**Report for JPMorgan Chase MLCOE internship – Question 1, Part 1**

Jaebum (Albert) Chung

January 1, 2026

**Introduction**

Literature Review

*The Limitations of "Plug" Variables in Financial Forecasting*

Financial statement forecasting is a foundational practice in corporate finance, yet traditional methodologies often compromise the integrity of the double-entry accounting system. As reviewed in Pareja et al. (2009) and Pareja (2007), a significant portion of existing literature relies on the use of "plugs": balancing variables forced to reconcile the difference between Assets and Liabilities plus Equity. Typically, analysts assign a specific account (often Cash or Revolving Debt) to absorb any discrepancies in the forecast.

While this approach ensures a balanced sheet mathematically, it is methodologically flawed for robust forecasting. By using a plug, the analyst effectively overrides the natural evolution of the balance sheet, obscuring the causal relationships between operating decisions and financial health. More critically, the use of a plug erodes the primary validation mechanism of the double-entry principle. If the balance sheet is forced to balance, it becomes impossible to detect logical errors or inconsistencies in the underlying code. As noted in the critiques within the Pareja framework, a model that balances by force rather than by logic fails to serve as a reliable tool for stress testing or error detection.

Proposed Methodology: A Structural "No-Plug" Approach

To address these limitations, this report implements a structural forecasting model based on the framework proposed by Pareja (2009). The core distinction of this method is the rejection of circular references and balancing plugs. Instead, the model relies on a fully integrated system where the Balance Sheet is derived endogenously from the Cash Budget and Income Statement.

In this approach, every line item is driven by specific causal factors (e.g., Receivables driven by Sales, Payables driven by Opex), and the ending cash balance is a calculated result of these flows rather than a residual plug.

We have further refined the Pareja (2009) framework to better reflect the treasury strategies of modern multinational corporations like Apple or General Motors. Where the original model might treat Marketable Securities solely as a reservoir for excess cash, our implementation models these securities as a distinct portion of the company's liquidity strategy. This adjustment allows for a more realistic depiction of how large firms manage capital allocation between operational cash and interest-bearing assets.

Testing and Validation Plan

The validity of this implementation is verified through a rigorous "Balance Check." Because the model eschews plugs, the fundamental accounting identity (*Assets = Liabilities + Equity*) serves as a strict test of correctness rather than a guaranteed outcome.

Our testing plan involves running the forecast over a multi-year horizon and calculating the delta between Total Assets and Total Claims at every time step. In this "no-plug" environment, a delta of zero confirms that all cash inflows and outflows have been correctly captured and routed through the Cash Budget. Any deviation from zero indicates a specific error in the logic (e.g., a "leak" in the system where an expense was recorded but cash was not deducted), allowing us to pinpoint and resolve structural flaws that traditional models would mask.

The following sections (**1. The Balance Sheet Elements**, **2. Mathematical Equations Governing the Balance Sheet Elements,** and **3. Time Series Representation of the Balance Sheet** are answers to the questions 1 and 2)

1. **The Balance Sheet Elements**

The elements of a balance sheet from Pareja09 are as follows:

Assets

Net fixed assets (NFA)

Advance payments to suppliers for purchases (AdvPP)

Account receivables (AR)

Inventory (Inventory)

Short-term investments (Invest)

Cash equivalent (Cash)

Liabilities

Account payables (AP)

Advance payments from customers for sales (AdvPS)

Short-term debt (STDebt)

Long-term debt (LTDebt)

Equity

Equity invested (EI)

Cumulated retained earnings (CRE)

Cumulated buyback of equity (CBB)

Net Income (NI)

On the balance sheets of real companies, there are likely to be more elements. For example, there could be current and non-current ‘other liabilities’ other than the 4 mentioned above, and there could be various other assets including deferred tax, non-current account receivables, Goodwill, and other intangible assets. For the purpose of this simple modeling exercise, we first identify the elements listed in the above balance sheet, and lump all unaccounted-for elements as shown below, using GM’s balance sheet as example:

* Assets
  + All unaccounted-for non-interest-bearing elements are lumped with “non-current assets” along with net PPE.
    - Deferred tax, non-current account receivables, goodwill, other assets
    - Depreciation rate based on net fixed asset will be reduced accordingly.
  + We use other current assets as the proxy for advance payments to suppliers for purchases.
  + Short-term and long-term investments in market securities are lumped into “investments in market securities.”
    - Earn a combined average interest.
  + We also combine cash equivalents with investments in market securities to for “total liquidity” because companies do not just hold onto a pile of cash, but rather invest a significant portion of cash in securities to earn interest, which is different from Pareja09’s logic of only investing in excess cash after meeting a minimum cash balance.
    - Total liquidity = cash equivalent + investments in market securities
* Liabilities
  + Advance payments from customers to be fulfilled next year are represented as deferred revenue in company balance sheets.
  + All unaccounted-for elements are lumped with long-term debt, to be labeled “non-current liabilities”
    - Includes non-current long-term debt, employee benefits, non-current deferred revenue, non-current accrued expenses, long-term capital lease obligation
      * These other non-current liabilities are like interest-free long-term debt
      * The total long-term liability is assumed to have the average maturity of AvgM years, and it generates an average LT interest (AvgLTInt) owed by the company
    - Something important to note is that the “long-term debt” in an actual balance sheet EXCLUDES the principal owed next year.
      * Non-current liabilitiest = Total LT liabilityt \* (1 – 1/AvgM)

