Income Statement and Balance Sheet Forecast

Marco Yuyuan Zhang *
December 26, 2024

Abstract

Financial statements are collections of interdependent fields that collectively illustrate a firm's financial position. A robust model for predicting financial statements should adhere to the identities established by accounting principles, such as the equation: assets = liabilities + equity. This report is organized as follows: Section 1 transforms a standard spreadsheet model of the balance sheet and income statement into a time series representation and implements this model using TensorFlow. Section 2 examines the model's performance and explores machine learning techniques that can enhance prediction accuracy.

^{*}PhD Candidate in Economics, HKU Business School, The University of Hong Kong. Email: marcoyzhang@gmail.com. This report is intended for the 2025 Machine Learning Center of Excellence Summer Associate – Time Series & Reinforcement Learning Internship at J.P. Morgan Chase.

A balance sheet consists of three important indicators that aggregate key information representing the financial positions of a firm: assets, liabilities, and equity. Balance sheet forecasting involves predicting the future values of all three variables as well as their important components such as current assets and retained earnings. The balance sheet is designed in a way such that the fundamental accounting identity – namely, assets = liabilities + equity – always holds. A reasonable forecast model should respect this and other accounting identities when making predictions. In other words, the predicted output should satisfies some constraints imposed by basic accounting principles.

Mathematically, a balance sheet can be represented by a time series model:

$$y(t+1) = f(x(t), y(t)) + n(t)$$
(0.1)

where y(t) is a random vector that collects all fields of a balance sheet. For example, $y(t) = (\operatorname{assets}_t, \operatorname{liabilities}_t, \operatorname{equity}_t)'$ is the most simplified version of a balance sheet (x, t) represents a vector of exogenous variables such as net income and/or cash flow from external financing. Exogenous variables are those that affect the balance sheet but not included in it. (x, t) is the noise term, reflecting the uncertainty in forecasting balance sheet fields. The forecasting function, denoted by (x, t), is the key element we aim to identify. It is crucial that the choice of (x, t) has to reflect all accounting identities in order to produce a reasonable predictive model.

Model 0.1 represents a one-step-ahead, multi-output problem which is straightforward to implement using Google's TensorFlow library in Python. [how to train model]

Balance sheet, income statement, and cash flow data are sourced from Yahoo Finance using the yfinance data API. I obtain the financial statements for all companies included in the Nasdaq 100 index – stocks of 100 largest non-financial companies listed on the Nasdaq Stock Exchange and one of the most widely tracked stock index around the world. Nasdaq 100 has a large concentration in blue chip stocks in technology and biotechnology sector. In other words, a big fraction of companies in Nasdaq 100 are industry-leading tech giants that have more stable financial health compared with small and medium cap companies.

[test forecast ability for balance sheet]

[check forecast respects accounting identities]

[forecast earnings using balance sheet model]

[ML techniques improve model. compare. random walk. linear. LSTM.]

Literature. Velez-Pareja (2010, 2011) provides a detailed description of the accounting identities built in standard financial statements. This balance sheet model formalizes the spreadsheet model in Velez-Pareja (2010) and Velez-Pareja (2011) as a time series representation and applied machine learning techniques to improve its predictability.

¹By default, all vectors are column vectors

1 Model

A simplified balance sheet model comprises four key equations: three that govern the evolution of total assets, liabilities, and equity, and one that represents the fundamental accounting identity: assets = liabilities + equity.

This four-equation balance sheet model can be easily cast into Model 0.1 where $y(t) = (\text{Total assets}_t, \text{Total liabilities}_t, \text{Total equity}_t)'$ and x(t) consists of all variables on the right hand side from equation 1.1 to 1.3. Lagged values of both y(t) and x(t) are useful for predicting y(t). Appendix B specifies a comprehensive model, as outlined by Velez-Pereja (2010). +

Forecasting Earnings

Earnings, or net income, is defined as profits after all expenses, taxes, and costs have been deducted from total revenue ²:

```
Net Income_t = Sales revenues_t - Cost of goods sold_t - A&S expenses_t - Depreciation_t + Return from short-term investments_t - Interest payments_t - Income Taxes_t (1.5)
```

2 Testing and Validation

[present tests here.]

