

UTILIZING NATURAL LANGUAGE PROCESSING AND OTHER TECHNICAL TOOLS TO
AUTOMATE THE FINANCIAL MODELING PROCESS WHILE MAINTAINING ROBUSTNESS
IN THE MODEL

© 2017
By: Robert Forrest Short

A thesis submitted to the faculty of The University of Mississippi in partial fulfillment of
the requirements of the Sally McDonnell Barksdale Honors College.

Oxford
May 2017

Approved by:

Advisor: Dr. Dawn Wilkins

Reader: Dr. Naeemul Hassan

Reader: Dr. Yixen Chen

Abstract

The objective of this thesis is to improve upon the most commonly utilized, entirely manual financial modeling process. This thesis gives guidance for implementing a set of technical tools that can be utilized in an individual's financial modeling process to automate repetitive and predictable tasks. The goal is not to provide a robust set of tools that can be implemented across all financial models; in fact, this may not be possible. The goal is to provide a basis for anyone who utilizes financial modeling to study their own modeling processes and adapt the given tools to fit into and improve their unique process.

Table of Contents

1. Financial Modeling	1
1.1 Overview	1
1.2 Modeling Process	2
1.3 Three-Statement Modeling Overview	3
2. Automation	5
2.1 Step One Overview: The Income Statement and Balance Sheet	5
2.2 Natural Language Processing	8
2.3 Python for NLP	9
2.4 Generally Accepted Accounting Principles	10
2.5 Step One: Automation	12
2.6 Step Two: Forecasting the Income Statement	16
2.7 Step Two: Forecasting the Balance Sheet	21
2.8 Step Three: Generating the Cash Flow Statement	24
3. Summary	30

List of Figures

Figure 1	Modeling Tree	3
Figure 2	Income Statement Model	6
Figure 3	Income Statement (10-K)	6
Figure 4	Balance Sheet Model	7
Figure 5	Balance Sheet (10-K)	8
Figure 6	Python Income Statement Extraction	13
Figure 7	Resulting Excel File	13
Figure 8	Corresponding 10-K	14
Figure 9	Forecasting Ratios	17
Figure 10	Income Statement Model with Projection	18
Figure 11	Revenue Projection Code	19
Figure 12	Revenue Projection Test	19
Figure 13	Growth Rates from Projections	21
Figure 14	Accounts Receivable Projections	22
Figure 15	Balance Sheet Projections Code	23
Figure 16	Cash Flow Statement Model	26
Figure 17	Cash Flow Direct Retrieval	27
Figure 18	Cash Flow Indirect Retrieval	27

Chapter 1

Financial Modeling

1.1 Overview

Financial modeling, on a high level, is the task of building a quantitative, mathematical representation of a real-world financial asset or situation. However, on a granular level, financial modeling is comprised of a set of predictable, repetitive tasks that coincide with the principles of corporate finance and in particular, the principles of accounting and valuation. The goal of this thesis will be to identify a subset of these tasks and offer a basis for their automation. This will begin with a study of basic modeling concepts and then lead into the technical techniques which could be highly effective if utilized properly within the financial modeling process.

Financial modeling, while used to represent a multitude of real world financial situations, has a singular purpose: to value value. When analyzing an investment, modeling a merger or acquisition or analyzing a company's operating performance the purpose of the analysis is always to identify value and understand how to acquire or grow that value and at what price. The price of value is dependent on the risk associated with it. Investments are made with the assumption that they will provide a return that is commensurate with or exceeds the risk taken. "The guiding principle of value creation is that companies create value by investing capital they raise from investors to generate future cash flows at a rate of return exceeding the cost of capital (the rate investors require to be paid for the use of their capital)." [3] Financial models are tools utilized in the process of estimating value, commonly known as valuation.

1.2 Modeling Process

On a high level, the financial modeling process can be broken down into four steps: (1) Understand the purpose of the analysis. Financial models are created to quantify value and analyze assets with the purpose of making informed decisions in a variety of situations. It is, therefore, imperative to understand the purpose of a model as it will shape the way in which, and what kind of analysis is done; (2) Understand the company or asset's history, the industry in which it competes and, most importantly, identify the key drivers that impact its performance. Significant drivers will vary between industries as well as between companies within specific industries, but, for the most part, the bulk of a company's performance can be attributed to five or fewer key drivers; (3) With an understanding of a company's history, performance, drivers, and industry, build the analysis in Microsoft Excel, and based on an understanding of the company and industry, make the necessary assumptions and extrapolate historical data into the future; (4) Based on the analysis and an understanding of what you expect the company to look like in the future, draw a conclusion that coincides with the purpose.

[9]

Based on this four-step process alone, one can tell that financial modeling is a process that involves a significant amount of assumptions that can be dependent on a significant number of factors. For that reason, they are often made on a case-by-case basis by financial professionals. The automation of this aspect of modeling is outside the scope of this thesis. Instead, this thesis will explore the tasks involved in the creation of the fundamental, three-statement model that are automatable and how their automation can be approached. This will involve a study of the financial statements

released by companies and how necessary data can be retrieved, organized and exported from those statements given their unstructured nature, as well as a look into working with this data in Excel in a semi-automated fashion. To begin, we will study the three-statement modeling process and identify the steps within the process that can theoretically be automated.

1.3 Three-Statement Modeling Overview

A financial statement model, also known as a three-statement model, is the foundational model from which other common models and analysis are developed. It is for this reason that it will be the primary focus of this thesis.

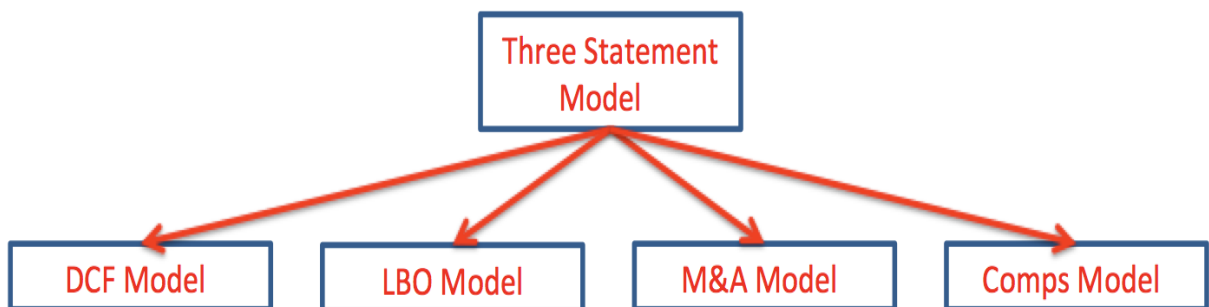


Figure 1: Modeling Tree

Figure 1 is representative of the financial modeling hierarchy, which is rooted in the three-statement model. A three-statement model is a representation of a financial asset's historical performance and, more importantly, an extrapolation of historical performance into the future. It is with these extrapolated numbers that a discounted cash flow (DCF) model, leveraged buyout (LBO) model, merger and acquisition (M&A) model or a comparable companies (Comps) model can be created in an attempt to quantify the value of an asset within particular financial situations.