= Total LT liabilityt \* (AvgM - 1) / AvgM

* + We group all other liabilities due next year as “current liabilities.”
    - Includes STDebt, current accrued expenses which include dealer and customer allowances, claims and discounts, product warranty, payrolls and benefits, and others, to be considered as “short-term loan”
      * Accrued expenses are like interest-free short-term debt.
      * Short-term loan (STLoant) generates an average ST interest (AvgSTIntt) owed by the company
    - Also includes current debt which includes principal payment from short-term debt and ALSO INCLUDES the current portion of the total LT liability
      * Current liabilitiest = Total ST liabilityt + Total LT liabilityt / AvgM

= Total ST liabilityt + Non-current liabilitiest / (AvgM – 1)

* Equity
  + Represented by Stockholders Equity (SEt)
    - SEt = SEt-1 + EFt + NIt – Dividendst-1 – BBt 
      * SEt-1 = stockholder equity last year
      * EFt = equity financing to cover portion of CapEx spending
      * NIt = net income this year
      * Dividendst-1 = Dividends earned last year, paid this year
      * BBt = stock buyback this year

So, our updated balance sheet, along with subitems for clarity, becomes:

Assets

Non-current assets (**NCAt**)

* + Net PPE, deferred tax, non-current account receivables, goodwill, other assets

Advance payments to suppliers for purchases (**AdvPPt**)

Account receivables (**ARt**)

Inventory (**Invt**)

Total liquidity (**TLt**)

* + Investments in market securities (**IMSt**)
    - Short-term and long-term investments
  + Cash equivalent (**Casht**)

Liabilities

Account payables (**APt**)

Advance payments from customers for sales (**AdvPSt**)

Non-current liabilities (**NLiabt**)

* + Non-current employee benefits, non-current deferred revenue, non-current accrued expenses, long-term debt and capital lease obligations MINUS next year’s principal, other non-current liabilities
  + **NLiabt** = Total LT liabilityt \* (1 – 1/**AvgM**)

Current liabilities (**CLiabt**)

* + ST debt, current accrued expenses, other current liabilities, and current portion of the long-term debt and capital lease obligations principals
  + **STLoant** = ST debtt + current accrued expensest + other current liabilitiest
  + **CLiabt** = **STLoant** + Total LT liabilityt / **AvgM** = **STLoant** + **NLiabt** / (**AvgM** – 1)

Equity

Stockholder equity (**SEt**)

* + **SEt** = **SEt-1** + **EFt** + **NIt** – **Dividendst-1** – **BBt**
    - SEt-1 : total stockholder equity last year
    - EFt : equity financing to cover portion of CapEx spending
    - NIt : net income this year
    - Dividendst-1 = Dividends earned last year, paid this year
    - BBt : stock buyback this year

1. **Mathematical Equations Governing the Balance Sheet Elements**

We refer to Pareja09 for constructing the mathematical equations that link all the elements on the balance sheet without circularity or plugs. First, we define two drivers that will predict the evolution of the balance sheet’s elements. They are:

* Salest : the total revenue generated by the company in year t
* Purchasest : the total purchases made by the company in year t

This is a different approach from Pareja09 where they modeled the unit price, the unit cost, and the volume sold based on market research data on volume’s price sensitivity. For the scope of this problem set, we lump these variables into sales and purchases whose evolutions are modeled independently based on historical data.

We begin by modeling the asset elements.

* 1. Assets
     1. *Non-current Assets*

Fixed assets undergo depreciation over their lifecycle. For a well-established company with a stable growth, we can assume a percentage of the fixed assets depreciate, and capital expenditures are made to make up for the depreciation and grow the assets by an amount proportional to its revenue from sales. If we take into account that we are using non-current assets (**NCA**) in our model, the effective depreciation rate we calculate will be lower than the actual depreciation.

NCAt = NCAt-1 – Depreciationt + CapExt (1)

where

Depreciationt = NCAt-1 \* %Depreciation (2)

CapExt = Depreciationt + Salest \* %Asset growth (3)

We keep track of %Depreciation (%Depr) and %Asset growth (%AG) as the variables to learn when we train our model.

* + 1. *Advance Payments*

Advance payments to suppliers for purchases (AdvPP) can be modeled as being a fraction of the total purchases being made next year:

AdvPPt = %AdvPP \* Purchasest+1 (4)

We add %Advance payments to suppliers (%AdvPP) to the variables to learn in our training.

