should assume a trend of convergence towards an equilibrium level. Of course, due to the uncertain nature of the economic world we cannot exactly foretell year by year, how fast or slow or complete this convergence process will be; and that is why it makes no sense to develop

deterministic projections. A stochastic approach to forecasting can so be helpful. Instead of making punctual projections, it is easier and more reasonable to manage them in probabilistic term, considering for each driver a distribution function of possible forecast values for each year of projection, and assuming an evolutionary pattern of its parameters (e.g., mean and standard deviation) through the years, simulating their convergence towards an equilibrium level.

The medium-long-term forecasting process could so be structured setting an evolutionary pattern of the two parameters that define the distribution function of the stochastic variable through the forecasting period, for example, the mean and the standard deviation. Typically, the mean pattern should follow an evolution related to the kind of economic relation described before, the kind of process (mean reverting process, long-term equilibrium convergence, steady-state, etc.) with its speed depending on the kind of variable considered. The standard deviation should follow the principle that forecasts related to a more distant point in time should be associated with a higher risk.

Figure 4 shows an example of the evolutionary pattern of the sales distribution function. A logistic pattern for sales is represented, highlighting how the evolution related to the launch of a new business line, typically characterized by three phases, can be stochastically modeled: a first one of slow growth when the new products/services enter the market; a second phase in which the demand grows exponentially and a final stage in which the market gets progressively saturated.

The evolutionary pattern can be developed at the industry level, analyzing the specific features of the sector, such as: barriers to entry, capital intensity, economies of scale, technology, market dimensions, consumer characteristics, product cycle, etc.

Figure 4. Sales evolutionary pattern: logistic growth

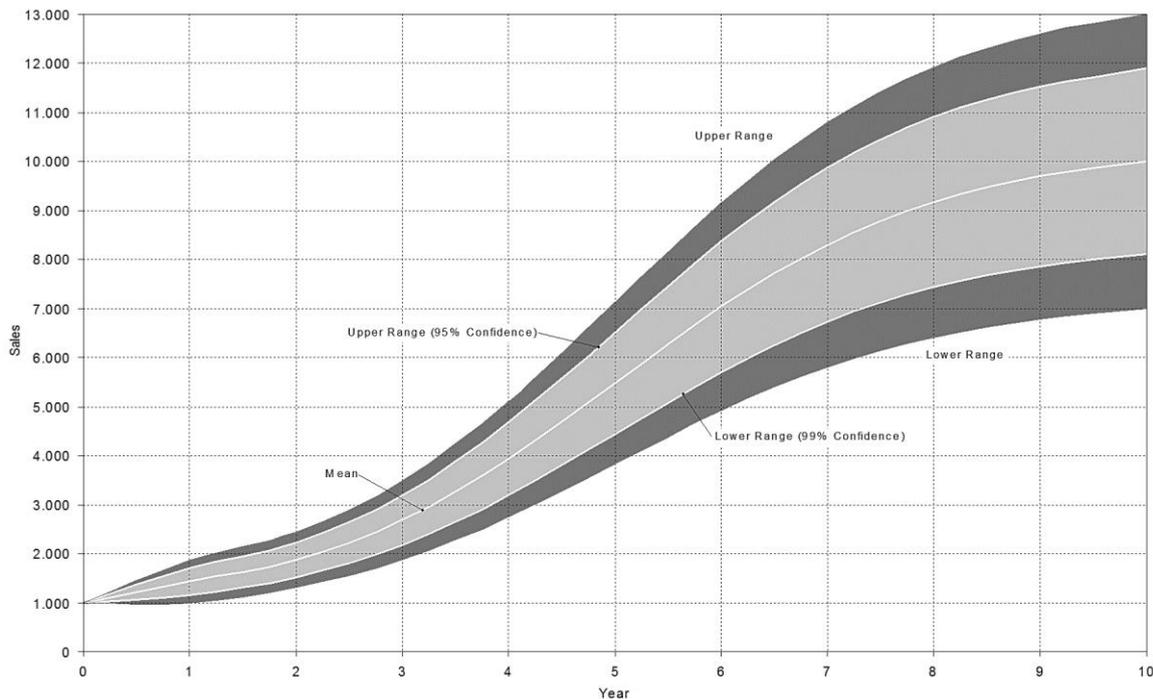
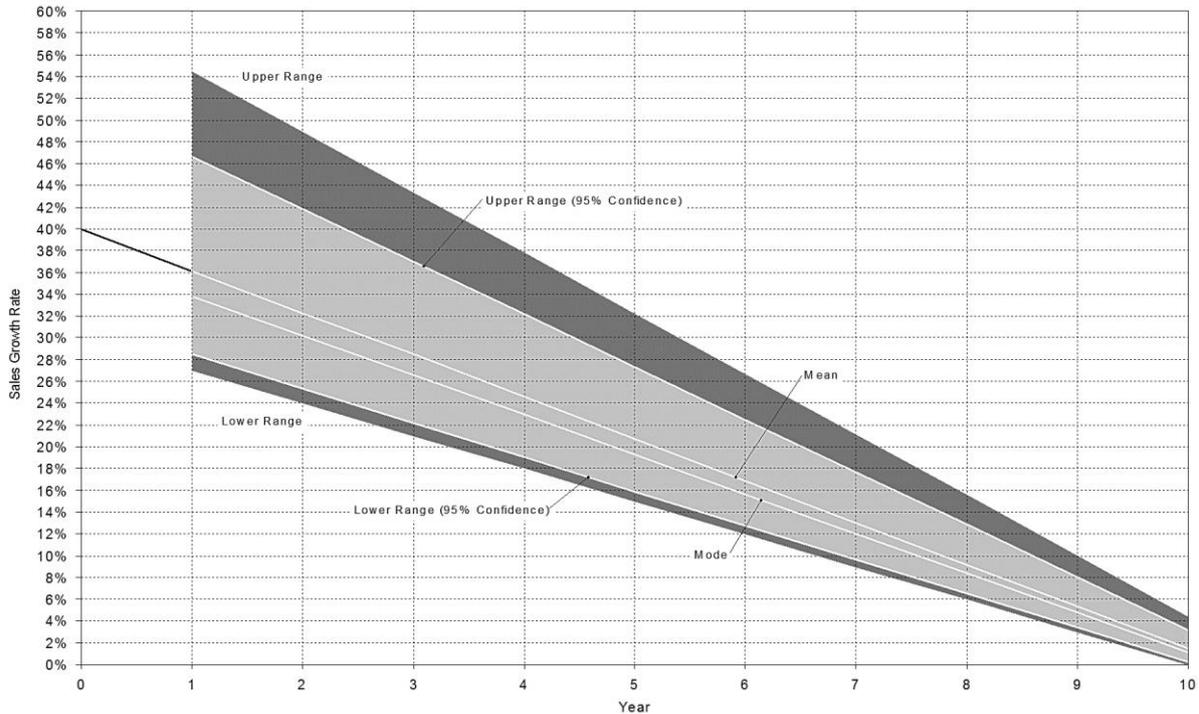