Reference

Velez-Pareja, I. (2010). Constructing Consistent Financial Planning Models for Valuation. IIMS Journal of Management of Science, 1.

Vélez-Pareja, I. (2011). Forecasting Financial Statements with No Plugs and No Circularity. The IUP Journal of Accounting Research & Audit Practices, 10(1).

²Henceforth, earnings and net income are used interchangeably.

Appendix

Question 2: Income Statement and Balance Sheet Forecast

Part 1

- 1) 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 respects these identities. For a short introduction to the problem, please consider the papers Velez-Pareja(09) and Velez-Pareja(10). For a much more detailed 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.
- 2) 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 governing 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?
 - 3) Implement the model in TensorFlow and Python.
- 4) 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-s
- 5) 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 respects the accounting identities and at least satisfies the asset = liability + equity identity as other relationships stated in the papers quoted here?
 - 6) Can you use your model to forecast earnings?
 - 7) What are the ML techniques we can use to make your model better?
- 8) 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?

Part 2

- a) Choose your favourite LLM to apply the problem of financial statement analysis.
- b) Let's try the task of balance sheet forecast using the same set of data as collected in part 1, does the LLM you picked perform better or worse than your model?
- c) Is it possible to combine your model in part 1 and LLM to create an ensemble model that performs better than the individual model in balance sheet forecast?
- d) Given the results of your analysis, pick a company you have analysed, what would you recommend to the CFO or CEO of this particular company given your results?

Full Model

Table 2 Nominal increase

$$Selling_t = (1 + Inflation \ rate_t)(1 + Real \ increase \ in \ selling \ price_t) - 1 \tag{.1}$$

$${\rm Purchasing}_t = (1 + {\rm Inflation} \ {\rm rate}_t)(1 + {\rm Real} \ {\rm increase} \ {\rm in} \ {\rm purchase} \ {\rm price}_t) - 1 \eqno(.2)$$

 $\label{eq:overhead} \text{Overhead expenses}_t = (1 + \text{Inflation rate}_t)(1 + \text{Real increase in overhead expenses}_t) - 1 \ \ \text{(.3)}$

$${\bf Payroll\ expenses}_t = (1 + {\bf Inflation\ rate}_t)(1 + {\bf Real\ increase\ in\ payroll\ expenses}_t) - 1 \eqno(.4)$$

$$\begin{cases} \text{Minimum cash required}_0 = \text{Minimum cash required for initial year} & t = 0 \\ \text{Minimum cash required}_t = \% \text{ of sales as } \cosh \times \text{Total sales}_t & t \geq 1 \end{cases}$$
 (.5)

Table 3 Forecasting volume, prices and sales revenues

Increase factor in volume
$$_t = 1 + \text{Increase}$$
 in sales volume (units) $_t$, $t \ge 1$ (.6)

$$\begin{cases} \text{Sales in units}_0 = b_0 \times \text{Price}^b & t = 0 \\ \text{Sales in units}_t = \text{Sales in units}_{t-1} \times \text{Increase factor in volume}_t & t \ge 1 \end{cases}$$

$$\begin{cases} \text{Selling price}_0 = \text{selling price} & t = 0 \\ \text{Selling price}_t = \text{Selling price}_{t-1} \times (1 + \text{selling}_t) & t \ge 1 \end{cases}$$

$$(.7)$$

$$\begin{cases} \text{Selling price}_0 = \text{selling price} & t = 0 \\ \text{Selling price}_t = \text{Selling price}_{t-1} \times (1 + \text{selling}_t) & t \ge 1 \end{cases} \tag{.8}$$

Total sales_t = Selling price_t × sales in units_t,
$$t \ge 1$$
 (.9)

Table 4 Forecasting Risk free rate and cost of debt and investment return

Risk free rate,
$$R_{f,t} = (1 + \text{inflation rate}_t)(1 + \text{Real interest rate}_t) - 1$$
 (.10)