* + 1. *Account Receivables*

Next, we have account receivables. A fraction of this year’s sales to customers are on credit and is receivable next year:

ARt = %AR \* Salest (5)

We add %Account receivables (%AR) to the variables to learn in our training.

* + 1. *Inventory*

Inventory can be modeled as a function of last year’s inventory, this year’s purchases, and this year’s total cost of goods sold (COGS):

Invt = Invt-1 + Purchasest – COGSt (6)

As done in Pareja09, we set the inventory target as a percentage of sales:

Invt = %Inv \* Salest (7)

such that we can rewrite (6) to represent COGSt :

COGSt = %Inv \* (Salest-1 – Salest) + Purchasest (8)

This follows the first-in-first-out (FIFO) principle, as the existing inventory is exhausted first when calculating the cost of goods sold. We add %Inventory (%Inv) to the variables to be trained.

* + 1. *Total Liquidity*

Finally, the last element in our assets is total liquidity (TL). Following Pareja09, we model the total liquidity available as a percentage of sales. Within the liquidity, we assume that a portion of it is invested in market securities (IMS):

TLt = %TL \* Salest (9)

Casht = %Cash \* TLt (10)

IMSt = (1 - %Cash) \* TLt (11)

We add %TL and %Cash to our variables to be trained.

Before we move onto liabilities, it is helpful to generate a Liquidity Budget to model the flow of cash in and out of cash balance and short-term investment into market securities. We have updated the name from Cash Budget to Liquidity Budget to reflect our use of total liquidity to represent both cash and market securities investment.

* 1. Liquidity Budget

We can break down flows into/out of liquidity balance into net liquidity balances (NLB) pertaining to the following activities: operating, capital expense, financing, external investment, and transaction with owners. The final liquidity is the sum of all NLBs.

* + 1. *Operating NLB*

Operating activity captures all inflows from sales (current year sales revenue, account receivables from last year, advance payments from customers for next year) and all outflows (current year purchases, account payables from last year, advance payments to suppliers for next year, operating expenses, income tax):

Inflowst = ARt-1 + AdvPSt + Salest \* (1 - %AR - %AdvPS) (12)

Outflowst = APt-1 + AdvPPt + Purchasest \* (1 - %AP - %AdvPP) + OpExt + ITt (13)

Operating NLBt = Inflowst – Outflowst (14)

where OpExt is the operating expense, and ITt is income tax. We can model OpExt as a function of sales and inflation:

OpExt = %OR \* Salest + OBt-1 \* (1 + Inflationt) (15)

where OB is the base operating expense that increases with inflation each year (rent, insurance), and %OR is the rate of increase proportional to this year’s sales revenue that impacts the variable portion of the operating expense. Inflation data can be obtained from U.S. Bureau of Labor Statistics.

We can model ITt as a percentage of net income (NIt):

ITt = %IT \* NIt

We add %IT, OBT\_start (base operating expense of the first element in a time series) and %OR to our variables to be trained.

* + 1. *Capital Expense NLB*

Investment is made into the non-current assets to:

1. keep non-current assets price the same by accounting for depreciation
2. grow non-current assets by some percentage.

We have already done the calculation in Eq. (3), which we reproduce here:

CapExt = Depreciationt + Salest \* %Asset growth from Eq. (3)

CapEx NLBt = -CapExt = -Depreciationt - Salest \* %Asset growth (16)

* + 1. *Financing NLB*

This is the net liquidity balance after obtaining more liquidity from new short-term loan (STLoant) and new long-term loan (LTLoant), and paying all debt obligations this year.

Financing NLBt = STLoant + LTLoant - current liabilitiest-1 - Avg. ST interestt - Avg. LT interestt (17)

where current liabilitiest-1 (CLiabt-1) are the principals from short-term liabilities and long-term liabilities paid this year, represented by:

CLiabt-1 = STLoant-1 + NLiabt-1 / (AvgM – 1) (17)

The average short-term interest (AvgSTIntt) and average long-term interest (AvgLTIntt) owed are calculated as follows:

AvgSTIntt = %AvgSTInt \* STLoant-1 = %AvgSTInt \* (CLiabt-1 – NLiabt-1 / (AvgM – 1)) (18)

AvgLTIntt = %AvgLTInt \* Total LT liabilityt-1 = %AvgLTInt \* (NLiabt-1 / (1 – 1/AvgM)) (19)

We add AvgM, %AvgSTInt and %AvgLTInt to our variables to learn. We derive the expression for STLoant and LTLoant in 2.3 Liabilities.

* + 1. *External Investment NLB*

We redeem external investments from last year with a fixed percentage of return (%MSReturn), and make new investments this year:

External investment NLBt = IMSt-1 \* (1 + % MSReturn) – IMSt (20)

Instead of only allocating the excess cash in market securities, we invest a proportion of total liquidity balance. Using Eq. (11):

External investment NLBt = (1 - %Cash) \* (TLt-1 \* (1 + % MSReturn) – TLt) (21)

We add %MSReturn to our training variables to be learned.