Figure 5 shows a different sales pattern modeling, in which a company currently experiments high growth, converges towards long-term sustainability, in line with the GDP growth rate. The

initial high potential variability of growth rate, characterized by a high mean in line with the latest records of the company, progressively converges towards the expected growth of the economy.

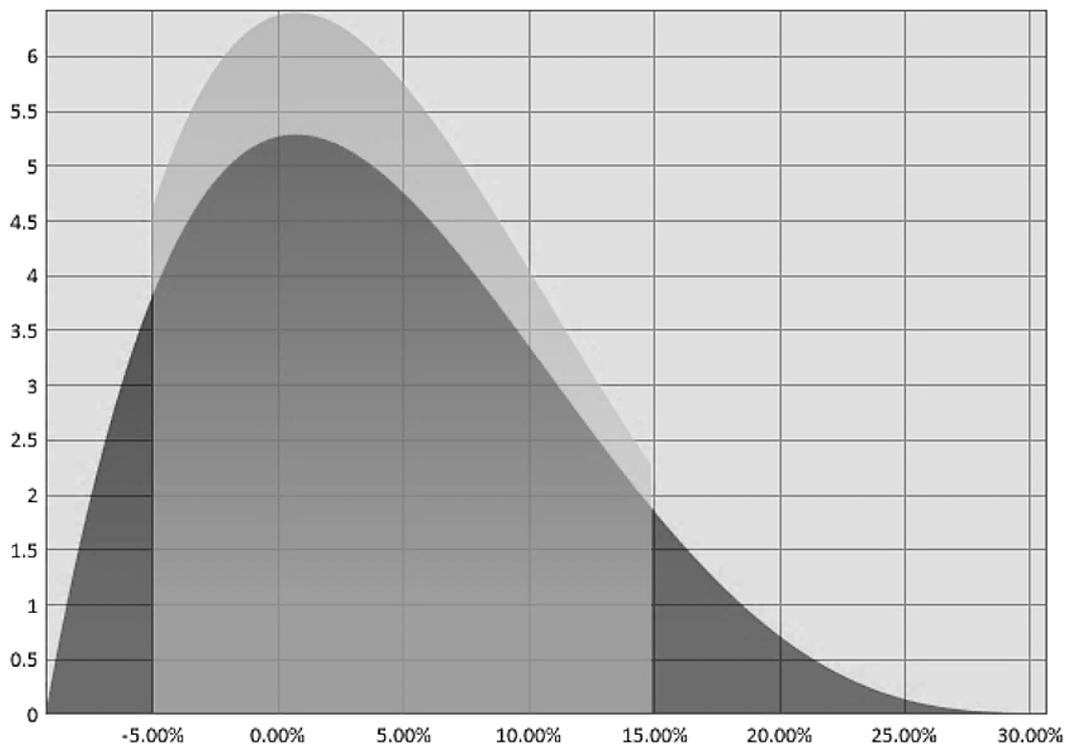
Figure 5. Sales evolutionary pattern: GDP long-term convergence



Within this evolutionary pattern more sophisticated features can be modeled, considering for instance cap and floor related to constraints

introduced by properly truncating the distribution functions, as shown in Figure 6.

Figure 6. Truncation of probability distribution



The assessment of an expected evolutionary pattern for the main value drivers described in the Appendix (sales, costs, capex, ... etc.) expresses the range of all the potential values that may occur for each driver during the business plan time. This pattern is functional to the definition of the stochastic variables and their probability distribution functions.

Making assumptions about the range of possible values can be easier than puzzling about a punctual forecast consistent with the traditional deterministic business planning approach. In this regard, the analysis of big data can provide useful information to assess and monitor the expected trend and variability of the key variables. Big data can be a valuable source of information. To draw practical indications for business planning purposes, they however need to be elaborated through sound financial models. In a context of uncertainty, these models should incorporate their informational contents into the economic forecasts of the company value drivers.

The forecasting information contents of big data may go beyond the short-term, but in this regard, they need constant refreshing. They can help in setting evolutionary patterns of the stochastic variables' distribution functions that are less strictly determined as an industry sector standard pattern and more company-specific.

Big data can tell a lot about the features of current clients of a company: geographic distribution, age, sex, employment, education, preferences, satisfaction, complaints, etc. All this information can substantially enrich the company's marketing data base and help to better define the company's "starting point", consistent with the cross-section data of Figure 2.