The three-statement modeling process can be broken down into three steps:

- (1) Inputting historical data from a company's income statement and balance sheet.
- (2) Projecting this historical data into the future, a process to which there is no singular approach; however, using common financial ratios is the most common method.
- (3) Generate a working cash flow statement from the projected income statement and balance sheet. This is a process in which a company's beginning cash balance, taken from the previous year's balance sheet, is reconciled with the ending cash balance, reflected on the current year's balance sheet. Creating the operating section of the cash flow statement, where this thesis will more directly focus, involves taking net income from the income statement, adding back non-cash charges, subtracting out capitalized uses of cash and accounting for changes in operations-based balance sheet accounts to arrive at the ending cash balance that is reflected on the current year's balance sheet.

In chapter two, we will begin to discuss the automation process that was found to be most effective during research and the three steps just presented will act as an outline. The most important requirements of each step will be discussed and relevant examples will be given along with a theory and basis for how certain processes, within each step, could be automated.

Chapter 2

Automation

This thesis will look at techniques to automate the first and third steps in the three-statement modeling process and will offer guidance around step two. While step two will be discussed, the conclusion has been reached, through research, that in order to create robust, professional models it is a step that must involve significantly more human interaction than its counterparts. The reasoning behind this conclusion will be discussed later, but we will begin by walking through the processes involved in step one, which is a unique step in that it requires extracting data from unstructured data, i.e. a company's financial statements.

2.1 Step One Overview: The Income Statement and Balance Sheet

To begin, the income statement is a financial statement that shows a company's financial performance for a given period of time (usually a year). The income statement takes all of a company's expenses, for a given period of time, and subtracts them from that same period's revenues to arrive at net income, also known as the bottom line.

Figure 2 shows the income statement, as it would be presented in a three-statement model, after the completion of step one.

INCOME STATEMENT			
Fiscal year	2011A	2012A	2013A
Fiscal year end date	<i>9/24/11</i>	<i>9/29/12</i>	<i>9/28/13</i>
Revenue	108,249	156,508	170,910
Cost of sales	(64,431)	(87,846)	(106,606)
Gross Profit	43,818	68,662	64,304
Research & development	(2,429)	(3,381)	(4,475)
Selling, general & administrative	(7,599)	(10,040)	(10,830)
Operating profit (EBIT)	33,790	55,241	48,999
Interest income	519	1,088	1,616
Interest expense	0	0	(136)
Other expense	(104)	(566)	(324)
Pretax profit	34,205	55,763	50,155
Taxes	(8,283)	(14,030)	(13,118)
Net income	25,922	41,733	37,037
Basic shares outstanding	924	935	925
Impact of dilutive securities	12	11	6
Diluted shares outstanding	937	945	932
Basic EPS	\$28.05	\$44.64	\$40.03
Diluted EPS	\$27.68	\$44.15	\$39.75

Figure 2: Income Statement Model

In contrast, Figure 3 shows the financial statement from which the data was taken:

	Years ended		
	September 28, 2013	September 29, 2012	September 24, 2011
Net sales	\$ 170,910	\$ 156,508	\$ 108,249
Cost of sales	106,606	87,846	64,431
Gross margin	64,304	68,662	43,818
Operating expenses:			
Research and development	4,475	3,381	2,429
Selling, general and administrative	10,830	10,040	7,599
Total operating expenses	15,305	13,421	10,028
Operating income	48,999	55,241	33,790
Other income/(expense), net	1,156	522	415
Income before provision for income taxes	50,155	55,763	34,205
Provision for income taxes	13,118	14,030	8,283
Net income	\$ 37,037	\$ 41,733	\$ 25,922
Earnings per share:			
Basic	\$ 40.03	\$ 44.64	\$ 28.05
Diluted	\$ 39.75	\$ 44.15	\$ 27.68
Shares used in computing earnings per share:			
Basic	925,331	934,818	924,258
Diluted	931,662	945,355	936,645
Cash dividends declared per common share	\$ 11.40	\$ 2.65	\$ 0.00

Figure 3: Income Statement (10-K)

When comparing Figures 2 and 3, it becomes evident that the model is essentially a copy of the published income statement. In fact, some well-respected modeling instructors suggest copying and pasting data from the published financial statements as a way of improving accuracy and efficiency. It is a process that is repetitive, simple and can be automated.

Continuing, the balance sheet is, again, a financial statement but it summarizes a company's assets, liabilities and owner's equity. The balance sheet, unlike the income statement, is a moment in time statement meaning that it shows a snapshot of a company's holdings at a specific point in time as opposed to showing the company's performance over a given period of time. One key point about the balance sheet is that it must at all times remain "balanced" which is determined by the accounting equation (assets = liabilities + stockholder's equity).

The balance sheet, as presented in three-statement model after the completion of step one, is shown in Figure 4, and it is contrasted with the financial statement data, shown in Figure 5, from which it was drawn.

BALANCE SHEET			
<i>Fiscal year</i>		2012A	2013A
Fiscal year end date		9/29/12	9/28/13
Cash & equivalents ST & LT market. securities		121,251	146,761
Accounts receivable		10,930	13,102
Inventory		791	1,764
Deferred tax assets		2,583	3,453
Other current assets (inc. non-trade receivables)		14,220	14,421
Property, plant & equipment		15,452	16,597
Acquired intangible assets (inc. Goodwill)		5,359	5,756
Other assets		5,478	5,146
Total assets		176,064	207,000
Accounts payable		21,175	22,367
Accrued expenses & def rev. (current & non-current)		20,015	23,916
Revolver		0	0
Long term debt		0	16,960
Other non-current liabilities		16,664	20,208
Total liabilities		57,854	83,451
Common stock / additional paid in capital		16,422	19,764
Treasury stock		0	0
Retained earnings / accumulated deficit		101,289	104,256
Other comprehensive income / (loss)		499	(471)
Total equity		118,210	123,549
<i>Balance check</i>		0	0

Figure 4: Balance Sheet Model

	September 28, 2013	September 29, 2012
ASSETS:		
Current assets:		
Cash and cash equivalents	\$ 14,259	\$ 10,746
Short-term marketable securities	26,287	18,383
Accounts receivable, less allowances of \$99 and \$98, respectively	13,102	10,930
Inventories	1,764	791
Deferred tax assets	3,453	2,583
Vendor non-trade receivables	7,539	7,762
Other current assets	6,882	6,458
Total current assets	73,286	57,653
Long-term marketable securities	106,215	92,122
Property, plant and equipment, net	16,597	15,452
Goodwill	1,577	1,135
Acquired intangible assets, net	4,179	4,224
Other assets	5,146	5,478
Total assets	<u>\$ 207,000</u>	<u>\$ 176,064</u>
LIABILITIES AND SHAREHOLDERS' EQUITY:		
Current liabilities:		
Accounts payable	\$ 22,367	\$ 21,175
Accrued expenses	13,856	11,414
Deferred revenue	7,435	5,953
Total current liabilities	43,658	38,542
Deferred revenue – non-current	2,625	2,648
Long-term debt	16,960	0
Other non-current liabilities	20,208	16,664
Total liabilities	<u>83,451</u>	<u>57,854</u>
Commitments and contingencies		
Shareholders' equity:		
Common stock, no par value; 1,800,000 shares authorized; 899,213 and 939,208 shares issued and outstanding, respectively	19,764	16,422
Retained earnings	104,256	101,289
Accumulated other comprehensive income/(loss)	(471)	499
Total shareholders' equity	<u>123,549</u>	<u>118,210</u>
Total liabilities and shareholders' equity	<u>\$ 207,000</u>	<u>\$ 176,064</u>

Figure 5: Balance Sheet (10-K)

It is once again noteworthy that Figures 4 and 5, despite some consolidation within the model, are essentially replicas of each other. The consolidation of certain line items will be revisited and addressed later.

