Machine Learning for the Measurement of Firm Value in Investment Banking: A Parametric Approach Based on Integral Equations

New York General Group
Nov. 2023

Abstract

In this paper, we propose a novel method for measuring the value of firms in investment banking using machine learning techniques and integral equations. We assume that the value of a firm is determined by its expected future cash flows, which depend on various macroeconomic and microeconomic factors. We use integral equations to model the relationship between these factors and the cash flows, and we employ machine learning algorithms to estimate the unknown parameters of the integral equations from historical data. We apply our method to a sample of publicly traded firms in the US and compare the results with the traditional discounted cash flow (DCF) method. We find that our method can capture the nonlinear and dynamic effects of the factors on the firm value, and can provide more accurate and robust estimates than the DCF method.

I. Introduction

The measurement of firm value is one of the most important tasks in investment banking, as it affects various decisions such as mergers and acquisitions, capital structure, dividend policy, and valuation of securities. However, the measurement of firm value is also challenging, as it requires forecasting the future cash flows of the firm, which are subject to uncertainty and volatility. Moreover, the future cash flows depend on a number of macroeconomic and microeconomic factors, such as interest rates, inflation, exchange rates, industry conditions, market competition, consumer preferences, technological innovation, and regulatory environment. These factors may have complex and nonlinear effects on the cash flows, and may also interact with each other over time.

One of the most widely used methods for measuring firm value is the discounted cash flow (DCF) method, which discounts the expected future cash flows of the firm by a constant discount rate, usually the weighted average cost of capital (WACC). The DCF method has the advantage of being simple and intuitive, but it also has several limitations. First, the DCF method assumes that the discount rate is constant and independent of the factors that affect the cash flows, which may not be realistic in practice. Second, the DCF method relies on the assumption that the cash flows follow a deterministic and stationary process, which may not capture the stochastic and dynamic nature of the cash flows. Third, the DCF method requires specifying the terminal value of the firm, which is often based on arbitrary assumptions or rules of thumb, such as the perpetual growth model or the exit multiple method.

In this paper, we propose a novel method for measuring firm value that overcomes the limitations of the DCF method. Our method is based on the use of integral equations and machine learning techniques. We assume that the value of a firm is determined by its expected future cash flows, which depend on various macroeconomic and microeconomic factors. We use integral equations to model the relationship between these factors and the cash flows, and we employ machine learning algorithms to estimate the unknown parameters of the integral equations from historical data. We apply our method to a sample of publicly traded firms in the US and compare the results with the DCF method.

The main contributions of our paper are as follows:

- We introduce a parametric approach based on integral equations for measuring firm value, which can capture the nonlinear and dynamic effects of the factors on the cash flows, and can also account for the uncertainty and volatility of the cash flows.
- We use machine learning techniques, such as neural networks and gradient boosting, to estimate the parameters of the integral equations from historical data, which can improve the accuracy and robustness of the estimates, and can also handle high-dimensional and complex data.
- We demonstrate the applicability and effectiveness of our method by applying it to a sample of publicly traded firms in the US, and we show that our method can provide more accurate and consistent estimates of firm value than the DCF method.

The rest of the paper is organized as follows. In section 2, we present the theoretical framework of our method, which consists of the integral equations model and the machine learning algorithms. In section 3, we describe the data and the empirical implementation of our method. In section 4, we present and discuss the results of our method and compare them with the DCF method. In section 5, we conclude and offer some directions for future research.

III. Theoretical Framework

In this section, we present the theoretical framework of our method, which consists of two main components: the integral equations model and the machine learning algorithms. We first describe the integral equations model, which is a general and flexible way of modeling the relationship between the factors and the cash flows. We then introduce the machine learning algorithms, which are used to estimate the parameters of the integral equations model from historical data.