The attributes of the company's client base that are more sensitive to the company's product (industry sector) demand must be selected. They represent the best drivers of the sales. Comparing the company's client base with the average industry sector one, we can assess its prospective points of strength or weakness and try to understand how better or worse it may evolve respect to the average pattern assumed for the industry. In other words, we can assume if the company's sales evolutionary pattern should be steeper or flatter, wider or narrower if compared to the industry mean.

The CAP number of years can be considered as a stochastic variable that may lead the convergence process of the company value drivers.

While modeling the company-specific evolutionary pattern, we can also consider the impact of the actions and strategies envisaged by the company's business plan, like for instance, targeting different clusters of clients, addressing deficiencies in customer assistance, changing marketing policy, etc.

The analysis of big data within the stochastic simulation can also be useful in assessing the correlation among stochastic variables, helping to measure the value of the parameters of the Monte Carlo simulation correlation matrix.

The use of big data integrated with stochastic simulation needs to be developed and customized on an industry basis and requires preliminary extensive econometric analyses to assess the drivers of each product demand.

8. STOCHASTIC APPROACH AND BIG DATA APPLICATIONS

Big data analysis, combined with a stochastic simulation approach, can provide a powerful support to business planning in distinct stages:

- Ex-ante during the business planning process, to help the management to make the best investment decisions and strategies over a multi-annual period, through:

- a) short-term budgeting forecasts;
- b) medium-long-term forecasts by supporting the modeling of the evolutionary patterns of key stochastic variables and setting correlation parameters;

- Ex-post, during the execution of the business plan, monitoring the evolution of big data and continuously updating the short-term forecast to adjust and refocus budgeting targets.

Considering that for most industries and companies fixed costs present a low variability and that, as mentioned before, variable costs and capex can be expressed in the forecast model as a function of sales (see Appendix A), with a limited variability of their elasticity to sales, it follows that the sales modeling is pivotal in the set-up of the stochastic simulation.

Therefore, in showing how to use big data for stochastic simulation, we will focus on that key variable. For sectors in which it is important to model costs too, a similar approach may be applied. For example, this could be the case of companies for which raw material with a high price variability (oil, gold, etc.) have a high impact on the profitability margin.

The use of big data for sales modeling as a stochastic variable first requires an appropriate break-down of the variable into its main components, that can be considered as independent drivers to manage in building up the company's strategy.

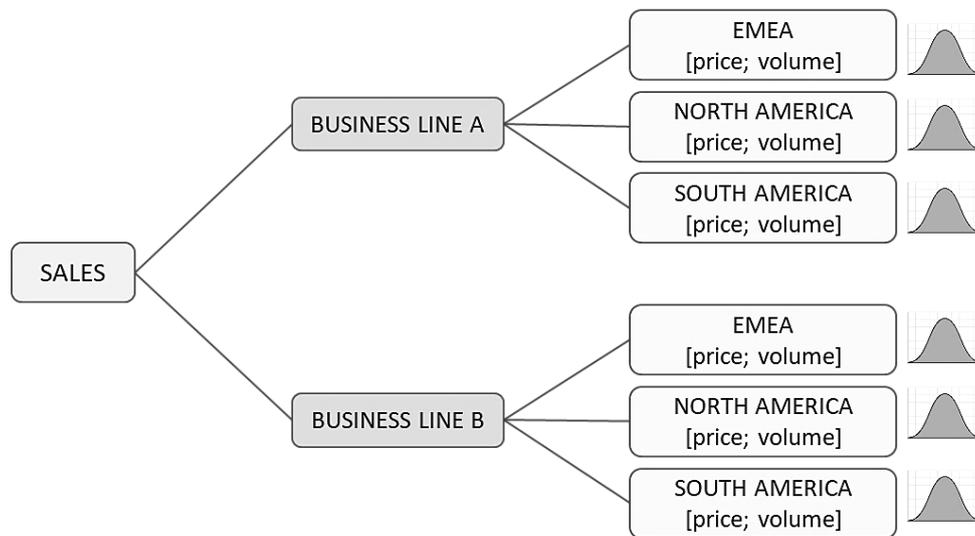
Demand forecasting is based on a new predictive analytics approach, and sophisticated information technology, which allows combining company data and economic information specifically matched to the market environment of product segments (Blackburn et al., 2015).

As an example, we may consider a company whose strategic target is to maximize the economic value added (EVA) generated within the business plan time horizon. For the sake of simplicity, we will consider only two product business lines (BL) (A, B) and three potential market areas: EMEA, North America, and South America. Let us also assume that the company cannot set the products prices (i.e., prices are determined by a competitive market) and that the production volume is essentially demand driven. In this context, we have for each sales cluster, given by a business line product and a market area, two uncertain variables, price and volume that determine the sales for that cluster and that can be managed as stochastic variables within the framework described. For each cluster, the price and volume variables will have different expected patterns, assessed through big data analyses that show ex-ante a different expected trend and risk (expected variability of the trend). By applying the stochastic approach to business planning, the management can find out which combinations of business line/market area break-down provide the

highest chances to reach the company EVA target; and thus, define the optimal company's strategy.

Figure 7 shows an example of sales ramification and prediction.

Figure 7. Sales ramification and prediction



More specifically, once the stochastic variables have been modeled, we can run several simulations, corresponding to different strategies characterized by different possible business line/market area combinations.

Then we can compare the cumulated EVA probability distributions and detect the strategy that ex-ante shows the highest probability to reach the target. The different distribution functions of sales clusters will provide different dynamics of the overall turnover of the company in the several simulations generated, according to the specific combination of the clusters assumed in the strategy (e.g., expand BL A in EMEA; reduce BL B in North America, etc.).