Return on ST investment_t = Risk free rate, $R_{f,t}$ + Risk premium for return on ST investment_t (.11)

Cost of debt,
$$K_{d,t} = \text{Risk}$$
 free rate, $R_{f,t} + \text{Risk}$ premium for cost of debt_t (.12)

Table 5 Depreciation schedule and investment in fixed assets

Beginning Net fixed $assets_0 = 0$ Beginning Net fixed $assets_t = Beginning Net fixed <math>assets_{t-1} + New fixed assets_{t-1} - Annual depreciation (.13)$

Annual depreciation for investment in year $0_1 = \frac{\text{New fixed assets}_0}{\text{Lineal depreciation (4 years)}}$ t = 1Annual depreciation for investment in year $0_t = \text{Annual depreciation for investment in year } 0_t = 1$ (.14)

Annual depreciation for investment in year $1_2 = \frac{\text{New fixed assets}_1}{\text{Lineal depreciation (4 years)}}$ $t = \frac{\text{New fixed assets}_1}{\text{Annual depreciation for investment in year } 1_t = \frac{\text{New fixed assets}_1}{\text{Lineal depreciation for investment in year } 1_t = \frac{\text{New fixed assets}_1}{\text{Lineal depreciation for investment in year } 1_t = \frac{\text{New fixed assets}_1}{\text{Lineal depreciation (4 years)}}$ $t = \frac{1}{12}$

Annual depreciation for investment in year $2_3 = \frac{\text{New fixed assets}_2}{\text{Lineal depreciation (4 years)}}$ $t = \frac{\text{New fixed assets}_2}{\text{Annual depreciation for investment in year 2}_t} = \frac{\text{New fixed assets}_2}{\text{Lineal depreciation for investment in year 2}_t}$ $t = \frac{(.16)}{}$

Annual depreciation for investment in year $3_4 = \frac{\text{New fixed assets}_3}{\text{Lineal depreciation (4 years)}}$ (.17)

 $\begin{cases} \text{Annual depreciation}_0 = 0 & t = 0 \\ \text{Annual depreciation}_t = \sum_{k=\max(0,t-4)}^{t-1} \text{Annual depreciation for investment in year } k_t & t \geq 1 \\ & \text{(.18)} \end{cases}$

 $\begin{cases} \text{Investment to keep fixed assets constant}_{-1} = \text{Fixed assets}_{0} & t = 0 \\ \text{Investment to keep fixed assets constant}_{t} = \text{Annual depreciation}_{t+1} & t \geq 1 \end{cases}$ (.20)

(up to second last period)

Investment in fixed assets for growth_t = Net fixed assets_{t-1}×Increase in sales volume (units)_{t+1} (.21)

New fixed assets t = 1 Investment in fixed assets for growth t + 1 Investment to keep fixed assets constant t = 1 (.22)

Net fixed assets_t = Beginning Net fixed assets_t + New fixed assets_t - Annual depreciation_t (.23)

Table 6a Inventory and purchases in units

Units sold_t = Sales in units_t
$$(.24)$$

$$\begin{cases} \text{Final inv.}_0 = \text{Initial investing (units)}_0 & t = 0 \\ \text{Final inv.}_t = \text{Units sold}_t \times \text{Inventory as \% of volume in units sold}_t & t \ge 1 \end{cases}$$
 (.25)

In.
$$inv._t = Final\ inv._{t-1}$$
 (.26)

$$Purchases_t = Units sold_t + Final inv._t - In. inv._t$$
 (.27)

Table 6b Inventory valuation and cost of goods sold in dollars

$$\begin{cases} \text{Unit } \cos t_0 = \text{Initial purchase price}_0 & t = 0 \\ \text{Unit } \cos t_t = \text{Unit } \cos t_{t-1} \times (1 + \text{Purchasing}_t) & t \ge 1 \end{cases}$$
 (.28)

In.
$$inv_{t} = Final\ inv_{t-1}$$
 (.29)

$$Purchases_t = Purchases_t \times unit cost_t \quad (purchase on RHS is unit)$$
 (.30)