II.I. Integral Equations Model: We assume that the value of a firm at time t, denoted by V(t), is equal to the expected present value of its future cash flows, denoted by C(t). That is,

$$V(t) = E_t[C(t)]$$

where E_t denotes the conditional expectation operator given the information available at time t. We further assume that the cash flows depend on a vector of macroeconomic and microeconomic factors, denoted by X(t), which may include interest rates, inflation, exchange rates, industry conditions, market competition, consumer preferences, technological innovation, and regulatory environment. We use an integral equation to model the relationship between the cash flows and the factors, as follows:

$$C(t) = \int_t^{\infty} f(X(s), s)\,ds$$

Machine Learning for the Measurement of Firm Value in Investment Banking: A Parametric Approach Based on Integral Equations

where f is an unknown function that represents the impact of the factors on the cash flows. The integral equation implies that the cash flows at time t are equal to the sum of the discounted impacts of the factors over the infinite horizon. The advantage of using an integral equation is that it can capture the nonlinear and dynamic effects of the factors on the cash flows, and it can also account for the uncertainty and volatility of the cash flows. Moreover, the integral equation can be easily generalized to incorporate different assumptions or specifications, such as different discount rates, different functional forms, different time horizons, or different stochastic processes.

The main challenge of using the integral equation model is that the function f is unknown and needs to be estimated from historical data. However, the estimation of f is not straightforward, as it involves solving an inverse problem, which is typically ill-posed and ill-conditioned. That is, the solution of f may not exist, may not be unique, or may be unstable. Therefore, we need to use some regularization techniques to obtain a stable and reliable estimate of f. In the next subsection, we introduce the machine learning algorithms that we use for this purpose.

II.II. Machine Learning Algorithms: Machine learning is a branch of artificial intelligence that deals with the design and analysis of algorithms that can learn from data and make predictions or decisions. Machine learning algorithms have been widely used in various fields and applications, such as computer vision, natural language processing, recommender systems, and bioinformatics. In this paper, we use machine learning algorithms to estimate the function f in the integral equation model from historical data. We use two types of machine learning algorithms: neural networks and gradient boosting.

Neural networks are computational models that consist of multiple layers of interconnected nodes, or neurons, that can process and transmit information. Neural networks can learn complex and nonlinear patterns from data by adjusting the weights and biases of the connections between the nodes. Neural networks have been shown to be universal function approximators, meaning that they can approximate any continuous function to any desired degree of accuracy, given enough nodes and data. Neural networks have also been successfully applied to various inverse problems, such as image reconstruction, signal processing, and parameter estimation.

Gradient boosting is a technique that combines multiple weak learners, or base models, into a strong learner, or ensemble model, by iteratively adding new base models that correct the errors of the previous ones. Gradient boosting can handle high-dimensional and heterogeneous data, and can also deal with missing values, outliers, and noise. Gradient boosting can also learn nonlinear and nonparametric functions from data, and can also incorporate prior knowledge or domain expertise into the learning process.

We use neural networks and gradient boosting to estimate the function f in the integral equation model as follows. We first divide the historical data into a training set and a test set. We then use the training set to train the neural network and the gradient boosting models, and use the test set to evaluate their performance. We use cross-validation and grid search to select the optimal hyperparameters of the models, such as the number of layers, the number of nodes, the activation function, the learning rate, the number of base models, and the loss function. We then compare the neural network and the gradient boosting models based on their accuracy, robustness, and interpretability, and select the best model for estimating the function f. We then use the estimated function f to compute the cash flows and the firm value for the sample of firms.

III. Data and Empirical Implementation

In this section, we describe the data and the empirical implementation of our method. We first explain the data sources and the sample selection criteria for the firms and the factors. We then describe the steps involved in estimating the function f in the integral equation model using the machine learning algorithms.

III.I. Data Sources and Sample Selection: We use two main sources of data for our analysis: one for the firm cash flows and value, and one for the macroeconomic and microeconomic factors. For the firm cash flows and value, we use the Compustat database, which provides financial information for publicly traded companies in the US. We obtain the data on earnings before interest and taxes (EBIT), depreciation, capital expenditure, working capital, and other assets from the income statement and the balance sheet of the firms. We also obtain the data on the market value of equity and the book value of debt from the Compustat database. We use these data to compute the operating free cash flow (OFCF) and the firm value (V) for each firm in each year, following the formulas described in section 2.1.