Following the economic relationships specified in the forecast model, all the other variables will be determined; and then elaborated to generate the projected income statements and balance sheets over the business plan period in each random scenario. Then by properly computing the company's projected

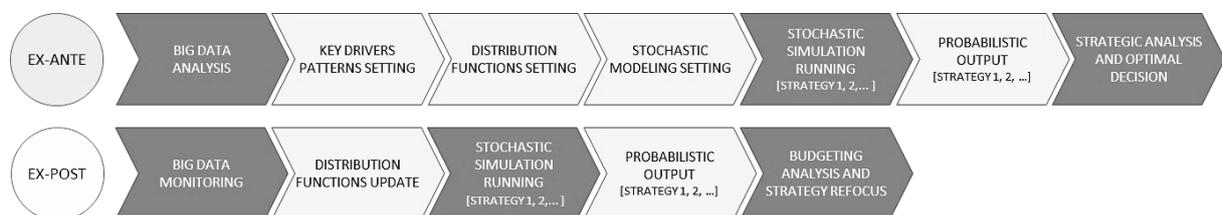
economics we can calculate EVA and, considering for each simulation all the random scenario simulated, its expected distribution functions for each simulation/strategy. By analyzing and comparing these distribution functions, the management can select the strategy that ex-ante is characterized by the highest probability to achieve the EVA target.

Subsequently to the definition of the multiannual strategic plan, during the periodical yearly (or interim) budget definition, the monitoring of the evolution of big data, jointly with an update of the stochastic simulation business planning approach, can help the management to adjust the sales break-down targets to refocus the strategy according to the additional available information.

The flexibility of the approach and its continuous re-elaboration allow to develop and put in place strategies with the highest chances of success.

The big data / stochastic analysis workflow is represented in Figure 8.

Figure 8. Big data / stochastic analysis workflow



9. REAL OPTIONS AND BOTTOM-UP BIG DATA FEEDBACKS

Figure 2 has shown that big data variability interacts with real options. A real option is the right - but not the obligation - to undertake certain business initiatives, such as deferring, abandoning, expanding, staging, or contracting a capital investment project. Real options are applied to capital budgeting decisions under uncertainty and are sensitive to the type of probability distribution (Peters, 2016) and

stochastic volatility (Huang et al., 2014).

Timely feedbacks from the real world are available through big data generation, capture, storing and processing.

The core concept is that if big data are continuously inserted in the business model, which can be instantaneously re-engineered, then real options are automatically embedded in business planning, which becomes more flexible and resilient. Real options may so reduce the volatility of the estimates, with a positive impact on business

planning and corporate evaluation. Lower volatility reduces the cost of capital (WACC), minimizing conflicts of interest among the stakeholders.

Real options describe the key tensions that managers face between commitment versus flexibility or between competition and cooperation (Trigeorgis & Reur, 2017). Corporate governance implications, concerning the composite relationships among stakeholders, evidently emerge and may be examined within a big data-driven stochastic pattern.

We may also consider more managerial flexibility in the pattern evolution illustrated in par. 6, reproducing somehow the effects of the real options eventually embedded in the company.

The flexibility embedded in the real options can be reproduced in a stochastic model based on Net Present Value (NPV), which incorporates big data-fed control variables and logic functions. Management-dependent variables can be foreseen according to expected outcomes and strategies can be fine-tuned following the stochastic evolution of key variables, sensitive for business planning and corporate valuation.

The advantage of considering real options with a stochastic simulation approach (compared to the traditional approach of separately evaluating each of them through options pricing models), lies in the fact that their effects in the overall company's projections context will be determined within a single forecasting model. Real options payoffs will be integrated with the dynamics of other variables (revenues, costs, investments), considering their interactions and feedback effects, and ultimately ensuring a higher level of consistency of the business plan projections.

In this context, the volatility of the option payoff is represented by the variability of the random scenarios generated through the stochastic process of the simulation. This approach allows to by-pass the problems of knowing in advance the volatility of the underlying asset or the shape of the distribution function of the option cash flows (so avoiding a priori simplistic normal symmetric assumptions); since its dynamics will be endogenously determined within the stochastic simulation model.

The stochastic approach to real options can be particularly helpful to assess the impact of multiple options, that are characterized by several sources of risks and are strongly state and path dependent; in fact, in these cases traditional option pricing models are inadequate and would require exotic pricing models that turn out to be quite complex and subject to many simplistic assumptions.

For example, a company's real option given by the flexibility of an easily switching business line production (according to the one that is most promising in each moment) might be considered in the stochastic approach by assigning a proper (negative) correlation to the stochastic variables that represent the business lines sales involved in the real option; the overall effect on the evolution of the company's total revenues will be to stabilize them. Similarly, a real option related to the possibility of easily scaling the production of new business lines (in case of success) can be modeled introducing a positive autocorrelation coefficient (temporal correlation between one year and the other) in the stochastic variable that represents the business line sales.

A real option related to flexibility in running off

an unsuccessful business line might be considered within the forecasting model by properly connecting the variable costs to the stochastic variables that determine the production level of that business line in the model (see the Appendix).

The number of real options to consider and the level of modeling sophistication will depend on their relevance in the cases under analysis. As a rule, all models should be taken at the simplest level of complexity compatible with the scope of the analysis. Only highly relevant real options capable of granting an exclusive competitive advantage, should be considered and modeled in, capable of granting an exclusive competitive advantage.

Flexibility embedded in real options can have several corporate governance implications, softening the conflicts of interest among composite stakeholders that are nurtured by rigid planning and inherent higher risk.

10. DISCUSSION

The model synthetically described has important corporate governance implications that depend on its criticalities and potential outcomes. An analysis of the findings of this study is preparatory to the corporate governance implications.

10. 1. Critical analysis of the findings

When the size of data to the process is big, conventional manipulation tools based on standard statistical and econometric techniques need to be complemented by machine learning techniques as decision trees, support vector machines, neural networks, and deep learning (Varian, 2014).