In summary, step one of the three-step modeling process is comprised of creating these two statements within Excel, a process which can be, at a minimum, partly automated using basic natural language processing techniques. Understanding these methods involves nothing more than basic knowledge of natural language processing and techniques commonly used for data retrieval.

2.2 Natural Language Processing

It is commonly believed that the beginning of the study of natural language processing (NLP) began in the 1950's bookmarked by the release of an article by Alan

Turing entitled “Computing Machinery and Intelligence.” Natural language processing is a field of data science that is utilized to work with text and unstructured data, where unstructured data is “data that does not have a pre-defined model or is not organized in a predefined manner.” [1] Natural language processing is a technique that uses a computer to process “natural language” and, through the identification of patterns, topics, and entities, provide structure to the data. The goal is to, more or less, interpret natural language as a human would. Natural language refers to the language that is used for communication by humans; it is the spoken and written language that we use every day. Natural language, which is unstructured, can be contrasted with programming languages, which can be described using a set of explicit rules. There are many techniques in NLP such as word statistics, entity extraction, topic identification, and sentiment analysis that are utilized to process and interpret large volumes of unstructured text.

2.3 Python for NLP

Python is described as an “interpreted, object-oriented, high-level programming language with dynamic semantics,” [1] and is often used for NLP because of the excellent functionality of the Python Natural Language Processing Toolkit (NLTK). The NLTK was created in 2001 in the Department of Computer and Information Science at the University of Pennsylvania. It is an open source project and therefore, since its creation it has been expanded on and developed by numerous contributors. The development of the NLTK has been guided by four principles: (1) Simplicity: building an intuitive framework providing a practical knowledge of NLP without getting bogged down in tedious “house-keeping.” (2) Consistency: To provide a uniform framework with consistent interfaces and data structures, and easily guessable methods. (3) Extensibility: to provide a structure into which new

software modules can be easily accommodated. (4) Modularity: to provide components that can be used independently without needing to understand the rest of the toolkit. [1]

Possibly the most commonly utilized technique in NLP is used every day by search engines, like Google. This technique finds desired information, within unstructured text, and presents the information. The goal of the natural language processing utilized in this thesis will be similar, in that, we will be searching for specific information found in unstructured text with the goal of extracting it automatically. We will, therefore, be focused on NLP functionality that enables extracting specific information from text as opposed to other functionality such as analyzing sentence structure or attempting to analyze the meaning of a sentence. Before looking at NLTK tools that can be used for data retrieval, it is necessary to study the generally accepted accounting principles (GAAP) that provide some structure to what is otherwise, unstructured data. These principles make it possible to accurately identify and capture the desired income statement and balance sheet data.

2.4 Generally Accepted Accounting Principles

Publicly traded companies, in accordance with regulations set forth by the Securities and Exchange Commission, release annual financial statements, commonly known as 10-K's. Within these financial statements, companies are required to include a consolidated balance sheet and income statement, amongst other items. This relevant financial data is set forth, in a structured manner, in accordance with generally accepted accounting principles. However, financial statements also include footnotes and management analysis that can be extensive in nature; it is not uncommon for a statement to be well over fifty pages. Despite a large amount of unstructured text, the structured nature of the financial components relevant to our purposes can be utilized to increase the accuracy of data extraction and allow for a single

code base to be utilized across all financial statements. This process, therefore, involves a study of Generally Accepted Accounting Principles and identifying, within an expansive guide of principles, known as the codification, which ones can be used to our advantage in extracting income statement and balance sheet data.

The guidelines that proved to be the most utilitarian in this process included ASC 205, *Presentation of Financial Statements*, and more specifically, ASC 205-10-45-1A, which states there are five things that must be shown during a reporting period. Of these five things, the one most relevant for our purposes states “Earnings (net income) for the period, (which may be presented as a separate statement or within a continuous statement of comprehensive income.” [12] This ensures that when parsing and extracting data from a company’s presentation of the income statement, net income can be used as an indicator that the final line has been reached. Unfortunately, there is no requirement regulating the title of the income statement. In fact, larger corporations very rarely use “income statement” as the title. Titles that are more common include “consolidated statement of operations”, “consolidated statement of income”, “consolidated statement of earnings” and “consolidated results of operations.” [12] It is, therefore, necessary to utilize a variety of “flags” when searching for the first line of the income statement. These flags, as will be shown, are sufficient for accurately retrieving income statement data using natural language processing techniques. Furthermore, ASC 210, *Balance Sheet*, provides guidance for the balance sheet presentation of all US GAAP reporting entities. There are certain titles required to be within the statement in some form according to S-X 5-02, which include: “Total Assets,” “Total Liabilities,” “Total Equity”, and lastly, “Assets” at the onset of the statement. If the balance sheet, shown previously in Figure 5, is referred back to, it becomes evident that these labels act as section

separators within the financial statement and are, therefore, adequate “flags” for data extraction. It should be noted that GAAP is a very complex, extensive document that is constantly changing, and the analysis done above is a highly simplified breakdown of GAAP principles. In fact, it does not even begin to scratch the surface of GAAP’s vastness nor does it provide adequate insight into its complexity. However, as will be partially demonstrated, the identified flags are sufficient for extracting the desired balance sheet and income statement information.