For the macroeconomic and microeconomic factors, we use various online data sources, such as the Federal Reserve Economic Data (FRED), the Bureau of Economic Analysis (BEA), the International Monetary Fund (IMF), and the World Bank. We obtain the data on interest rates, inflation, exchange rates, industry conditions, market competition, consumer preferences, technological innovation, and regulatory environment from these sources. We use these data to construct a vector of factors (X) for each year, following the definitions and measurements described in section 2.1.

We restrict our sample to firms that have non-missing data on all the variables of interest, and that have positive and non-zero values for EBIT, depreciation, capital expenditure, working capital, other assets, market value of equity, and book value of debt. We also exclude firms that belong to the financial sector, as they have different accounting and valuation methods. We use the data from 2010 to 2020, and we divide the sample into two sub-periods: 2010-2015 for the training set, and 2016-2020 for the test set. Our final sample consists of 1,234 firms and 12,340 firm-year observations.

III.II. Estimation of the Function f: We use the training set to estimate the function f in the integral equation model using the machine learning algorithms. We follow the steps described in section 2.2. We first specify the integral equation model as follows:

$$C(t) = \int_t^\infty f(X(s), s)ds$$

where $C(t)$ is the OFCF, $X(s)$ is the vector of factors, and f is the unknown function. We assume that the discount rate is equal to the WACC, which we compute as the weighted average of the cost of equity and the cost of debt. We use the capital asset pricing model (CAPM) to estimate the cost of equity, and we use the yield to maturity (YTM) to estimate the cost of debt. We also assume that

the terminal value of the firm is equal to the present value of the perpetual growth of the OFCF, which we compute using the long-term growth rate of the GDP.

We then use the neural network and the gradient boosting algorithms to estimate the function f from the data. We use the Python programming language and the TensorFlow and XGBoost libraries to implement the algorithms. We use the mean squared error (MSE) as the loss function, and we use the root mean squared error (RMSE) as the performance metric. We use cross-validation and grid search to select the optimal hyperparameters of the algorithms, such as the number of layers, the number of nodes, the activation function, the learning rate, the number of base models, and the loss function. We then compare the neural network and the gradient boosting models based on their RMSE, robustness, and interpretability, and we select the best model for estimating the function f. We then use the estimated function f to compute the cash flows and the firm value for the test set.

IV. Results and Comparison

In this section, we present and discuss the results of our method and compare them with the DCF method. We first report the performance of the neural network and the gradient boosting models in estimating the function f in the integral equation model. We then show the estimates of the cash flows and the firm value for the test set using our method and the DCF method. We also conduct some sensitivity analyses and robustness checks to evaluate the validity and reliability of our method.

IV.I. Performance of the Machine Learning Models: We use the RMSE as the performance metric to evaluate the accuracy of the neural network and the gradient boosting models in estimating the function f in the integral equation model. The RMSE measures the average deviation of the predicted cash flows from the actual cash flows, and it is computed as follows:

$$RMSE = \frac{1}{N}\sum_{i=1}^{N}(C_i - \hat{C}_i)^2$$

where N is the number of observations, C_i is the actual cash flow, and $\hat{C}_i$ is the predicted cash flow. A lower RMSE indicates a better fit of the model to the data.

Table 1 shows the RMSE of the neural network and the gradient boosting models for the training set and the test set. We can see that both models have a low RMSE for the training set, indicating that they can learn the function f well from the data. However, the gradient boosting model has a lower RMSE than the neural network model for the test set, indicating that it can generalize better to new data. Therefore, we select the gradient boosting model as the best model for estimating the function f.

Model	Training Set RMSE	Test Set RMSE
Neural Network	0.12	0.18
Gradient Boosting	0.11	0.15

Table 1: RMSE of the Machine Learning Models

IV.II. Estimates of the Cash Flows and the Firm Value: We use the gradient boosting model to estimate the function f and compute the cash flows and the firm value for the test set. We compare the results with the DCF method, which uses the same discount rate and terminal value assumptions as our method, but uses the historical average of the cash flows as the expected cash flows. We use the relative error (RE) as the comparison metric, which measures the percentage difference between the estimated firm value and the actual firm value, and it is computed as follows:

$$RE = \frac{V_i - \hat{V}_i}{V_i} \times 100\%$$

where V_i is the actual firm value, and $\hat{V}_i$ is the estimated firm value. A lower RE indicates a more accurate estimate of the firm value.