* + 1. *Transaction with owners NLB*

In Pareja09’s model, owners can make additional investments to finance a portion of long-term debt, which counts as an inflow to liquidity. The company pays dividends to owners and can also engage in stock buybacks:

Owners transaction NLBt = Equity financingt – Dividendst-1 – Stock buybackst (22)

where

Equity financingt = LTLoant \* %EF / (1 - %EF) (23)

Dividendst-1 = %Payout ratio \* Net incomet-1 (24)

Stock buybackst = %Buyback ratio \* Depreciationt (25)

We add %Equity financing (%EF), %Payout ratio (%PR) and %Buyback ratio (%BB) to our variables to learn. We will define the equations for Net incomet in 2.4. Income Statement.

* + 1. *Final NLB*

We obtain the final net liquidity balancet by summing all the different activities’ NLBt:

Final NLBt = Operating NLBt + CapEx NLBt + Financing NLBt + External investment NLBt  (26)

+ Owners transaction NLBt

This is how we obtain the net liquidity balance at time t without plugs or circularity. The unknown variables in the equation can be trained because we have the target liquidity balance given by Eq. (9):

TLt = %TL \* Salest from Eq. (9)

* 1. Liabilities
     1. *Account Payables*

Account payables are modeled as a percentage of total purchases made this year, payable next year to the suppliers:

APt = %AP \* Purchasest (27)

We add %AP to our variables to be trained.

* + 1. *Advance Payments*

Advance payments from customers for sales are modeled as a fraction of next year’s sales:

AdvPSt = %AdvPS \* Salest+1 (28)

We add %AdvPS to our variables to be trained.

* + 1. *Current Liabilities*

To model the current liabilities, we first model the effective short-term loan (STLoant) issued this year. Short-term loan covers any excess cash needed to cover operational expenses and last year’s short-term loan obligation (principal + interest) due this year and bring the total liquidity to this year’s liquidity target. We also account for the return from investment in market securities here because the size of investment is directly proportional to cash, **instead of** investing excess cash after ST and LT loan calculations which is done in Pareja09:

Short-term cash deficitt

= TLt – TLt-1 – Operating NLBt + STLoant-1 \* (1 + %AvgSTInt) – %MSReturn \* IMSt-1 (29)

STLoant = max(0, Short-term cash deficitt) (30)

Then, we calculate the effective long-term loan (LTLoant) issued this year to finance CapExt, last year’s current portion of the long-term liability, interest on long-term liability, and dividends to bring the liquidity to this year’s target after accounting for the equity financing:

LTLoant = max(0, Short-term cash deficitt – STLoant – CapEx NLBt + NLiabt-1 / (AvgM – 1) + AvgLTIntt + Dividendst-1 + Stock buybackst) \* (1 - %EF) (31)

Current liabilities are summation of the effective short-term loan and the current portion of non-current liabilities, and are represented by:

CLiabt = STLoant + NLiabt-1 / (AvgM – 1) (32)

* + 1. *Non-current Liabilities*

Non-current liabilities include the new effective LTLoant issued this year, and last year’s non-current liabilities minus the current portion of this year’s non-current liabilities.

NLiabt = Total LT liabilityt \* (1 – 1/AvgM) (33)

where

Total LT liabilityt = LTLoant + NLiabt-1 (34)

* 1. Equity

This year’s equity accounts for the additional equity financing this year, net income from this year, dividends paid out this year, and stock buyback this year:

SEt = SEt-1 + EFt + NIt – Dividendst-1 – BBt (35)

where

EFt = LTLoant / (1 - %EF) \* %EF (36)

Dividendst-1 = %PR \* NIt-1 (37)

BBt = %BB \* %Depreciation \* NCAt-1 (38)

We have most of what we need to learn the training variables except dividends and income tax, which will come from modeling the net income in the income statement.

* 1. Income Statement

We have all the components needed from the previous sections to construct the elements related to net income.

EBITDAt (earnings before interest, taxes, depreciation, and amortization) can be calculated by:

EBITDAt = Salest – COGSt – OpExt (39)

Earnings before tax is calculated by subtracting interest payments, depreciation (and amortization):

EBTt = EBITDAt – Depreciationt – AvgSTIntt – AvgLTIntt + %MSReturn \* IMSt-1 (40)

Net income (NIt) is determined by subtracting the income tax:

NIt = EBT \* (1 - %IT) (41)

Next year’s dividends were already modeled in Eq. (24)

Dividendst = %PR \* NIt from Eq. (24)

We now have everything we need to represent every element in the balance sheet as a time series.

1. **Time Series Representation of the Balance Sheet**

We re-write every element in the balance sheet as a time series that have been simplified by combining the equations above. The primary drivers are Salest and Purchasest, which drive every element given the variables to be trained are learned.