2.5 Step One: Automation

Figures 6 and 7 contain a code snippet utilizing these flags and the python NLTK to extract the desired income statement financial data as well as a screenshot of the resulting Excel file:


```

import nltk
import xlswriter
from urllib import urlopen
from bs4 import BeautifulSoup

#Reading in 10-K from online SEC filing
#All filings can be found on SEC.gov or on specific company websites
url = "http://investor.apple.com/secfiling.cfm?filingid=1193125-14-383437"
html = urlopen(url).read()

#Beautiful soup it used to get the text while ignoring the html
web_str = BeautifulSoup(html).get_text()

#Utilizing markers, as discussed, to capture Income statement data
#Net income can also be utilized to mark the end of the Income statement
start = web_str.find("CONSOLIDATED STATEMENTS OF OPERATIONS")
end = web_str.find("Cash dividends declared per common share")
last_sent = len("Cash dividends declared per common share")
is = web_str[start:end+last_sent]

#Tokenizing the Income Statement (IS) data and formatting it
#The tokenized data is shown below
is_tokens = nltk.word_tokenize(is)
isCommasRemoved = bs.replace(',', '')

# [u'CONSOLIDATED', u'STATEMENTS', u'OF', u'OPERATIONS', u'(In', u'millions', u'except', u'number', u'of', u'shares',
# u'which', u'are', u'reflected', u'in', u'thousands', u'and', u'per', u'share', u'amounts)', u'Years', u'ended',
# u'September', u'27', u'2014', u'September', u'28', u'2013', u'September', u'29', u'2012', u'Net', u'sales', u'$',
# u'182795', u'$', u'170910', u'$', u'156508', u'Cost', u'of', u'sales', u'112258', u'106606', u'87846', u'Gross',
# u'margin', u'70537', u'64304', u'68662', u'Operating', u'expenses:', u'Research', u'and', u'development', u'6041',
# u'4475', u'3381', u'Selling', u'general', u'and', u'administrative', u'11993', u'10830', u'10040', u'Total',
# u'operating', u'expenses', u'18034', u'15305', u'13421', u'Operating', u'income', u'52503', u'48999', u'55241',
# u'Other', u'income/(expense)', u'net', u'980', u'1156', u'522', u'Income', u'before', u'provision', u'for', u'income',
# u'taxes', u'53483', u'50155', u'55763', u'Provision', u'for', u'income', u'taxes', u'13973', u'13118', u'14030',
# u'Net', u'income', u'$', u'39510', u'$', u'37037', u'$', u'41733', u'Earnings', u'per', u'share:', u'Basic',
# u'$', u'6.49', u'$', u'5.72', u'$', u'6.38', u'Diluted', u'$', u'6.45', u'$', u'5.68', u'$', u'6.31', u'Shares',
# u'used', u'in', u'computing', u'earnings', u'per', u'share:', u'Basic', u'6085572', u'6477320', u'6543726', u'Diluted',
# u'6122663', u'6521634', u'6617483', u'Cash', u'dividends', u'declared', u'per', u'common', u'share']

#Extracting the numbers from the data
is_array_numbers = [int(s) for s in commasRemoved.split() if s.isdigit()]

#Writing the data to an excel workbook
workbook = xlswriter.Workbook('arrays.xlsx')
worksheet = workbook.add_worksheet()

array = [bs_array[i:i+3] for i in range(0, len(is_array_numbers), 3)]
del array[:2]

row = 1

for col, data in enumerate(array):
    worksheet.write_row(col, row, data)

workbook.close()

```

Figure 6: Python Income Statement Extraction

	A	B	C	D
1		182795	170910	156508
2		112258	106606	87846
3		70537	64304	68662
4		6041	4475	3381
5		11993	10830	10040
6		18034	15305	13421
7		52503	48999	55241
8		980	1156	522
9		53483	50155	55763
10		13973	13118	14030
11		39510	37037	41733
12		6085572	6477320	6543726
13		6122663	6521634	6617483

Figure 7: Resulting Excel File

Furthermore, in Figure 8, is the 10-K from which the data, shown in Figure 7, was extracted. As is evident and as would be expected, the numbers are identical. The code above does not import the line labels, shown on the left-hand side of Figure 8, into Excel because line labels are often changed in models. However, the line labels can be exported with a slight modification to the above code.

Net sales	\$ 182,795	\$ 170,910	\$ 156,508
Cost of sales	112,258	106,606	87,846
Gross margin	70,537	64,304	68,662
Operating expenses:			
Research and development	6,041	4,475	3,381
Selling, general and administrative	11,993	10,830	10,040
Total operating expenses	18,034	15,305	13,421
Operating income	52,503	48,999	55,241
Other income/(expense), net	980	1,156	522
Income before provision for income taxes	53,483	50,155	55,763
Provision for income taxes	13,973	13,118	14,030
Net income	\$ 39,510	\$ 37,037	\$ 41,733

Figure 8: Corresponding 10-K

There are some important formatting differences to note between the Excel models, shown previously in Figures 4 and 2, which were created manually in Excel and the Excel file created automatically using Python, shown in Figure 7. All of these differences will not be addressed; however, I will address the differences that hurt the integrity of the model, starting with the lack of hard-coded cells. As can be seen in the previous figures of the financial statements created manually, some data is blue and other data is black. A black font color indicates that the cell contains a formula; a blue font indicates that the data was hard-coded. The data extracted and exported into Excel is, of course, all going to initially be hard-coded. For example, the line item “Net income” from the income statement, shown at the bottom of Figure 8 and represented in row 11 of Figure 7, will not contain a formula subtracting expenses from revenues as it should; the numbers will instead be hard-coded into the cells. Therefore, it is necessary, within the Excel file created using Python, to replace

hard-coded data, in line items such as “Net Income”, with appropriate formulas. This will allow Excel to calculate this data and for the model to be robust and reactive.

Another important difference to note is the fact that some line items have been consolidated. In particular, the extracted income statement data includes a single line item entitled “Other income/(expense), net”, shown in Figure 8, while the income statement created manually, shown in Figure 2, splits a similar line into three separate lines entitled “Interest Income”, “Interest Expense” and “Other Expense”. Splitting this line item into three separate items involved digging through the company’s financial statements and finding the footnote that specified what individual sources of income and expense comprised the single line item and breaking it down into its component parts. The breaking down of this line is necessary but understanding the reasoning behind it is not crucial to this thesis.

These differences are pointed out with the purpose of supporting the conclusion, hinted at earlier and pertaining to the automation of financial modeling. Through research, it has been concluded that due to factors including, but not limited to, differing financial statement line consolidation from company to company, varying drivers of growth from company to company and, to a greater extent, industry to industry and the absence of a standardized method of predicting future results, it is necessary for human interaction to facilitate the modeling process. While technical tools can be utilized for data extraction as well as many other modeling tasks, some of which will be demonstrated, it is necessary for someone, with modeling expertise, to perform tasks such as studying the format of a company’s financials, as they were presented in the 10-K, and understanding what line items must be broken out for the model to function properly upon completion. As we continue to step two of the three-statement modeling process, the necessity of human interaction will

become more evident. There are, however, steps in the process that can utilize available tools, leading to a “semi-automated” modeling process that is more accurate and efficient than the process currently utilized by financial professionals.

As has already been demonstrated, simple natural language processing techniques can be utilized for quick and efficient data extraction. While some human interaction may still be necessary at the onset of this extraction, that does not counter the fact that efficiency and accuracy have already been improved. Simple tools to quickly correct the problems discussed above, including the lack of hard-coded cells and the incorrect color scheme, could be created, but the process of creating them is relatively simple and not crucial to the goals of this thesis and, therefore, we will move forward to step two in the three-statement modeling process, projecting historical data into the future.

2.6 Step Two: Forecasting the Income Statement

Once historical data has been entered into the model, it becomes necessary to forecast this data into the future. This step is necessary because whether a financial model is being used to analyze an investment, model a merger or acquisition or analyze a company’s operating performance, the purpose of the analysis is always to identify value and understand how to acquire or grow that value. Therefore, a model that does not take into account financial results beyond the reported historical periods implies that there is no value in future earnings. While it is true that, due to the time value of money, earnings today are more valuable than earnings that will be realized five years from now, it does not take an expert to understand that these future earnings are worth something, and in order to approximate that value, a company’s performance must be projected into the future.

The most common method for forecasting financial statements into the future is to study a company's historical performance and with a working knowledge of the industry at hand and an understanding of the drivers behind operating performance, develop a set of forward-looking financial ratios that. These will include ratios such as revenue growth rates, gross profit margins, operating margins, as well as margins relating to certain expenses such as research and development expenses. Creating these financial ratios for forecasting is a standard process in modeling, but there is variance in the process of determining appropriate ratio values for future periods.

The forecasting process can be demonstrated using the “Growth rates & margins” table associated with the income statement model we have been utilizing.