Final inv._t = Unit
$$cost_t \times Final inv._t$$
 (.31)

$$COGS_t = In. inv._t + Purchases_t - Final inv._t$$
 (.32)

Table 7 Administrative and selling A&S expenses

$$\begin{cases} \text{Overhead expenses}_0 = \text{Estimated overhead expenses}_0 & t = 0 \\ \text{Overhead expenses}_t = \text{Overhead expenses}_{t-1} \times (1 + \text{Overhead expenses}_t) & t \geq 1 \end{cases}$$
 (.34)

$$\begin{cases} \text{Payroll expenses}_0 = \text{Administrative and sales payroll} & t = 0 \\ \text{Payroll expenses}_t = \text{Payroll expenses}_{t-1} \times (1 + \text{Payroll expenses}_t) & t \ge 1 \end{cases}$$
 (.35)

Advertising expenses_t = Total sales_t × Promotion and advertising as a fraction of sales (.36)

A & S expenses_t = Sales commissions_t+Overhead expenses_t+Payroll expenses_t+Advertising expenses_t (.37)

Table 8a Sales and purchases disaggregated according to the timing of flows

Inflow from current year_t = Total sales_t \times (1 – Account receivable as % of sales – advanced payment from (.39)

Credit sales_t = Total sales revenues_t × Account receivable as % of sales (.40)

Payment in $advance_t = Total sales revenues_t \times advanced payment from customers as % of next year sale (.41)$

$$Total purchases_t = Purchases_t$$
 (.42)

Purchases paid the same $year_0 = Total \ purchases_0$ Purchases paid the same $year_t = Total \ purchases \times (1 - Accounts \ payable \ as \% \ of \ purchases - Adva (.43)$

Purchases on
$$\operatorname{credit}_0 = \operatorname{Purchases}_0 - \operatorname{Purchases}$$
 paid the same year_0 $t = 0$
Purchases on $\operatorname{credit}_t = \operatorname{Total} \operatorname{purchases}_t \times \operatorname{Account}$ payable as % of purchases $t \ge 1$
(.44)

Payment in $advance_t = Total purchases_t \times Advance payments to suppliers as a % of next year purchases (.45)$

Table 8b Flows from sales and purchases

Sales revenues from current
$$year_t = Inflow from current year_t$$
 (.46)

Account Receivable_{$$t$$} = Credit sales _{$t-1$} (.47)

$$Advance payments_t = Payment in advance_{t+1}$$
 (.48)

Total inflows_t = Account Receivable_t+Sales revenue from current year_t+Advance payments_t (.49)

Purchases paid the current $year_t = Purchases paid the same <math>year_t$ (.50)

Payment of Accounts Payable_t = Purchases on credit_{t-1} (.51)

Table 9a Cash budget module 1 operating activities

Inflows from sales_t = Total inflows_t
$$(.52)$$

Total inflows_t = Inflows from sales_t
$$(.53)$$

$$Payments for purchases_t = Total outflows_t$$
 (.54)

Administrative and Selling expenses_{$$t$$} = A&S expenses _{t} (.55)

$$Income taxes_t = Income Taxes_t$$
 (.56)

(RHS from table 12)

Total cash outflows_t = Payments for purchases_t+Administrative and selling expenses_t+Income Taxes_t (.57)

Operating
$$NCB_t = Total inflows_t - Total cash outflows_t$$
 (.58)

net cash balance NCB

Table 9b Cash budget module 2 investing activities

Investment in fixed assets_{$$t$$} = New fixed assets _{t} (.59)

NCB of investment in assets_{$$t$$} = $-$ Investment in fixed assets _{t} (.60)

Table 9c Cash budget module 3 external financing

$$ST Loan_{t} = \begin{cases} 0 \\ - \text{(cumulated NCB}_{t-1} + \text{Operating NCB}_{t} - \text{Total ST loan payment}_{t} - \text{Minimum cash required} \end{cases}$$
(.62)

$$LT Loan_{t} = \begin{cases} 0 \\ -\left(\text{cumulated NCB}_{t-1} + \text{NCB after Capex}_{t} + \text{ST Loan}_{t} - \text{Total loan payment}_{t} - \text{Payments}_{t} \end{cases}$$
(.63)

$$Principal ST Loan_{t} = ST loan_{t-1}$$
 (.64)