To improve forecasts, we need to go back to theory and modeling. The necessity to formulate long-term forecasts raises an issue of economic coherence between the hypotheses of the model and the theoretical principles of economics and finance.

For example, by exploiting consumers' comments on the social media, it is possible to "learn" the relationship between a forerunner key indicator of the sales dynamic in a specific product; in this way, by observing the evolution of the indicator it is possible to achieve accurate forecasts for a key variable such as sales.

The revenue model is possibly the biggest critical factor in business planning. Profits are hard to forecast, and they are strongly linked to the corporate evaluation methods, driven by profit-sensitive parameters as EBITDA or operating cash flows.

The revenue model represents the pivotal element of business planning. The estimate of sales growth (Gupta et al., 2013) is so central and can be improved both with big data and with stochastic modelling.

Sales can be subdivided into their core elements, as:

- volumes and prices;
- typologies of products or services;
- geographical markets;
- trends driven by organic growth or external acquisitions.

Stochastic variables of revenues are to be split into their strategic sections (business line, product, geographic area, etc.).

Mark-to-market feedbacks of the revenues ease the constant re-engineering and updating of the stochastic forecasts. The goal may be represented by machine learning processes where big data are instantaneously captured and processed, igniting artificial intelligence patterns that automatically refresh predictions.

Evolutionary patterns of the segmented revenues need to be examined following big data informative drivers together with their stochastic trends (probability assumptions, correlations, etc.).

Big data can improve granularity of sales, whereas stochastic modelling represents a forecast pattern that estimates probability scenarios within an interval of confidence.

The real options issue has just been mentioned in par. 9 and deserves further analysis. Theoretical and empirical research should focus on the merging process between real options and big data, with their biunivocal relationship. In other terms, it should be clarified if big data trigger real options or if options fuel data. Options may be conveniently exercised when their underlying value is properly known, and from this perspective data influence options. Conversely, the very existence of a valuable real option may convey input information that fuels big data.

Once the causal relationship between data and options is settled, further analysis should be dedicated to their impact on the budgeting process in a stochastic scenario. A theoretical link, again to be analyzed in deeper terms, may be represented by the circumstance that both options and stochastic processes may follow binomial patterns.

Big data have interesting pros, but also some drawbacks. Despite the enthusiasm that they have generated regarding the forecasting issues, data cannot eliminate uncertainty, regardless of the quantity and quality of the available information.

Forecasts may be improved, and the risk may be reduced within a reasonable probabilistic range, using a broader set of information, but perfect foresight remain a chimera. The idea of developing a universal Turing learning machine or Master Algorithm just by properly feeding it with the necessary amount of data, is, according to Hosni & Vulpiani (2017) conceptually wrong and illusory.

Moreover, even if the large availability of data can easily bring out new significant statistical relationships among data time series, assessing a correlation between two variables does not mean to have discovered a causal relationship among them.

Excess of data (especially if unchecked, and potentially referable to fake news) may sometimes be misleading. The forecasting process and a consistent use of data should be adapted to the specific circumstances under analysis and particularly to the time horizon of the forecast.

10. 2. Corporate governance implications

Corporate governance issues traditionally refer to balance sheet accounts that reproduce an instantaneous “screenshot”, giving past evidence of the book value of equity and debt and of the asset composition. Little if any evidence is typically dedicated to dynamic estimates where the value is forecast over a consistent horizon. Estimates are however crucial to forecast conflicts of interest

among stakeholders that can so be minimized. The impact of big data on corporate governance is analyzed in Moro Visconti (2017).

The main finding of this study is that estimates can be improved combining stochastic methodologies with big data that can be helpful especially in the updating and reformulation of short-term forecasts. The risk may so be reduced.

Any forecast of market value scenarios, consistent with a stochastic approach, can greatly contribute to improving the knowledge of corporate governance concerns. And big data add depth, and immediacy to forecasting, as evidenced in Table 1.

Estimation of value and free cash flows (Richardson, 2006) is a central corporate governance issue, since the seminal study of Jensen (1986). Managerial discretion and remuneration policies, information asymmetries, investment policies, or debt servicing capacity are just some of the main governance implications of budgeting.

Financial accounting information is critical for corporate control mechanisms (Bushman & Smith, 2001), whereas financial reporting transparency reduces governance-related agency conflicts among managers, directors, and shareholders, as well as conflicts between shareholders and creditors (Armstrong et al., 2010).

The link between free cash flow and sales is investigated in Brush et al. (2000), whereas Dittmar & Mahrt-Smith (2007) analyze the governance-driven impact of cash holdings on future operating performance. Over-investment is concentrated in firms with the highest levels of free cash flow (Richardson, 2006).

Conversion from book to market value is typically actuated with corporate evaluation methodologies, as those described in par. 4., and is used for many applications like the calculation of the market-driven cost of capital.

10. 3. New research avenues

New research avenues should consider the feasibility of machine learning techniques, where artificial intelligence dominates the process that automatically collects data to generate increasingly accurate forecasts.

Another research avenue is represented by the impact of big data and stochastic estimates on strategic management and optimal project selection (Meredith & Mantel, 2015), related to capital budgeting, a classic branch of corporate finance. Analysis of uncertainty, the management of risk and consequent conflicts are typical of any project activity planning where budgeting, scheduling and cost estimation drive resource allocation strategies. Any improvement in the budgeting process can have important benefits for the stakeholders, softening corporate governance criticalities.

Patterns of growth represent the engine of capital budgeting and company forecasting. While some suggest that the growth path followed by the enterprise is linear or predictable, others claim that the growth is opportunistic or unpredictable (Gupta et al., 2013). Growth unpredictability, increasingly related to unconventional firms with innovative business plans, can be better described once again with stochastic models fed in real time by big data and real options. This represents a further research

theme, with evident governance implications.