<u>Growth rates & margins</u>								
Revenue growth	NA	44.6%	9.2%	5.0%	5.0%	4.5%	4.9%	5.4%
Gross profit as % of sales	40.5%	43.9%	37.6%	37.1%	37.4%	37.4%	37.4%	37.4%
R&D margin	2.2%	2.2%	2.6%	2.9%	3.1%	3.1%	3.1%	3.1%
SG&A margin	7.0%	6.4%	6.3%	6.8%	6.8%	6.8%	6.8%	6.8%
Tax rate	24.2%	25.2%	26.2%	26.3%	26.0%	26.0%	26.0%	26.0%

Figure 9: Forecasting Ratios

As can be seen in Figure 9, each row within the “Growth rates & margins” table corresponds to a specific line item in the income statement, shown in Figure 2. The first three lines from the left show historical data (historical growth rates and margins from the most recent years) and the remaining lines show projections (what these growth rates and margins are expected to be in future years). The projections are utilized to forecast their respective line items into the future. The resulting income statement, after the completion of this step, is shown in Figure 10.

F16		=E16*(1+F37)				
A	B	F	G	H	I	J
12	INCOME STATEMENT					
13	Fiscal year	2014P	2015P	2016P	2017P	2018P
14	Fiscal year end date	9/30/14	9/30/15	9/30/16	9/30/17	9/30/18
15						
16	Revenue	179,381	188,389	196,830	206,514	217,643
17	Cost of sales	(112,830)	(117,932)	(123,216)	(129,278)	(136,245)
18	Gross Profit	66,550	70,458	73,614	77,236	81,398
19	Research & development	(5,202)	(5,840)	(6,102)	(6,402)	(6,747)
20	Selling, general & administrative	(12,198)	(12,810)	(13,384)	(14,043)	(14,800)
21	Operating profit (EBIT)	49,150	51,807	54,128	56,791	59,852
22	Interest income	1,541	1,614	1,694	1,786	1,875
23	Interest expense	(329)	(329)	(317)	(317)	(252)
24	Other expense	(324)	(324)	(324)	(324)	(324)
25	Pretax profit	50,038	52,768	55,181	57,936	61,152
26	Taxes	(13,160)	(13,720)	(14,347)	(15,063)	(15,899)
27	Net income	36,878	39,048	40,834	42,873	45,252
28						
36	Growth rates & margins					
37	Revenue growth	5.0%	5.0%	4.5%	4.9%	5.4%

Figure 10: Income Statement Model with Projections

Take note of the formula of the highlighted cell, which demonstrates how the financial ratios are utilized to generate future income statement data. Cell “E16”, not shown because it holds historical data, contains the most recent year’s historical revenue. Therefore, to make the first projection, the last year of historical revenue is multiplied by one plus the anticipated growth rate, contained within cell “F37”, to arrive at forecasted revenue for the first projection year, presented in the highlighted cell (“F16”).

The basic process for forecasting is simple enough and can be automated. For example, the code in Figure 11, written in Visual Basic and implemented as a macro in Excel, uses a simple algorithm to generate revenue growth rate projections and then inputs the proper formulas into the projection year cells. The results produced in a specific test case are simplified and shown in Figure 12.

```

Sub RevenueProjectionScratch()

'The user makes the current cell the cell for which growth data should be calculated.
Worksheets("Model").Activate
Dim currentCell As String
currentCell = ActiveCell.Value

'Capturing historical revenue numbers
Dim predictor1 As String
predictor1 = ActiveCell.Offset(columnOffset:=1).Address(False, False)
Dim predictor2 As String
predictor2 = ActiveCell.Offset(columnOffset:=2).Address(False, False)

'Inputting string data where appropriate
ActiveSheet.Cells(35, 2).Value = currentCell & " growth"
ActiveSheet.Cells(35, 3).Value = "NA"

'Inputting proper revenue growth formula
ActiveSheet.Cells(35, 4).Formula = "=" & predictor2 & "-" & predictor1 & ")/" & predictor1
Worksheets("Model").Range("D35:E35").FillRight
Worksheets("Model").Range("D35:J35").NumberFormat = "0.0% "

'Calculating the percent change of the historical data which will be used to forecast.
Dim percentChange As Double
percentChange = (Worksheets("Model").Cells(35, 5).Value - Worksheets("Model").Cells(35, 4).Value) _
/ Worksheets("Model").Cells(35, 4).Value

'An algorithm generates estimates for what revenue growth will look like in the future and inputs the numbers
'into the appropriate cells
'The logic behind this algorithm is that revenue growth
'will continue in the same direction and at the same magnitude as is seen in the historical numbers.
|
'This code is utilized to flatline revenue growth after a 50% change from the historical numbers.
Dim halfLastHistoricalGrowth, OneAndAHalfHistoricalGrowth As Double
halfLastHistoricalGrowth = Worksheets("Model").Cells(35, 5).Value / 2
OneAndAQuarterLastHistoricalGrowth = Worksheets("Model").Cells(35, 5).Value * 1.25

Dim counter As Integer
For counter = 5 To 9
    If Worksheets("Model").Cells(35, counter) * (1 + percentChange) > halfLastHistoricalGrowth And _
Worksheets("Model").Cells(35, counter) * (1 + percentChange) < OneAndAQuarterLastHistoricalGrowth Then
        Worksheets("Model").Cells(35, counter + 1).Value = Worksheets("Model").Cells(35, counter).Value * (1 + percentChange)
    Else
        Worksheets("Model").Cells(35, counter + 1).Value = OneAndAQuarterLastHistoricalGrowth
    End If
End If
Next counter

Worksheets("Model").Range("D35:J35").NumberFormat = "0.0% "

End Sub

```

Figure 11: Revenue Projections Code

	Historical		Projected					
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
Revenue	134,800	145,000	155,000	164,743	174,181	183,276	191,998	200,326
Revenue growth	NA	7.6%	6.9%	6.3%	5.7%	5.2%	4.8%	4.3%

Figure 12: Revenue Projection Test

The algorithm generates projected revenue growth rates by using historical rates to estimate the direction and rate growth is taking place. Furthermore, the algorithm will “flat line” the projected numbers after growth rates deviate from historical numbers by more than fifty percent.

This method will produce accurate models in any situation in which the historical data trend is representative of the future. However, there are certainly cases in which this method of prediction could lead to inaccurate analyses. For example: if a company has a

large spike in historical revenue numbers due to reasons unrelated to operations, this would skew the results produced by the algorithm, leading to higher than realizable revenue projections.