Interest ST loan_{$$t$$} = IP _{t} (.65)

(from Table 11a)

$$Total ST loan payment_t = Interest ST loan_t + Principal ST loan_t$$
 (.66)

$$Interest LT loan_t = Total Interest_t (from Table 11b) (.68)$$

Total loan payment_t = Total ST loan payment_t + Principal LT loan_t + Interest LT loan_t (.69)

Table 9d Cash budget module 4 transactions with owners

$$IE_t = \frac{LT \, \text{loan}_t}{\% \, \text{of financing with debt, the rest is financed by equity}} \times (1 - \% \, \text{of financing with debt, the rest is financed})$$

$$(.71)$$

$$Div_t = \text{Next year dividends}_{t-1}$$
 (.72)

$$RS_t = \text{Annual depreciation}_t \times \text{Stock Repurchases as a \% of depreciation}_t$$
 (.73)

Payments to owners_t =
$$Div_t + RS_t$$
 (.74)

NCB with owners_t =
$$IE_t$$
 – Payments to owners_t (.75)

NCB of previous $modules_t = NCB$ with $owners_t + NCB$ of financing activities $_t + NCB$ after $Capex_t$ (.76)

Table 9e Cash budget module 5 discretionary transactions

Redemption of ST investments_{$$t$$} = ST Investments _{t} (.77)

Return from ST Investments_t = Return on ST Investments_t \times Redemption of ST Investments_t (.78)

Total inflow from ST Investments_t = Return from ST Investments_t+Redemption of ST Investments_t (.79)

$$ST Investments_{t} = \begin{cases} 0 \\ Cumulatively NCB_{t-1} + NCB \text{ of previous modules}_{t} + Total \text{ inflow from ST Investration} \end{cases}$$

$$(.80)$$

NCB of discretionary transactions_t = Total inflow from ST Investments_t – ST Investments_t (.81)

Year $NCB_t = NCB$ at previous modules_t + NCB of discretionary transactions_t (.82)

$$Cumulated NCB = Minimum required cash_{t-1}$$
 (.83)

Table 10 Checking the cumulated NCB and the minimum cash target

$$\begin{cases} \text{Calculated Cumulated NCB}_0 = \text{Year NCB}_0 \\ \text{Calculated Cumulated NCB}_t = \text{Calculated Cumulated NCB}_{t-1} + \text{Year NCB}_t \quad t \geq 1 \end{cases}$$

Check with MCT = Calculated Cumulated NCB_t – Minimum cash require, (.85)

Table 11a Short-term loan schedules

$$BB_t = EB_{t-1} \tag{.86}$$

$$IP_t = EB_{t-1} \times Kd_t \tag{.87}$$

$$PP_t = \frac{EB_{t-1}}{Short-term loan 2 (1 year)}$$
 (.88)

$$\begin{cases} EB_0 = ST \log_0 \\ EB_t = EB_{t-1} - PP_t + ST \log_t \quad t \ge 1 \end{cases}$$
 (.89)

$$Kd_t = Cost \text{ of debt, } Kd_t$$
 (.90)

Table 11b Long-term loan schedules

$$BB LT debt_t = EB LT debt_{t-1}$$
 (.91)

$$LT loan yr 0 = LT loan_0 (.92)$$

$$\begin{cases} \text{PP loan yr } 0_1 = \frac{\text{LT loan yr 0}}{\text{Long-term (LT) years Loan 3 (M years)}} & t = 1 \\ \text{PP loan yr } 0_t = \text{PP loan yr 0}_1 & t \geq 2 \end{cases} \tag{.93}$$

New loan yr
$$1 = LT loan_1$$
 (.94)

$$\begin{cases} \text{PP loan yr } 1_2 = \frac{\text{New loan yr 1}}{\text{Long-term (LT) years Loan 3 (M years)}} & t = 2 \\ \text{PP loan yr } 1_t = \text{PP loan yr } 1_2 & t \geq 3 \end{cases} \tag{.95}$$