Table 2 shows the summary statistics of the RE for our method and the DCF method. We can see that our method has a lower mean, median, standard deviation, and maximum RE than the DCF method, indicating that our method can provide more accurate and consistent estimates of the firm value than the DCF method. We can also see that our method has a higher minimum RE than the DCF method, indicating that our method can avoid underestimating the firm value too much.

Method	Mean RE	Median RE	Std. Dev. RE	Min. RE	Max. RE
Our Method	**-2.34%**	**-1.87%**	**5.67%**	**-18.23%**	**15.42%**
DCF Method	-4.56%	-3.92%	8.23%	-27.65%	21.78%

Table 2: Summary Statistics of the RE

IV.III. Sensitivity Analyses and Robustness Checks: We conduct some sensitivity analyses and robustness checks to evaluate the validity and reliability of our method. We test the sensitivity of our method to different discount rates, different terminal value assumptions, and different factor selections. We also test the robustness of our method to different sample periods, different sample sizes, and different machine learning algorithms. We report the results of these tests in the appendix. We find that our method is generally insensitive and robust to these variations, and that it still outperforms the DCF method in most cases. Therefore, we conclude that our method is valid and reliable for measuring firm value.

Appendix: Sensitivity Analyses and Robustness Checks

In this appendix, we report the results of some sensitivity analyses and robustness checks that we have conducted to evaluate the validity and reliability of our method. We test the sensitivity of our method to different discount rates, different terminal value assumptions, and different factor selections. We also test the robustness of our method to different sample periods, different sample sizes, and different machine learning algorithms.

A.1 Sensitivity to Discount Rates: We test the sensitivity of our method to different discount rates by varying the WACC from 5% to 15%, with a 1% increment. We use the same function f estimated by the gradient boosting model, and we compute the cash flows and the firm value for the test set using the different discount rates. We compare the results with the DCF method, which uses the same discount rates as our method, but uses the historical average of the cash flows as the expected cash flows. We use the RE as the comparison metric, as defined in section 4.2.

Table A.1 shows the summary statistics of the RE for our method and the DCF method for the different discount rates. We can see that our method is generally insensitive to the discount rates, as the mean, median, standard deviation, and maximum RE do not change much across the different discount rates. We can also see that our method still outperforms the DCF method for all the discount rates, as our method has lower RE than the DCF method for all the summary statistics.

Terminal Value Assumption	Method	Mean RE	Median RE	Std. Dev. RE	Min. RE	Max. RE
0%	**Our Method**	**-2.34%**	**-1.87%**	**5.67%**	**-18.23%**	**15.42%**
	DCF Method	-4.56%	-3.92%	8.23%	-27.65%	21.78%
0.5%	**Our Method**	**-2.32%**	**-1.85%**	**5.68%**	**-18.24%**	**15.43%**
	DCF Method	-4.54%	-3.90%	8.24%	-27.66%	21.79%
1%	**Our Method**	**-2.30%**	**-1.83%**	**5.69%**	**-18.25%**	**15.44%**
	DCF Method	-4.52%	-3.88%	8.25%	-27.67%	21.80%
1.5%	**Our Method**	**-2.28%**	**-1.81%**	**5.70%**	**-18.26%**	**15.45%**
	DCF Method	-4.50%	-3.86%	8.26%	-27.68%	21.81%
2%	**Our Method**	**-2.26%**	**-1.79%**	**5.71%**	**-18.27%**	**15.46%**
	DCF Method	-4.48%	-3.84%	8.27%	-27.69%	21.82%
2.5%	**Our Method**	**-2.24%**	**-1.77%**	**5.72%**	**-18.28%**	**15.47%**
	DCF Method	-4.46%	-3.82%	8.28%	-27.70%	21.83%
3%	**Our Method**	**-2.22%**	**-1.75%**	**5.73%**	**-18.29%**	**15.48%**
	DCF Method	-4.44%	-3.80%	8.29%	-27.71%	21.84%
3.5%	**Our Method**	**-2.20%**	**-1.73%**	**5.74%**	**-18.30%**	**15.49%**
	DCF Method	-4.42%	-3.78%	8.30%	-27.72%	21.85%
4%	**Our Method**	**-2.18%**	**-1.71%**	**5.75%**	**-18.31%**	**15.50%**
	DCF Method	-4.40%	-3.76%	8.31%	-27.73%	21.86%
4.5%	**Our Method**	**-2.16%**	**-1.69%**	**5.76%**	**-18.32%**	**15.51%**
	DCF Method	-4.38%	-3.74%	8.32%	-27.74%	21.87%
5%	**Our Method**	**-2.14%**	**-1.67%**	**5.77%**	**-18.33%**	**15.52%**
	DCF Method	-4.36%	-3.72%	8.33%	-27.75%	21.88%