It is for this reason that projections, particularly revenue projections, must be made with oversight. An expert, who understands a particular industry and the ongoing economic trends affecting that industry, is often necessary to make accurate projections. In particular, projections, such as revenue growth, that are often made based on industry-specific drivers. For example, models relating to consumer products may take a company's internal projections of how many units will be sold in the coming years and use those numbers, combined with the retail prices of units, to back into revenue projections. Or a company in the power and electric industry will use the company's projections of how many watts of power will be generated, combined with an analysis of what the price of electricity is expected to be, to form projections. In instances such as these, where revenue projections are created by the user, automation can take place in other processes. For example, the code in Figure 13 takes the projected revenue numbers and automatically inserts the proper formulas to calculate revenue growth rates within the forecasting ratios table shown in Figure 9. In fact, this code, with a simple modification, could be utilized to create any line item that inputs growth rate formulas based on already projected numbers.


```

Sub RevenueProjection()
|
'The user makes the current cell the cell for which growth data should be calculated.
Worksheets("Model").Activate
Dim currentCell As String
currentCell = ActiveCell.Value

'Capturing historical revenue numbers
Dim predictor1 As String
predictor1 = ActiveCell.Offset(columnOffset:=1).Address(False, False)
Dim predictor2 As String
predictor2 = ActiveCell.Offset(columnOffset:=2).Address(False, False)

'Inputting string data into appropriate cells
ActiveSheet.Cells(35, 2).Value = currentCell & " growth"
ActiveSheet.Cells(35, 3).Value = "NA"

'Inputting proper revenue growth formula
ActiveSheet.Cells(35, 4).Formula = "=" & predictor2 & "-" & predictor1 & ")/" & predictor1
Worksheets("Model").Range("D35:J35").FillRight
Worksheets("Model").Range("D35:J35").NumberFormat = "0.0% "

End Sub

```

Figure 13: Growth Rates from Projections

Observations made during research have led to the conclusion that step two in the three-statement modeling process is not an ideal candidate for automation because of the necessity for human interaction in situations like those discussed above. However, it is not the case that there are no steps within the entirety of the process that cannot be automated. As has been discussed, consistency and efficiency within the modeling process are greatly improved with semi-automation. Step two is no exception; there are certain repeatable, automatable processes can be implemented using code similar to that shown in Figures 11 and 13.

2.7 Step Two: Forecasting the Balance Sheet

We have been focusing on projecting the income statement, but, for the purpose of showing processes that can be automated with more precision, let's move on to projecting the balance sheet. The balance sheet, as discussed, shows a moment in time "snapshot" of a company's assets, liabilities and owner's equity. What is interesting about the balance sheet is, by nature, it is a lot more consistent than the income statement. Take property plant and equipment (PP&E), an asset account that includes items such as land, machinery and buildings, as an example: companies, particularly ones who have been operating for some

time, are much more likely to continually upgrade or discard their PP&E from year to year than they are to do a complete overhaul and purchase all new PP&E in one year. For this reason, this account and other accounts like it are stable and predictable. Another example is accounts receivable, an asset account that shows how much a company is owed. Companies have become very good at issuing credit and understand at what level to maintain accounts receivable in order generate the most profit. If their credit policies are too stringent, they will lose business and profitability, but if, on the other hand, they are not stringent enough, they will have too large of a bad debt expense.

As a result of this consistency, balance sheet items can often be predicted without an understanding of industry trends and current economic conditions, making this process a good candidate for automation. Forecasting accounts receivable, for example, has been done in every model I have seen by generating an accounts receivable margin (accounts receivable /sales) and roughly averaging the margin for projection purposes. How this is represented in a model is shown in Figure 14:

Accounts receivable								
Beginning of period				13,102	13,454	14,318	14,959	15,695
Increases / (decreases)				352	864	642	736	846
End of period		10,930	13,102	13,454	14,318	14,959	15,695	16,541
AR as % of sales		7.0%	7.7%	7.5%	7.6%	7.6%	7.6%	7.6%

Figure 14: Accounts Receivable Projections

The line titled “End of period” represents what is contained in the accounts receivable account and is reported in a company’s financial statements at the end of the fiscal year, and the highlighted cell is the first projected accounts receivable amount. This projection is a result of simply multiplying “AR as % of sales” by revenue for the given year. As can be observed, “AR as a % of sales” is stable from year to year, the reason being is companies issue credit with the goal of hitting certain yearly targets.

Generating this sub-table and linking the correct data back into the balance sheet is quite simple and can be done by grabbing the correct data, and implementing the correct formulas. This is done automatically using the code shown in Figure 15.

```
Sub BalanceSheetProjections()

'The appropriate income statement line item that will be utilized for forecasting
'would either be hard coded and/or the user could be prompted.
'I hard coded for simplicity.
Dim incomeStatementAccount As String
incomeStatementAccount = Worksheets("Model").Cells(16, 4).Address(False, False)

'The user signifies the balance sheet account to be forecasted simply by making it the active cell.
'End of period numbers are linked to the balance sheet account being forecasted.
Worksheets("Model").Cells(126, 4) = "=" & ActiveCell.Offset(columnOffset:=2).Address(False, False)
Worksheets("Model").Range("D126:E126").FillRight

'The appropriate formula, calculating the balance of the balance sheet account as a percentage of the income sheet line,
'is created within the appropriate cells.
Worksheets("Model").Cells(128, 4) = "=" & ActiveCell.Offset(columnOffset:=2).Address(False, False) & "/" & incomeStatementAccount
Worksheets("Model").Range("D128:E128").FillRight
Worksheets("Model").Range("D128:J128").NumberFormat = "0.0%"

'An algorithm generates estimates for what these percentages will look like in the future and inputs the numbers
'into the appropriate cells
'The logic behind this algorithm is that the balance sheet account balance as a percentage of the income statement line item
'will continue in the same direction and at the same rate it is in the historical numbers.
Dim startingHistoricalPercentage As Double
Dim nextHistoricalPercentage As Double
Dim percentChangeInHistoricals As Double
startingHistoricalPercentage = Worksheets("Model").Cells(128, 4).Value
nextHistoricalPercentage = Worksheets("Model").Cells(128, 5).Value
percentChangeInHistoricals = (nextHistoricalPercentage - startingHistoricalPercentage) / startingHistoricalPercentage

'The algorithm furthermore insures that the estimate values will flatline once a 50%
'change from the most recent historical data has taken place
Dim maxPercentProjection As Double
Dim minPercentProjection As Double
maxPercentProjection = (nextHistoricalPercentage * 1.5)
minPercentProjection = (nextHistoricalPercentage * 0.5)

Dim counter As Integer
For counter = 5 To 9
    If Worksheets("Model").Cells(128, counter).Value * (1 + percentChangeInHistoricals) < maxPercentProjection _
    And Worksheets("Model").Cells(128, counter).Value * (1 + percentChangeInHistoricals) > minPercentProjection Then
        Worksheets("Model").Cells(128, counter + 1).Value = Worksheets("Model").Cells(128, counter).Value * _
            (1 + percentChangeInHistoricals)
    End If
Next counter

Worksheets("Model").Range("F128:J128").Font.Color = RGB(0, 0, 255)

'Creates the appropriate formulas to calculate the projected accounts receivable amounts based on the estimates made by the algorithm
Worksheets("Model").Cells(126, 6) = "=" & Worksheets("Model").Cells(16, 6).Address(False, False) & "*" & Worksheets("Model").Cells(128, 6).Address(False, False)
Worksheets("Model").Range("F126:J126").FillRight
Worksheets("Model").Range("F126:J126").NumberFormat = "000,000"

End Sub
```

Figure 15: Balance Sheet Projections Code

This code is not robust in its implementation; however, it is designed to serve as a proof of concept. Similar methods can be utilized to project the balance sheet almost in its entirety, as well as select income statement lines.

In summary, step two in the three-statement modeling process, much like step one, may not be fully automatable, but there is a large amount of work that can be

automated, leading to increased efficiency and accuracy. The theory underlying this thesis is that the proper amount of human interaction combined with technical tools to facilitate the modeling process will lead to a drastic increase in efficiency and accuracy within models while still enabling financial experts to create something that is robust and fully representative of the asset at hand.