A further research stream can be represented by artificial intelligence and machine learning techniques, fueling a process that automatically collects big data to generate increasingly accurate forecasts in a stochastic environment. This represents the goal of optimal modeling, where business planning and inherent corporate evaluation may be dynamically inserted.

11. CONCLUSION

This study has shown that estimates for business planning and corporate appraisals can be improved combining stochastic methodologies with big data that can be helpful especially in the continuous updating and reformulation of short-term forecasts. The risk, so important for corporate governance concerns, may so be reduced.

Financial accounting information is critical for corporate control mechanisms, whereas financial reporting transparency and accurateness reduce governance-related agency conflicts among corporate stakeholders (shareholders, debtholders, managers, directors, employees, etc.).

Corporate evaluation methodologies normally convert book input data to market value indicators and are mainly based on discounted cash flow patterns or market multipliers of the Enterprise Value/EBITDA. These well-established methods may conveniently be complemented with big data and stochastic estimates.

This is the first study, as far as the authors are concerned, addressing corporate evaluation issues, so crucial for corporate governance concerns, using a combination of big data and stochastic estimates, and so introducing a process innovation.

REFERENCES

1. Al-Barznji, K., & Atanassov, A. (2017). Review of big data and big data mining for adding big value to enterprises. *Science, Engineering & Education*, 2(1), 50-57.
2. Archak, N., Ghose, A., & Panagiotis Ipeirotis, G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8), 1485-1509. <https://doi.org/10.1287/mnsc.1110.1370>
3. Armstrong, C. S., Guaya, W. R., & Weberb, J. P. (2010). The role of information and financial reporting in corporate governance and debt contracting. *Journal of Accounting and Economics*, 50(2-3), 179-234. <https://doi.org/10.1016/j.jacceco.2010.10.001>
4. Bancel, F., Mittoo, U. R. (2014). The gap between the theory and practice of corporate valuation: Survey of european experts. *Journal of Applied Corporate Finance*, 26(4), 106-117. <https://doi.org/10.1111/jacf.12095>
5. Blackburn, R., Lurz, A., Priese, B., Göb, R., & Darkow, I. (2015). A predictive analytics approach for demand forecasting in the process industry. *International Transactions in Operational Research*, 22(3), 407-428. <https://doi.org/10.1111/itor.12122>
6. Brush, T. H., Bromiley, P., & Hendrickx, M. (2000). The free cash flow hypothesis for sales growth and firm performance. *Strategic Management Journal*, 21(4), 455-472. [https://doi.org/10.1002/\(SICI\)1097-0266\(200004\)21:4<455::AID-SMJ83>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0266(200004)21:4<455::AID-SMJ83>3.0.CO;2-P)
7. Bushmana, R. M., & Smith, A. J. (2001). Financial accounting information and corporate governance. *Journal of Accounting and Economics*, 32(1-3), 237-333. [https://doi.org/10.1016/S0165-4101\(01\)00027-1](https://doi.org/10.1016/S0165-4101(01)00027-1)
8. Cario, M. C., & Nelson, B. L. (1997). *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix*. Evanston, Illinois, US: Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University.
9. Chatfield, C. (2003). *The analysis of time series: An introduction* (6th ed.). Boca Raton, Florida: Chapman & Hall/CRC Press.
10. Chen, Y., Sun E. W., & Lin, Y. (2018). Coherent quality management for big data systems: A dynamic approach for stochastic time consistency. *Annals of Operations Research*, 1-30. <https://doi.org/10.1007/s10479-018-2795-1>
11. Daas, P. J. H., Puts, M. J., Buelens, B., & van den Hurk, P. A. M. (2015). Big data as a source for official statistics. *Journal of Official Statistics*, 31(2), 249-262. <https://doi.org/10.1515/jos-2015-0016>
12. Dittmara, A., & Mahrt-Smith, J. (2007). Corporate governance and the value of cash holdings. *Journal of Financial Economics*, 83(3), 599-634. <https://doi.org/10.1016/j.jfineco.2005.12.006>
13. Duan, L., & Xiong, Y. (2015). Big data analytics and business analytics. *Journal of Management Analytics*, 2(1), 1-21. <https://doi.org/10.1080/23270012.2015.1020891>
14. Franks, B. (2014). *The analytics revolution: How to improve your business by making analytics operational in the big data era*. Hoboken, New Jersey: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118936672>
15. Gupta, P. D., Guha, S., & Krishnaswami, S. S. (2013). Firm growth and its determinants. *Journal of Innovation and Entrepreneurship*, 2(1), 15. <https://doi.org/10.1186/2192-5372-2-15>
16. Hartmann, C., Hahmann, M., Lehner, W., & Rosenthal, F. (2015). Exploiting big data in time series forecasting: A cross-sectional approach. Paper presented at *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. <https://doi.org/10.1109/DSAA.2015.7344786>
17. Hazen, B., Boone, C., Ezell, J., & Jones-Farmer, J. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154, 72-80. <https://doi.org/10.1016/j.ijpe.2014.04.018>
18. Hosni, H., & Vulpiani, A. (2017). Forecasting in light of big data. *Philosophy & Technology*, 1-13. <https://doi.org/10.1007/s13347-017-0265-3>
19. Huang, B., Cao, J., & Chung, H. (2014). Strategic real options with stochastic volatility in a duopoly model. *Chaos, Solitons & Fractals*, 58, 40-51. <https://doi.org/10.1016/j.chaos.2013.11.005>
20. Jensen, M.C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review*, 76(2), 323-329.
21. Jin, X, Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research*, 2(2), 59-64. <https://doi.org/10.1016/j.bdr.2015.01.006>

