1. A literature review of potential methods for solving the problem

A balance sheet forecast is a financial tool used by organizations to project their future financial position by estimating key components such as assets, liabilities, and equity. It is integral for effective financial planning, enabling businesses to anticipate their financial health and make informed strategic decisions. Samonas, M. (2015) underscores that financial modeling, including balance sheet forecasting, is a means to simplify and abstract complex economic realities to support informed decision-making​. The accuracy of a balance sheet forecast is paramount, as it not only informs budgeting and resource allocation but also guides capital management and risk assessment efforts. Brigham, E. F. (1982) highlights the methodological importance of precision in modeling assumptions to ensure the reliability of financial projections​.

The significance of balance sheet forecasting lies in its ability to create a coherent financial strategy that aligns with both current performance and future objectives. The forecasting process involves a thorough analysis of historical data, identification of key financial drivers, and integration with income statement forecasts to ensure consistency across financial statements (Brigham, E. F., & Houston, J. F. (2013).). Francis, J., Olsson, P., & Schipper, K. (2008) further elaborates on the creative yet constrained nature of theoretical modeling, emphasizing that scenario planning and ratio analysis are indispensable for anticipating and navigating market fluctuations​.

Critical components include the estimation of current and long-term assets, current and long-term liabilities, and equity, which together provide a comprehensive view of an organization's financial stability. However, balance sheet forecasting is not without challenges. The dynamic nature of financial variables, the necessity for accurate historical data, and the influence of external market conditions can complicate the forecasting process (Koller, T., Goedhart, M., & Wessels, D. (2010); Higgins, R. C. (2016)). These challenges align with Penman, S. H. (2013) discussion of the iterative process in economic model refinement to address interdependencies and potential inaccuracies in assumptions​. Additionally, interdependencies among financial statements and potential inaccuracies in underlying assumptions may lead to significant discrepancies in projected outcomes.

1. Propose your choice of methods and the reason for the choice

To ensure reliable and adaptable financial projections, it is essential to combine precision in assumptions with flexibility in addressing economic variability. This can be achieved through the integration of historical data analysis and dynamic modeling techniques. First, historical data analysis provides a strong foundation for identifying key financial trends and patterns. As highlighted by Brigham (1982), precise assumptions grounded in accurate historical data are critical to the reliability of financial forecasts. By analyzing past performance, this method reduces uncertainty and creates a realistic starting point for projections. Second, dynamic modeling techniques, such as ARIMA or VAR, address the volatility of external variables like interest rates and inflation. These methods enhance the model's adaptability, allowing it to respond effectively to economic fluctuations and market conditions, thereby improving its relevance in dynamic environments.

1. Present the answers of the question in clear and understandable language.
2. We would like to forecast the balance sheet of a company.  Unfortunately, the different fields of a balance sheet are not independent.  Hence we have to construct a model that respect these identities.  For a short introduction of the problem, please consider the papers Velez-Pareja(09) and Velez-Pareja(10).  For a much more detail exposition of the problem, please consult Shahnazarian(04) and the textbook “financial forecasting, analysis and modelling” by Samonas, as well as other standard accounting textbooks.

The key to forecasting a balance sheet is maintaining the accounting equation (Assets = Liabilities + Equity) while handling the interdependence of fields. We need to create an integrated model linking the balance sheet to the income statement and cash flow statement, ensuring dynamic consistency.

One branch of financial forecasting research focuses on using iterative methods to resolve circular dependencies, such as those between debt levels and interest expenses. This approach emphasizes the validation of models with historical data to ensure accuracy and consistency. Additionally, it includes testing the model under different scenarios to assess its robustness and adaptability to various assumptions.

However, Vélez-Pareja critiques the common practice of using "plugs" in financial forecasting. While this method is easy to implement, it hides potential errors in the financial statements. For example, even if data inputs or calculations are incorrect, the plug mechanism masks the issue by making the balance sheet appear consistent. This approach undermines the principles of double-entry accounting and increases the risk of undetected errors. They proposed a method without using plug and circularity. The key idea is to adhere strictly to accounting principles, which ensures that any inconsistency or error is immediately detectable. This approach avoids arbitrary adjustments (plugs) and instead relies on a systematic, modular design. The proposed model uses a cash budget (CB) as its foundation, which records all cash inflows and outflows in detail. This allows the model to dynamically calculate cash surpluses or deficits and determine financing needs (e.g., short-term or long-term loans) logically. By structuring the model step-by-step—starting with the cash budget, followed by the income statement, and then the balance sheet—it eliminates the need for circular references.

1. Construct a very simple model of the balance sheet based on the tools of Velez-Pareja(09) and Velez-Pareja(10).  Please write down the mathematical equations government the evolution of the fields of balance sheet.  Is it possible to model this problem as a time series?  How do we handle the accounting identities?