2.8 Step Three: Generating the Cash Flow Statement

The final step in the three-statement modeling process is one that seems most interesting because it is the step in which the accounting is the most intriguing and, as one learns the process behind modeling the cash flow statement, it seems the most automatable of all three steps in the three-statement modeling process. Quite frankly, it is surprising there are no preexisting tools utilized to facilitate the process. It should be noted that the cash flow statement, from an accounting standpoint, is often considered the most difficult statement to understand, but the conclusions of this thesis can be drawn without a comprehensive understanding of the statement. Therefore, for simplification, we are only going to focus on the operating section of the statement.

To summarize, the cash flow statement is designed to present cash-basis information about a company's operating, financing and investing activities. This is a necessity because the data presented on a company's income statement, in accordance with GAAP, is on an accrual basis, meaning revenues and expenses are recognized when they are incurred, regardless of when cash is exchanged. A good example of the utilization of accrual-based accounting is when inventory is purchased. When inventory is purchased and paid for in cash, it is not immediately expensed. It is capitalized and only expensed when sold, leading to the creation of "Cost of Goods Sold" on the income

statement. However, if only fifty percent of the inventory purchased, using cash, during a specified period is sold within the same specified period, it creates a rift between how much a company made according to “Net Income” on the income statement and how much actual cash is available for shareholders. It is for this reason that companies are valued based on how much cash they generate and not their reported bottom lines, and, therefore, the cash flow statement is a necessity for investors to make informed decisions.

The operating section, for financial analyses purposes, is considered the most important section of the cash flow statement. The reason for this is investors often desire to value a company’s operations without regard to their capital structure. Cash from operations is used in multiple forms of analysis but, more specifically, it is used in leveraged buyout analysis to predict cash flow available to pay down debt (Cash Flow from Operations – Capital Expenditures). Because of its importance to financial practitioners and because the completion of the operating section incorporates all the processes necessary to complete the remaining sections of the statement, we will focus solely on the operating section. To begin, a completed cash flow statement, as presented within a model, will be shown for reference and we will walk through the process of putting it together.

CASH FLOW STATEMENT					
<i>Fiscal year</i>	2014P	2015P	2016P	2017P	2018P
<i>Fiscal year end date</i>	<i>9/30/14</i>	<i>9/30/15</i>	<i>9/30/16</i>	<i>9/30/17</i>	<i>9/30/18</i>
Net income	36,878	39,048	40,834	42,873	45,252
Depreciation and amortization	7,149	7,535	7,677	7,786	8,001
Stock based compensation	2,214	2,213	2,426	2,545	2,682
Accounts receivable	(352)	(864)	(642)	(736)	(846)
Inventory	(41)	(82)	(85)	(97)	(111)
Accounts payable	538	1,507	1,094	1,255	1,442
Accrued expenses & def revenues	(58)	68	(109)	139	(15)
Other current assets (inc. non-trade receivables)	0	0	0	0	0
Deferred tax assets (DTAs)	0	0	0	0	0
Other assets	0	0	0	0	0
Other non current liabilities	3,000	3,000	3,000	3,000	3,000
Non-cash (PIK) interest	0	0	0	0	0
Cash from operating activities	49,327	52,426	54,196	56,766	59,406

Figure 16: Cash Flow Statement Model

The operating section of the cash flow statement, shown in Figure 16, has been color coded by discernable section. The top line (blue) is always net income because, as discussed, the goal of the operating section is to reconcile “Net income” from an accrual-basis to a cash-basis. The second section (red) encompasses non-cash charges, these are expenses on the income statement but do not represent an outlay of cash. Depreciation and amortization (D&A), which represents wear and tear on a company’s property plant and equipment, is the most commonly seen item within this section because D&A is the accounting method used to expense PP&E over its useful lifetime. However, this equipment is, often times, paid for upfront and, therefore, D&A charges do not represent cash outlays. The last section (green) accounts for changes in a company’s operations-based balance sheet accounts. Consider “Inventory”, a balance sheet account, for example: when the change (the difference between the inventory balance at the start of the period and the inventory balance at the end of the period) in the inventory account is positive it means that more inventory was purchased during the period than was sold. Because inventory is a capitalized expense and only expensed when sold, this creates a discrepancy between net income and cash that is

accounted for by subtracting out the increase in inventory.

Grasping the cash flow statement beyond a rudimentary level is not highly important for our purposes, however, it is important to remember that all of the information needed to create the cash flow statement has already been generated in steps one and two, and therefore, the creation of this statement simply involves understanding what information needs to be retrieved and linking the associated cash flow statement cell with its counterpart in other parts of the model.

Automating the creation of the cash flow statement can be done in multiple ways, but I would suggest the code shown in figures 17 and 18 as a starting point.

```
Sub CashFlowDirectRetrieval()  
Worksheets("Model").Activate  
  
Dim activeCellRow As Integer  
activeCellRow = activeCell.Row  
  
For counter = 15 To 47  
    If Worksheets("Model").Cells(counter, 2).Value = activeCell.Value Then  
        activeCell.Offset(columnOffset:=4).Formula = "=" & Worksheets("Model").Cells(counter, 6).Address(False, False)  
        Worksheets("Model").Range("F" & activeCellRow & ":" & "J" & activeCellRow).FillRight  
        Worksheets("Model").Range("F" & activeCellRow & ":" & "J" & activeCellRow).NumberFormat = "#,##0_);(,##0);@"  
    End If  
Next counter  
  
End Sub
```

Figure 17: Cash Flow Direct Retrieval

```
Sub IndirectRetrieval()  
Worksheets("Model").Activate  
  
'Captures the activeCell which is selected by the user  
'The data within the activeCell is used to search for and find the corresponding data  
'within the model and generate the appropriate links within the cash flow statement.  
Dim activeCellRow As Integer  
activeCellRow = activeCell.Row  
  
'Searches the model for the appropriate data  
'It is worth noting that I specified the general location where the algorithm should  
'look. This is not robust and could be greatly improved for practice.  
  
'Furthermore, some data that is retrieved within this section must be entered as a positive  
'while other must be entered as a negative. Therefore, the user would have to specify this, or  
'a table that specifies which accounts are entered as a positive and which are entered as a  
'negative could be created and referenced.  
  
For counter = 122 To 219  
    If Worksheets("Model").Cells(counter, 2).Value = activeCell.Value Then  
        activeCell.Offset(columnOffset:=4).Formula = "=" & Worksheets("Model").Cells(counter + 2, 6).Address(False, False)  
        Worksheets("Model").Range("F" & activeCellRow & ":" & "J" & activeCellRow).FillRight  
        Worksheets("Model").Range("F" & activeCellRow & ":" & "J" & activeCellRow).NumberFormat = "#,##0_);(,##0);@"  
    End If  
Next counter  
End Sub
```

Figure 18: Cash Flow Indirect Retrieval

The algorithm behind the code is simple; it takes the line item label the user has entered from

the cash flow statement and then proceeds to move through specified areas of the model looking for the equivalent item. Once this item has been found, it links the corresponding data back into the cash flow statement. One important note to make is that the algorithm has to account for the differences in the discernable sections of the cash flow statement. For example, the blue, red and orange sections, presented in Figure 16, only require finding the corresponding line items within the income statement or balance sheet. The first Sub (Sub CashFlowDirectRetrieval ()), shown in Figure 16, handles these operations. Take net income, the first line of the cash flow statement, for example: Once the algorithm finds “Net Income” within the income statement, its job is easy because it simply has to link the corresponding cells in the cash flow statement to the data directly to the right of “Net income”.