22. Jumi, K, Wookey, L., & Kwan-Hee, Y. (2016). Business driving force models for big data environment. Paper presented at the *International Conference on Big Data and Smart Computing*. <https://doi.org/10.1109/BIGCOMP.2016.7425928>
23. Karim, L., Boulmakoul, A., Mabrouk, A., & Lbath, A., (2017). Deploying real time big data analytics in cloud ecosystem for hazmat stochastic risk trajectories. *Procedia Computer Science*, 109, 180-187. <https://doi.org/10.1016/j.procs.2017.05.322>
24. Kostyuk, A., Mozghovyi, Y., Govorun, D. (2018). Corporate governance, ownership and control: A review of recent scholarly research. *Corporate Board. Role, Duties and Composition*, 14(1), 50-56. <https://doi.org/10.22495/cbv14i1art4>
25. Latif, K., Bhatti, A. A., & Raheman, A. (2017). Earnings quality: A missing link between corporate governance and firm value. *Business & Economic Review*, 9(2), 255-280. <https://doi.org/10.22547/BER/9.2.11>
26. Lau, R. Y. K., Zhang, W., & Xu, W. (2018). Parallel aspect-oriented sentiment analysis for sales forecasting with big data. *Production and Operations Management (forthcoming)*. <https://doi.org/10.1111/poms.12737>
27. Mauboussin, M. J., & Johnson, P. (1997). Competitive advantage period: The neglected value driver. *Financial Management*, 26, 67-74.
28. Mauboussin M. J. (2006). *More than you know: finding financial wisdom in unconventional places*. New York: Columbia University Press.
29. Meredith, J. R., & Mantel, S. J. (2015). *Project management: A managerial approach* (9th ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.
30. Montesi, G., Papiro, G. (2014). Risk analysis probability of default: A stochastic simulation model. *Journal of Credit Risk*, 10(3), 29-86.
31. Moro Visconti, R. (2017). Public private partnerships, big data networks and mitigation of information asymmetries. *Corporate Ownership & Control*, 14(4-1), 205-215. <https://doi.org/10.22495/cocv14i4c1art3>
32. Moro Visconti, R., Larocca, A., & Marconi, M. (2017). *Big data-driven value chains and digital platforms: From Value Co-Creation to Monetization*. Retrieved from the World Wide Web: <https://ssrn.com/abstract=2903799>
33. Somani, A. K., & Deka, G. C. (2017). *Big data analytics: Tools, technology for effective planning* (book chapter). Boca Raton: Chapman and Hall/CRC Press.
34. Mundlak, Y. (1978). On the pooling of time series and cross-section data. *Econometrica*, 46(1), 69-85. <https://doi.org/10.2307/1913646>
35. Nelsen, R. (2006). *An introduction to copulas*. Heidelberg, Berlin: Springer-Verlag GmbH.
36. Nijeholt, M. L., & Grift, Y. (2007). *Goodwill, excess returns, and determinants of value creation and overpayment, January* (Working paper 07-31, University of Utrecht).
37. Olson, D. L., & Wu, D. D. (2017). Data mining models and enterprise risk management. In D. L. Olson & D. D. Wu (Ed.), *Enterprise risk management models* (pp. 119-132). Heidelberg, Germany: Springer. <https://doi.org/10.1007/978-3-662-53785-5>
38. Peters, L. (2016). Impact of probability distributions on real options valuation. *Journal of Infrastructure Systems*, 22(3). [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000289](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000289)
39. Richardson, S. (2006). Over-investment of free cash flow. *Review of Accounting Studies*, 11(2-3), 159-189. <https://doi.org/10.1007/s11142-006-9012-1>
40. Robert, C. P., & Casella, G. (2004). *Monte Carlo statistical methods* (2nd ed.). New York: Springer Publishing Company.
41. Rubinstein, R. Y., & Kroese, D. P. (2017). *Simulation and the Monte Carlo method*. Hoboken, New Jersey: John Wiley & Sons, Inc.
42. Sonnier, G. P., McAlister, L., & Rutz O. J. (2011). A dynamic model of the effect of online communications on firm sales. *Marketing Science*, 30(4), 702-716. <https://doi.org/10.1287/mksc.1110.0642>
43. Tallon, P. P. (2013). Corporate governance of big data: Perspectives on value, risk, and cost. *Computer*, 46(6), 32-38. <https://doi.org/10.1109/MC.2013.155>
44. Tanner, J. F. (2014). *Analytics and dynamic customer strategy: Big profits from big data*. Hoboken, New Jersey: John Wiley & Sons, Inc. Retrieved from the World Wide Web: <https://doi.org/10.1002/9781118919767>
45. Trigeorgis, J., & Reuer, J. J. (2017). Real options theory in strategic management. *Strategic Management Journal*, 38(1), 42-63. <https://doi.org/10.1002/smj.2593>
46. Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3-28. <https://doi.org/10.1257/jep.28.2.3>
47. Vasarhelyi, M. A., Kogan, A., & Tuttle, B.M. (2015). Big data in accounting: An overview. *Accounting Horizons*, 29(2), 381-396. <https://doi.org/10.2308/acch-51071>
48. Vernimmen, P., Quiry, P., Dallochio, M., Le Fur, Y., & Salvi, A. (2014). *Corporate finance: Theory and practice* (4th ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.
49. Walker, R. (2015). *From big data to big profits*. Oxford, UK: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199378326.001.0001>
50. Wong, W. K. (2007). Stochastic dominance and mean-variance measures of profit and loss for business planning and investment. *European Journal of Operational Research*, 182(2), 829-843. <https://doi.org/10.1016/j.ejor.2006.09.032>
51. Xiaohui, Y., Yang, L., Huang, J. X., & Aijun, A. (2012). Mining online reviews for predicting sales performance: A case study in the movie domain. *Ieee Transactions on Knowledge and Data Engineering*, 24(4), 720-734.
52. Xu, J., Huang, E., Chen, C., & Lee, L. H. (2015). Simulation optimization: A review and exploration in the new era of cloud computing and big data. *Asia-Pacific Journal of Operational Research*, 32(3), 1-34. <https://doi.org/10.1142/S0217595915500190>
53. Xu, X., & Birge, J. R. (2006). Equity valuation, production, and financial planning: A stochastic programming approach. *Naval Research Logistics*, 53(7), 641-655. <https://doi.org/10.1002/nav.20182>
54. Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research*, 69(5), 1562-1566. <https://doi.org/10.1016/j.jbusres.2015.10.017>