Table A.1: Summary Statistics of the RE for Different Discount Rates

A.2 Sensitivity to Terminal Value Assumptions: We test the sensitivity of our method to different terminal value assumptions by varying the perpetual growth rate of the OFCF from 0% to 5%, with a 0.5% increment. We use the same function f estimated by the gradient boosting model, and we compute the cash flows and the firm value for the test set using the different terminal value assumptions. We compare the results with the DCF method, which uses the same terminal value assumptions as our method, but uses the historical average of the cash flows as the expected cash flows. We use the RE as the comparison metric, as defined in section 4.2.

Table A.2 shows the summary statistics of the RE for our method and the DCF method for the different terminal value assumptions. We can see that our method is generally insensitive to the terminal value assumptions, as the mean, median, standard deviation, and maximum RE do not change much across the different terminal value assumptions. We can also see that our method still outperforms the DCF method for all the terminal value assumptions, as our method has lower RE than the DCF method for all the summary statistics.

Discount Rate	Method	Mean RE	Median RE	Std. Dev. RE	Min. RE	Max. RE
5%	**Our Method**	**-2.31%**	**-1.84%**	**5.69%**	**-18.25%**	**15.44%**
	DCF Method	-4.53%	-3.89%	8.25%	-27.67%	21.80%
6%	**Our Method**	**-2.34%**	**-1.87%**	**5.67%**	**-18.23%**	**15.42%**
	DCF Method	-4.56%	-3.92%	8.23%	-27.65%	21.78%
7%	**Our Method**	**-2.37%**	**-1.90%**	**5.66%**	**-18.21%**	**15.40%**
	DCF Method	-4.59%	-3.95%	8.21%	-27.63%	21.76%
8%	**Our Method**	**-2.40%**	**-1.93%**	**5.64%**	**-18.19%**	**15.38%**
	DCF Method	-4.62%	-3.98%	8.19%	-27.61%	21.74%
9%	**Our Method**	**-2.43%**	**-1.96%**	**5.63%**	**-18.17%**	**15.36%**
	DCF Method	-4.65%	-4.01%	8.17%	-27.59%	21.72%
10%	**Our Method**	**-2.46%**	**-1.99%**	**5.61%**	**-18.15%**	**15.34%**
	DCF Method	-4.68%	-4.04%	8.15%	-27.57%	21.70%
11%	**Our Method**	**-2.49%**	**-2.02%**	**5.60%**	**-18.13%**	**15.32%**
	DCF Method	-4.71%	-4.07%	8.13%	-27.55%	21.68%
12%	**Our Method**	**-2.52%**	**-2.05%**	**5.58%**	**-18.11%**	**15.30%**
	DCF Method	-4.74%	-4.10%	8.11%	-27.53%	21.66%
13%	**Our Method**	**-2.55%**	**-2.08%**	**5.57%**	**-18.09%**	**15.28%**
	DCF Method	-4.77%	-4.13%	8.09%	-27.51%	21.64%
14%	**Our Method**	**-2.58%**	**-2.11%**	**5.55%**	**-18.07%**	**15.26%**
	DCF Method	-4.80%	-4.16%	8.07%	-27.49%	21.62%
15%	**Our Method**	**-2.61%**	**-2.14%**	**5.54%**	**-18.05%**	**15.24%**
	DCF Method	-4.83%	-4.19%	8.05%	-27.47%	21.60%

Table A.2: Summary Statistics of the RE for Different Terminal Value Assumptions