* 1. Variables to be trained
* %AG : Asset growth rate as a percentage of non-current assets
* %Depr : Depreciation rate as a percentage of non-current assets
* %AdvPS : Advance payments from customers as a percentage of next year’s sales
* %AdvPP : Advance payments to suppliers as a percentage of next year’s purchases
* %AR : Account receivables as a percentage of this year’s sales
* %AP : Account payables as a percentage of this year’s purchases
* %Inv : Inventory target as a percentage of this year’s sales
* %TL : Total liquidity as a percentage of this year’s sales
* %Cash : Cash as a percentage of total liquidity
* %IT : Income tax as a percentage of net income this year
* %OR : Variable portion of OpEx that increases in proportion to this year’s sales
* OBT\_start : Baseline OpEx that increases with this year’s average inflation
* %AvgSTInt : Average short-term interest on the effective short-term loan
* %AvgLTInt : Average interest on total long-term liabilities
* AvgM : Average maturity of total long-term liability
* %MSReturn : Percentage return on the investment in market securities
* %EF : Equity financing as a percentage of total excess cash needed after accounting for long-term investments and long-term obligations
* %PR : Dividends payout ratio as a percentage of net income this year, paid next year
* %BB : Stock buyback as a percentage of asset depreciation this year
  1. Variables from External Environment

Since we assume a flat short-term and long-term interest for loans and return on market securities, the only external variable is inflation.

* Inflationt : Average inflation at year t
  1. Assets
* Non-current assets
  + NCAt = NCAt-1 \* (1 + %AG)
* Advance payments to suppliers for purchases
  + AdvPPt = %AdvPP \* Purchasest+1
* Account receivables
  + ARt = %AR \* Salest
* Inventory
  + Invt = %Inv \* Salest
* Total liquidity (**TLt**)
  + Target TLt = %TL \* Salest
    - Investments in market securities
      * IMSt = (1 - %Cash) \* TLt
    - Cash equivalent
      * Casht = %Cash \* TLt
  + Should be equivalent to: Final TLt = Operating NLBt + CapEx NLBt + Financing NLBt + External investment NLBt + Owners transactions NLBt
  1. Liabilities
* Account payables
  + APt = %AP \* Purchasest
* Advance payments from customers for sales
  + AdvPSt = %AdvPS \* Salest+1
* Current liabilities
  + CLiabt

= STLoant + NLiabt / (AvgM – 1)

where

* + - STLoant = max(0, STCashDeficitt)
    - STCashDeficitt = TLt – TLt-1 – Operating NLBt + STLoant-1 \* (1 + %AvgSTInt) – %MSReturn \* IMSt-1
    - STLoant-1 = CLiabt-1 - NLiabt-1 / (AvgM – 1)
    - Operating NLBt = ARt-1 + AdvPSt + Salest \* (1 - %AR - %AdvPS) – (APt-1 + AdvPPt + Purchasest \* (1 - %AP - %AdvPP) + OpExt + ITt)
    - APt-1 = %AP \* Purchasest-1
    - AdvPPt = %AdvPP \* Purchasest+1
    - OpExt = %OR \* Salest + OBt-1 \* (1 + Inflationt)
    - ITt = %IT \* NIt
    - ARt-1 = %AR \* Salest-1
    - AdvPSt = %AdvPS \* Salest+1
* Non-current liabilities
  + NLiabt

= (LTLoant + NLiabt-1) \* (AvgM – 1)/AvgM

where

* + - LTLoant = (max(0, STCashDeficitt – STLoant – CapEx NLBt + NLiabt / (AvgM – 1) + AvgLTIntt + Dividendst-1 + Stock buybackst) \* (1 - %EF)
    - CapEx NLBt = NCAt-1 \* %Depr + %AG \* Salest
    - Dividendst-1 = %PR \* NIt-1
    - NIt-1 : net income from last year

= (1 - %IT) \* (Salest-1 – COGSt-1 – OpExt-1 – Depreciationt-1 – AvgSTIntt-1 – AvgLTIntt-1 + %MSReturn \* (1 - %Cash) \* TLt-1)

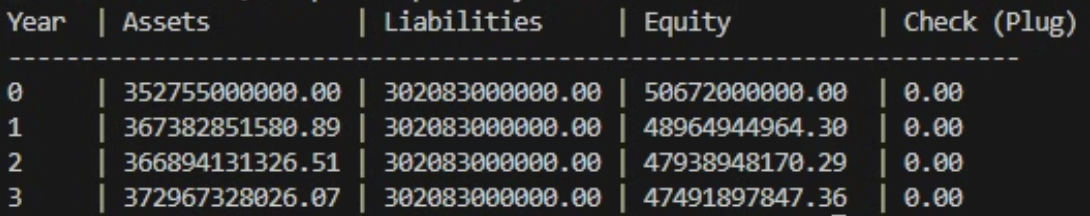
* + - TLt-1 = Total liquidity last year.
    - COGSt-1 = %Inv \* (Salest-2 – Salest-1) + Purchasest-1
    - OpExt-1 = %OR \* Salest-1 + OBt-2 \* (1 + Inflationt-1)
    - Depreciationt = NCAt-1 \* %Depr
    - AvgSTIntt = %AvgSTInt \* (CLiabt-t – NLiabt-t / (AvgM – 1))
    - AvgLTIntt = %AvgLTInt \* (NLiabt-1 / (1 – 1/AvgM))
    - Stock buybackst = %BB \* Depreciationt
  1. Equity
* Stockholder equity
  + SEt