The model of the balance sheet should capture:

* 1. Asset:
  2. Liability:
  3. Equity
  4. Net Income
  5. Cash budget

1. Implement the model in Tensorflow and python

I uploaded the code in Github.

1. You can get income statement and balance sheet data from yahoo finance.  This blog post may help you.  <https://rfachrizal.medium.com/how-to-obtain-financial-statements-from-stocks-using-yfinance-87c432b803b8>

The income data from Yahoo Finance is limited to the past 5 years. Here, I used data from WRDS instead.

1. Choose some companies to apply your model to.  How are you going to train your model?  How can you test if your model is good at forecasting the balance sheet of the company?  How can you ensure that your forecast at least respect the accounting identities, and at least satisfying the asset = liability + equity identity as other relationship stated in the papers quoted here?

**How are you going to train your model?** Due to the limited availability of time-series data for a single company, we will train the model using data from multiple companies, dividing the dataset by company rather than time. This approach allows us to leverage a larger dataset, capturing a diverse range of financial behaviors across companies, which enhances the model’s generalizability. By incorporating data from various companies, we can identify common patterns and relationships that apply broadly across industries or within specific sectors.

**How can you test if your model is good at forecasting the balance sheet of the company?**To test if the model is good at forecasting the balance sheet, we would evaluate both the directionality and quantitative accuracy of its predictions. Directionality testing ensures the model captures correct trends, such as assets increasing with revenue growth or liabilities decreasing with reduced financing, by comparing the predicted changes with actual changes in sign and overall trend. Quantitative accuracy is assessed using metrics like Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE), which measure the size of the error between predicted and actual values. Combining these approaches allows us to validate that the model not only aligns with realistic financial behaviors but also delivers precise forecasts. This dual evaluation ensures robustness in both high-level trends and detailed values.

**How can you ensure that your forecast at least respect the accounting identities, and at least satisfying the asset = liability + equity identity as other relationship stated in the papers quoted here?** I follow the methodology proposed by Vélez-Pareja strictly, avoiding using arbitrary balancing figures (plugs) and calculating each field logically based on drivers and dependencies. Besides,

1. Can you use your model to forecast earnings?

Yes, the balance sheet forecasting model can be extended to forecast earnings by integrating it with the income statement. Earnings, often represented as Net Income, are a direct output of the income statement and are intrinsically linked to other components of the financial statements.

1. What are the ML techniques we can use to your model to make it better?

Random Forest is particularly effective for predicting equity investment across a large dataset, such as S&P 500 companies (n=500) over different time period for different companies (t > 20). It can handle panel data efficiently, capturing patterns and differences among these companies without requiring sequential time dependencies. For instance, representing features such as firm age, industry, market capitalization or other financial indicators, Random Forest identifies key drivers of equity investment and other input variables for forecasting the financial statements. This method is robust to the heterogeneity of financial data across 500 companies, works well with non-linear relationships, and is computationally efficient, making it ideal for large-scale horizontal comparisons in a given period.

1. Hint: simulation is highly related to prediction.  Suppose that you can simulate y(t+1) given y(t).  The prediction problem is very simple to implement numerically.  A general form of the model can be written as y(t+1) = f( x(t), y(t) ) + n(t),  where n(t) is some noise term to be specified, and x(t) are additional sets of variables that are relevant for the simulation.  What should x(t) be?

The variables x(t) should include all relevant external and internal factors, such as macroeconomic indicators, industry trends, or derived variables like growth rates, to make the simulation more accurate. Recently, the use of investor attention metrics, such as website search behavior (e.g., EDGAR filings or Google search trends), has gained significant interest in prediction models.

1. Testing plans and testing results for checking your implementation is correct and your results are valid.

First, following the method in Vélez-Pareja, the identity, Assets = Liabilities + Equity, is going to be satisfied. To test the accuracy of prediction, we would evaluate both the directionality and quantitative accuracy of its predictions among all the sample companies. Directionality testing ensures the model captures correct trends, such as assets increasing with revenue growth or liabilities decreasing with reduced financing, by comparing the predicted changes with actual changes in sign and overall trend. Quantitative accuracy is assessed using metrics like Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE), which measure the size of the error between predicted and actual values.

Reference:

Samonas, M. (2015). Financial forecasting, analysis, and modelling: a framework for long-term forecasting. John Wiley & Sons.

Brigham, E. F. (1982). *Financial management: Theory and practice*. Atlantic Publishers & Distri.

Brigham, E. F., & Houston, J. F. (2013). *Fundamentals of financial management*. South-Western Cengage Learning.

Francis, J., Olsson, P., & Schipper, K. (2008). Earnings quality. *Foundations and Trends® in* Accounting, 1(4), 259-340.

Koller, T., Goedhart, M., & Wessels, D. (2010). Valuation: measuring and managing the value of companies. John Wiley & Sons.

Higgins, R. C. (2016). Analysis for financial management. McGraw-Hill.

Penman, S. H. (2013). Financial statement analysis and security valuation. McGraw-hill.