New loan yr
$$2 = LT loan_2$$
 (.96)

$$\begin{cases} \text{PP loan yr 2}_3 = \frac{\text{New loan yr 2}}{\text{Long-term (LT) years Loan 3 (M years)}} & t = 3\\ \text{PP loan yr 2}_t = \text{PP loan yr 2}_3 & t \ge 4 \end{cases} \tag{.97}$$

New loan yr
$$3 = LT loan_3$$
 (.98)

$$\begin{cases} \text{PP loan yr 3}_4 = \frac{\text{New loan yr 3}}{\text{Long-term (LT) years Loan 3 (M years)}} & t = 4 \\ \text{PP loan yr 3}_t = \text{PP loan yr 3}_4 & t \ge 5 \end{cases} \tag{.99}$$

New loan yr
$$4 = LT loan_4$$
 (.100)

Total Interest_t = EB LT debt_{t-1} × Kd_t
$$t \ge 1$$
 (.101)

New debt
$$LT_t = New loan yr t$$
 (.102)

Total PP LT_t =
$$\sum_{k=0}^{t}$$
 PP loan year k (.103)

$$EB LT debt_t = BB LT debt_t + New debt LT_t - Total PP LT_t$$
 (.104)

Table 12 Income Statement

$$Sales revenues_t = Total sales_t$$
 (.105)

$$COGS_t = COGS_t \tag{.106}$$

RHS from Table 6b

$$Gross Income_t = Sales revenues_t - COGS_t$$
 (.107)

$$A\&S expenses_t = A\&S expenses_t$$
 (.108)

RHS from Table 7

$$Depreciation_t = Annual depreciation_t (.109)$$

$$EBIT_t = Gross\ Income_t - A\&S\ expenses_t - Depreciation_t \tag{.110}$$

Return from ST investment_t = Return from ST investments_t (.112)

RHS from Table 9e

$$EBT_t = EBIT_t + Return from ST investment_t - Interest payments_t$$
 (.113)

$$Income \ Taxes_t = \begin{cases} 0 & EBT_t \le 0 \\ EBT_t \times Corporate \ tax \ rate & otherwise \end{cases}$$
 (.114)

$$Net Income_t = EBT_t - Income Taxes_t$$
 (.115)

Next year Dividends_t = Net Income_t
$$\times$$
 Payout ratio (.116)

$$CRE_t = CRE_{t-1} + Net Income_{t-1} + Next year Dividends_{t-1}$$
 (.117)

Table 13 The Balance Sheet

$$Cash CB_t = Cumulated NCB_t$$
 (.118)

$$AR IT_t = Credit sales_t (.119)$$

Inventory
$$IT_t = Final inv._t$$
 (.120)

$$APP_t = Advance payment to suppliers_t$$
 (.121)

ST investment
$$CB_t = ST$$
 investments._t (.122)

$$Current \ assets_t = Cash \ CB_t + AR \ IT_t + Inventory \ IT_t + APP_t + ST \ investment \ CB_t \quad (.123)$$

Net fixed assets
$$IT_t = Net fixed assets_t$$
 (.124)

$$Total_t = Net fixed assets IT_t + Current assets_t$$
 (.125)

$$AP IT_t = Purchases on credit_t$$
 (.126)

$$APR_t = Advance payments_t (.127)$$

Short-term debt
$$CB_t = EB_t$$
 (.128)

$$Long-term debt CB_t = EB LT debt_t$$
 (.130)

Total Liabilities_t = Long-term debt
$$CB_t + Current \ liabilities_t$$
 (.131)

Equity investment
$$CB_t = Equity$$
 investment $CB_{t-1} + IE_t$ (.132)

Retained earnings
$$IS_t = CRE_t$$
 (.133)

Current year
$$NI_t = Net Income_t$$
 (.134)

 $\label{eq:linear_line$

 $Check_t = Liabilities and equity_t - Total_t$ (.137)