= SEt-1 + LTLoant / (1 - %EF) \* %EF + NIt – %PR \* NIt-1 – %BB \* %Depr \* NCAt-1

We see in all elements’ equations that they are driven by Sales and Purchases and elements from previous years, indicating no circularity or plugs. The net liquidity balance calculation and the new short-term and long-term borrowing equations ensure that the assets and liabilities + equity are increased/decreased the same. Thus, assets = liabilities + equity will always be satisfied. We can check with test values in our simple TensorFlow model.

This forward model has been implemented in Python in **simple\_financial\_model.py** using data from Apple’s financial statements. The financial statements were first obtained from Yahoo finance as suggested by Question 4 (we wrote the script in **yahoofinance.py**), but we could only access past 4 years of data, which is not sufficient for training in the following section. We managed to obtain older financial data by directly pulling from Apple’s 10-K filings. The script that we wrote and used for the data extraction is found in **sec-edgar-download-extractor.py.**

The following is the output of the code when run with: python .\simple\_financial\_mode.py :



Even without using any plugs, we see that the assets = liabilities + equity identity is satisfied because our model follows the double-entry principle with any inflows and outflows on the balance sheet as done by Pareja09.

1. **Training the Model**

Out of the variables to be trained, there are quite a few that are easily trained independently from other variables by looking at balance sheets and income statements. We first note the two drivers for our model:

* Salest : Total revenue in balance sheet
* Purchasest : COGSt + (Invt – Invt-1)

We can train the rest of the variables like the following:

* %AG : Learn from change in all non-current assets in balance sheet
* %Depr : Learn from depreciation in income statement and non-current assets in balance sheet
* %AR : Learn from Sales and account receivables data in balance sheet
* %AP : Learn from Purchases and account payables data in balance sheet
* %AdvPS : Learn from current deferred revenue data and next year’s Sales in balance sheet
* %AdvPP : Learn from Purchases and current other assets in balance sheet
* %Inv : Learn from inventory data in balance sheet
* %Cash : Learn from short-term investments in balance sheet
* %TL : Learn from revenue and Cash Equivalents And Short Term Investments in balance sheet
* %IT : Learn from net income and tax provision in income statement
* %OR, OBT\_start : Learn from inflation data and operating expense in income statement
* %PR : Learn from common stock dividends in Cash Flow and net income in income statement
* %BB : Learn from Repurchase Of Capital Stock and depreciation in Cash Flow

Each of the variables above can be related to an element from Cash Flow, Balance Sheet, or Income Statement, and can be trained with a simple mean-square-error minimization for each.