On the other hand, the lines in the cash flow statement that account for changes in operations based balance sheet accounts (the green section), require a more in-depth search. This is handled by the second Sub (Sub IndirectRetrieval()) shown in Figure 17. The reason a different type of search must take place is that the change in account balances is calculated separately from the line items presented in the income statement or balance sheet. Take accounts receivable as an example: referring back to Figure 14, which was utilized when discussing the process of automating balance sheet projections, there is a row labeled “increase / (decrease)”, shown in the figure, which was ignored while discussing projections because it was irrelevant. The purpose of this row, as it turns out, is to be linked back to the cash flow statement. Because of the necessity of sections such as this one, for projecting certain balance sheet and income statement items and calculating the increase/decrease in account balances, there will

be multiple locations in the model where the line item label being searched for can be found. It is, therefore, necessary for the algorithm to seek the row that calculates the increase/decrease in the account balance the specific cash flow statement line item refers to while ignoring rows that may have the same or similar labels elsewhere in the model. For simplicity, in the code shown above, it has been specified for the algorithm in which sections of the model to find what data. This does, however, create a lack of robustness that would have to be addressed before this algorithm could be fully implemented. The solution could be as simple as entering a text marker on each line within the green section of the cash flow statement that the algorithm can use to differentiate the section from its counterparts.

Chapter 3

Summary

To summarize, automating step three in the three-statement modeling process requires a certain amount of consistency within the model. Although there will be certain consistencies across all financial models, there is often variation in the formats and methods used. Predictions, for example, may be calculated within separate tabs, each financial statement may have its own tab, or the entire model may be done on a single page in Excel. Inconsistencies in formatting, as well as other issues discussed throughout this thesis, have led to the conclusion that there is no one-size-fits-all solution to improve the modeling process.

Systems currently available that attempt to generate a one-size-fits-all automated process eliminate any ability to create robust models across industries with varying drivers and valuation techniques. Failed attempts to create a comprehensive system that eliminates the need for human interaction in the modeling process is what has led many to assume that automation within modeling will lead to an inaccurate analysis that is more trouble to fix than it would be to create an accurate analysis from scratch.

An outsider cannot create an automated modeling process; experts, with an understanding of the specific methods they utilize, must identify the elements of their modeling processes that can be automated and improved with technical tools. Furthermore, it is necessary that the experts utilizing the tools have an in-depth understanding of how the algorithms within them generate data, allowing for

adaptability when situations arise that may require some portion of the analysis to be done differently than is the norm.

The necessity for models to be adaptable in situations that are outside the norm is why it is necessary to utilize a “semi-automated” process in which a modeling expert moves from step to step in the three-statement modeling process recognizing along the way which tools can be beneficial, at what points the user must interact with and adapt the model to the situation at hand and how those adaptations will need to be accounted for when using the same technical tools later in the process.

The methods that have been laid out in this thesis are a good base to be adapted for modeling within specific industries or situations. The processes advocated can be summarized as follows:

(1) The necessary financial data is autonomously extracted using basic natural language processing techniques.

(1b) The user interacts with the data consolidating and formatting the data as needed.

(1c) Technical tools assist in the formatting by completing tasks such as color coding cells and inputting formulas where necessary.

(2) With an understanding of the forecasting process, the user decides what tools available can be utilized to forecast predictable elements, while simultaneously interacting in the forecasting of elements that require unique assumptions and financial expertise.

(3) Lastly, the user, with an understanding of the way in which the data has been manipulated and forecasted, enters into the model the line labels of the elements that

need to be retrieved and linked within the cash flow statement before running a process in which the data is found and the appropriate links are made instantaneously and automatically.

While some may question if the time spent in creating the technical tools necessary is worthwhile since there must remain interaction at all levels, it is important to remember that the three-statement model is just the base from which many other types of analysis are developed. This automation process does not stop upon the completion of the three-statement model. The ideas presented in this thesis can be expanded into discounted cash flow, leveraged buyout, comparable companies and other modeling processes. As these techniques are spread across all modeling processes, having a robust semi-automated process with a set of tools at the modeler's disposal can immensely increase efficiency, accuracy as well as consistency while still leaving room for financial experts to input the knowledge-backed assumptions that are necessary for a robust, professional model.

Bibliography

- [1] Bird, Steven, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. First ed. Sebastopol: O'Reilly Media, 2009. Print.
- [2] Rosenbaum, Joshua, and Joshua Pearl. *Investment Banking*. Hoboken, NJ: Wiley, 2013. Print.
- [3] McKinsey & Company Inc., Tim Koller, Marc Goedhart, and David Wessels. *Valuation: Measuring and Managing the Value of Companies*. Fifth ed. Hoboken, NJ: Wiley, 2010. Print.
- [4] Brealey, Richard A., Stewart C. Myers, and Franklin Allen. *Principles of Corporate Finance*. Ninth ed. New York: McGraw-Hill, 2008. Print.
- [5] Kieso, Donald E., Jerry J. Weygandt, and Terry D. Warfield. *Intermediate Accounting*. Fifteenth ed. S.l.: John Wiley, 2013. Print.
- [6] *Introduction to Natural Language Processing*. Dir. Brain Sacash. Perf. Brian Sacash. *Introduction to Natural Language Processing*. Udemy, n.d. Web. Nov. 2016. <<https://www.udemy.com/natural-language-processing>>.
- [7] *Merger Bundle: Crack Excel with Modeling & Best Practices*. Dir. Training the Street. Perf. Training the Street. *Training The Street*. Udemy, n.d. Web. Nov. 2016. <<https://www.udemy.com/training-the-street-bundle>>.
- [8] *Corporate Valuation*. Dir. Training the Street. Perf. Training the Street. *Training The Street*. Udemy, n.d. Web. Nov. 2016. <<https://www.udemy.com/training-the-street-bundle>>.
- [9] *Excel and Financial Modeling Fundamentals*. Dir. Brian DeChesare. Perf. Brian DeChesare. *Breaking into Wall Street*. N.p., n.d. Web. Nov. 2016 <<http://www.breakingintowallstreet.com>>.
- [10] Staff, Investopedia. "Financial Modeling." *Investopedia*. N.p., 13 Jan. 2016. Web. Jan.-Feb. 2017. <<http://www.investopedia.com/terms/f/financialmodeling.asp>>.
- [11] "FASB Accounting Standards Codification®." *FASB Accounting Standards Codification®*. N.p., n.d. Web. Feb.-Mar. 2017. <<https://asc.fasb.org/>>.
- [12] Price Waterhouse Cooper. "Financial Statement Presentation." *Financial Statement Presentation*. Price Waterhouse Cooper, 2016. Web. Feb.-Mar. 2017. <<http://www.pwc.com/us/en/cfodirect/publications/accounting-guides/financial-statement-presentation-accounting-guide.html>>.

[13] "Financial Modeling Courses and Investment Banking Training." *Wall Street Prep*. Wall Street Prep, n.d. Web. Jan.-Feb. 2017.
<<http://www.wallstreetprep.com/>>.