APPENDIX

Sales-driven forecasting model

A sales-driven forecasting model is often the best kind of modeling for business plan projections. The set-up of such a model requires the definition of functional relations between sales and the main company's value drivers. There are important economic factors, such as economies of scale that tie the level of production to all the most important value drivers; in fact, sales affect not only operating costs, through the operating leverage, but also working capital and capex. Think about the *learning by doing* effect, which represents the main scale economy factor in many industries.

Within a stochastic simulation model, to ensure consistency in the generation of random scenarios, it is necessary to functionally relate the key variables to the most important and uncertain variable: sales. Two

simple modeling for relating operating cost and capex to sales are reported below.

A.1. Performing operating leverage

A straightforward way to model operating leverage is to assume that operating costs depend in part (θ) on the cost in the previous period and in part ($1-\theta$) on sales. In formal terms:

$$\text{Operating Cost}_t = \theta \cdot \text{Operating Cost}_{t-1} + (1-\theta) \cdot (1-s_t)$$

Where θ is included in the range [0,1] and s_t is the sales rate of change.

By dividing all by sales and through some simple transformations we obtain the following relation:

$$\frac{\text{Operating Cost}_t}{\text{Sales}_t} = \frac{\theta}{1-s_t} \cdot \frac{\text{Operating Cost}_{t-1}}{\text{Sales}_{t-1}} + \frac{(1-\theta) \cdot \text{Operating Cost}_{t-1}}{\text{Sales}_{t-1}}$$

Or:

$$\frac{\text{Operating Cost}_t}{\text{Sales}_t} = \frac{\text{Operating Cost}_{t-1}}{\text{Sales}_{t-1}} \cdot \frac{1+s_t \cdot (1-\theta)}{1+s_t}$$

If $\theta=0$ then operating cost are assumed to be totally variable and can be projected as a constant proportion of sales. If $\theta=1$ then operating cost are assumed as totally fixed (independently from sales) and projected as a constant value (in that case an increase/reduction in sales implies a reduction/increase in the operating cost/sales ratio).

Within the model θ can be managed as a deterministic or stochastic variable if the sales/cost relation is uncertain or unstable, modeling operating leverage dynamics. The forecast assumptions of this parameter can be based on the company's and/or industry sector historical records, through a back-testing optimization process aimed at minimizing the error between forecast achieved and effective historical records.

A.2. Capex = f (sales)

The relation between the effective expected capital expenditures, or Capex, and the sales can be set by first modeling net fixed asset (NFA) as a function of sales. We can define the investment function as a relation between NFA and the sales growth, through an elasticity coefficient; in formal terms:

$$NFA_t = NFA_{t-1} \left(\frac{\text{Sales}_t}{\text{Sales}_{t-1}} \right)^b$$

Where b is the fixed assets elasticity coefficient respect to sales growth. If $b=1$ then NFA are assumed to have a constant ratio with sales, meaning that the investment growth rate matches the sales growth rate. If $b > 1$ then the investment growth rate is assumed to be higher than the sales growth; similarly, if $b < 1$ the investment growth rate is assumed to be lower than the sales growth, showing economies of scale.

Once the value of NFA has been determined, the CAPEX can be obtained by considering that the NFA at time t can also be expressed as:

$$NFA_t = NFA_{t-1} + CAPEX_t - DA_t$$

where DA is the value of depreciation and amortization. If the CAPEX is gradually carried out through the year, DA can be defined as:

$$DA_t = \alpha_t \left(GFA_{t-1} - RETIR_t + \frac{CAPEX_t}{2} \right)$$

where α is the expected rate of depreciation and amortization, GFA are the gross fixed assets, $RETIR$ is the value of retirements (i.e., the share of fixed assets fully depreciated that must be deducted from GFA every year). Swapping equation [III] into [I] we can obtain the Capex as:

$$CAPEX_t = \frac{2 \cdot (NFA_t - NFA_{t-1} + \alpha_t \cdot GFA_{t-1} - \alpha_t \cdot RETIR_t)}{2 - \alpha_t}$$

$RETIR$ can be determined in several ways. One way is through a closed formula; for example, if NFA in each period is equal to the sum of all the future depreciations necessary to fully depreciate the outstanding GFA :

$$NFA_t = \sum_{i=1}^n \alpha \cdot \left(GFA_t - i \cdot \frac{GFA_t}{n} \right)$$

And resolving the CAPEX equation to obtain n , we can then determine $RETIR$ as GFA divided by the number of years necessary to its full depreciation:

$$RETIR_t = \frac{GFA_t}{n}$$

Another way to determine $RETIR$ is to multiply the retirement rate (r) by the GFA of the previous period, recurring to an historical estimate of the retirement rate:

$$RETIR_t = GFA_{t-1} \cdot r_t$$

Considering that, in the absence of disinvestments, retirements in each period are equal to the value of depreciation and amortization minus the change in accumulated depreciation, estimates of r can be obtained dividing the historical values of this difference by GFA values.