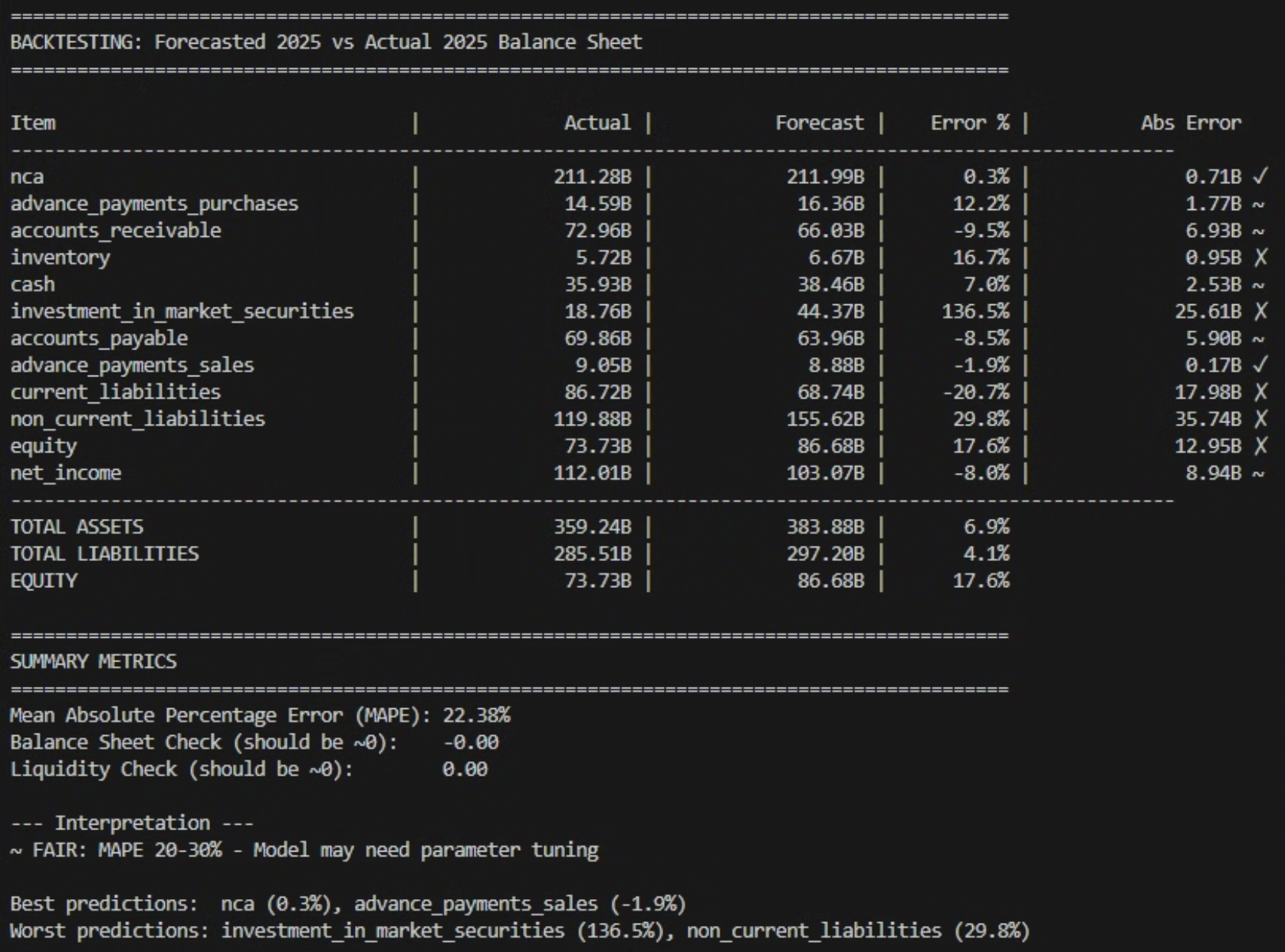
The following variables (i.e. structural parameters) can be trained together using the above trained variables:

* %AvgSTInt : Average short-term interest on the effective short-term loan
* %AvgLTInt : Average interest on total long-term liabilities
* AvgM : Average maturity of total long-term liability
* %MSReturn : Percentage return on the investment in market securities
* %EF : Equity financing as a percentage of total excess cash needed after accounting for long-term investments and long-term obligations

Using net income, stockholders equity and current and non-current liabilities data, we can train AvgM, %AvgSTInt, %AvgLTInt, %MSReturn, and %EF concurrently via simple gradient while minimizing mean-square error between predicted and actual net income, stockholders equity, current liabilities, and non-current liabilities.

The code has been implemented in trainable\_financial\_model.py.

In order to test the model’s prediction capability, we train the parameters from Apple’s 2018~2024 data, predict 2025 data, and compare it with the actual data. The following is the output after running **trainable\_financial\_model.py**:



It shows a modest prediction ability, which is unsurprising because we have lumped various balance sheet elements together to form several elements to be more consistent with models described in Pareja07 and Pareja09, as we have explained in **1. The Balance Sheet Elements**. Also, there is a limitation of assuming a liquidity balance (cash + investment in market securities) that varies linearly with sales, especially for a cash-rich company like Apple whose cash balance and stock buybacks are governed by the recent buyback programs in order to reduce its cash balance instead of being proportional to asset depreciation, which is done here as done by Pareja09.

Our model can also forecast earnings because net income is automatically calculated when predicting the elements on the balance sheet, and is outputted as “net\_income” as shown above.

1. **Applying Machine Learning**

As our model is rather a simple model based on Pareja09, there is a lot of improvements that could make the model more performant. A short list of potential improvements is as follows:

* Predicting purchases and sales of future years using Long Short-Term Memory and macroeconomic indicators (e.g., inflation, interest rate)
* Implementing Bayesian model to account for probability distribution of accounting variables (e.g., company policy for capital expenditure, operating expense prediction, cash balance targets) to account for the noise term, n(t).
* Implementing neural networks for each accounting variables that take into consideration macroeconomic signals (e.g., incorporate consumer price index, GDP growth, and interest rates to predict advance purchases percentage)
* Incorporating the company’s stock buyback target if it has been disclosed publicly into training the buyback ratio with a neural network

For our model, we choose to implement Bayesian learning with Variational Inference method to train the operational expense because

1. OpEx is influential in the net income value
2. OpEx, in our model, is heavily influenced by sales revenue, which is one of the main drivers in our model, up to 23%

Applying Variational Inference allows us to calculate the posterior distribution of OpEx given the past OpEx, sales, and inflation data instead of a singular value that a simple linear regression gives you. It allows us to predict the OpEx, and therefore the earnings and other balance sheet elements, with a confidence interval.

Most of the code from trainable\_financial\_model.py remains unchanged except how we model the OpEx during training. The following updates are made:

1. Bayesian Conversion: Replaced the deterministic %OR and OBT\_start variables with Variational Posteriors.

* We learn a Mean (µ) and a Standard Deviation(σ) for each of these variables.
* During training/inference, we sample their values from their distributions:
  + Value = µ + σ \* ϵ, where ϵ ～𝒩(0, 1)

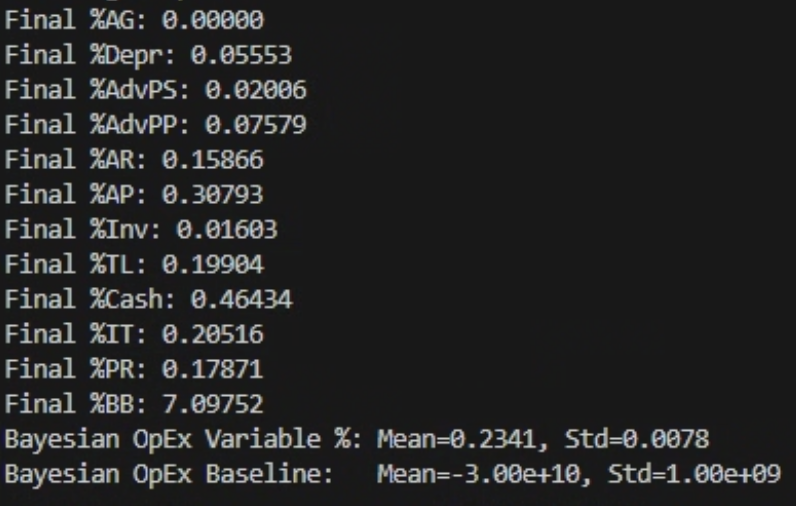
1. Loss Function: Replaced the simple Mean Squared Error (MSE) for OpEx with Negative Log Likelihood + KL Divergence.

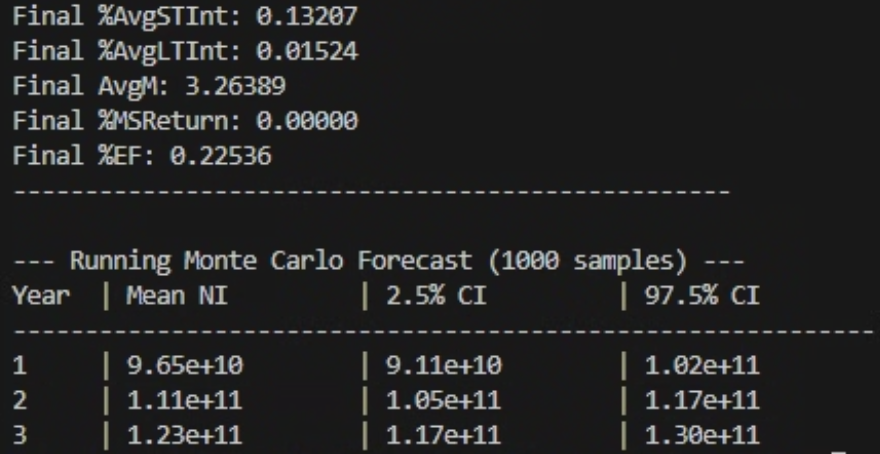
* This balances fitting the data (Likelihood) with keeping the distribution close to our prior beliefs (KL).

1. Monte Carlo Simulation: Added a run\_monte\_carlo\_forecast function. Instead of predicting one single Net Income line, we run the forecast 1,000 times, sampling different OpEx values each time, to generate Confidence Intervals.

So, to answer the ‘Hint’, we are improving the model such that when we predict the earnings (i.e. net income), we can now model n(t) with the confidence intervals given by the Monte Carlo simulation. After we have trained all the variables, they become the very function that take in this year’s y(t) and input x(t) to predict y(t+1). The input x(t) includes the primary drivers, which are sales and purchases time series, and inflation, which is the external environment variable that we include in the model for predicting the operating expense.

We implemented the updated Bayesian model in **trainable\_financial\_model\_bayesian\_vi.py**, and the output is reproduced below:





References

[Pareja(09)] Vélez Pareja, Ignacio, (2009), “Constructing Consistent Financial Planning Models for Valuation” (August 15), Available at SSRN: <http://ssrn.com/abstract=1455304>

[Pareja(07)] Vélez Pareja, Ignacio, (2007), “Forecasting Financial Statement with No Plugs and No Circularity”, Available at SSRN: <http://ssrn.com/abstract=1031735>

[Benninga(06)]