Cash Management and Forecasting at Scale

Vrushank Vora, Gus Ostow
Papert Labs Inc.
{vvora, gus}@papertlabs.com

Carlos Vega
Tesorio
carlos@tesorio.com

ABSTRACT

Forecasts for future cash flows serve as the foundation for critical business decisions made by operators, executives, and investors. In complex, fast-paced business environments, lack of transparency in the flow of financial information is common, which leads to inaccurate forecasts and misguided decision making. According to *The Economist*, since 2015, General Electric's stock price and profits have decreased by nearly 50% as a result of "muddled flow of financial information" [3]. In this paper, we present a general purpose machine learning system for cash management and cash flow forecasting that adds transparency to existing processes. Our platform integrates with business's Enterprise Resource Planning (ERP) systems (e.g. NetSuite, SAP, Intuit's Quickbooks, etc.) to forecast: 1.) Payment Dates for Outstanding Invoices; 2.) Future Invoices and Bills for the business; and 3.) 12-month ahead Account Receivables (AR) and Account Payables (AP). Compared to existing methodologies, we present results that are more accurate, update real-time, and require almost no human intervention. To the best of our knowledge, our work represents the first large-scale application of machine learning to transaction level datasets of companies invoices and bills. Building models using such fine-grained transaction level data allows us to build explainable forecasting models and align high-level decision making with day-to-day operations.

1. INTRODUCTION

What will be the Account Receivables and Account Payables amount in next 12 months?

Which vendors are likely to bill in the next 3 weeks?

Which customers consistently pay way after their due date?

How many outstanding invoices will come in late?

These are few of the basic questions that executives and investors should be able to answer about business operations to make effective operational and financing decisions. When answers to such questions are inaccurate or non-existent, executive strategy is misaligned with day-to-day actions of middle managers and employees. Operational blunders are likewise imminent [3].

In financial operations management, profit is an opinion, cash flow is fact. Cash flow is the fundamental measurement that determines the viability of a business. On a forward looking basis, the value of a company is by definition the present value of all its future cash flows. However, evolving regulation, such as Revenue Recognition Standard ASC 606, is making it harder to understand the correlation between a Profit and Loss statements, changes in the Balance Sheet and variations in cash. The smartest companies, such as Amazon, have realized that tracking cash flow carefully gives them the freedom to invest for the long-term and secure their success.

With the growing complexity of business operations, the market for Enterprise Resource Planning (ERP) systems is expected to grow to $41.69 billion by 2020 [1]. While there are few dominant ERP vendors such as NetSuite and SAP, it is an extremely fragmented market. Customers routinely complain about implementation overruns, rigidity and obsolescence of software, and data quality issues. All of these issues make any sort of Machine Learning and Data Science extremely difficult. Furthermore, any individual business would not only lack the technical expertise, but would be unable to leverage the transfer of knowledge from a vast network of datasets from similar businesses to learn effective global machine learning models. In this paper, we present an application of large-scale machine learning and time-series forecasting to forecast: 1.) Payment Dates for Outstanding Invoices (invoice time to pay); 2.) Future Invoices and Bills for the business; and 3.) 12-month ahead Account Receivables (AR) and Account Payables (AP). When predicting invoice time to pay, our platform effectively learns to transfer knowledge across businesses, allowing it to achieve reasonable accuracies even with companies very few invoices. We likewise use these transaction level models to generate short-term cash flow forecasts that are not only more accurate but also more explainable.

2. RELATED WORK & EXISTING BUSINESS PROCESSES

While this work stands at the intersection of past literature in corporate finance, statistical machine learning, and time-series forecasting, to the best of our knowledge, our work represents the first large-scale application of machine learning to transaction level datasets of companies invoices and bills for cash management and forecasting. Building models using such fine-grained transaction level data allows us to build explainable forecasting models and align high-level decision making with day-to-day operations.

2.1 Predicting Payment Dates for Outstanding Invoices

Zeng et al. and Peiguang have previously applied machine learning methods to predict delinquent invoices and customers [9] [4]. Both of their works treat the problem as a classification problem. Their works discretized the target, difference between the due date of the invoice and the actual payment date of the invoice, into bins. Treatment of the problem as a classification problem limits utility of this task into the upstream tasks for predicting cash flows in a given month and day. We rectify this by learning models that machine learn the parameters of a Weibull distribution and generate a distribution over every single invoice and when it will get paid. Moreover, from a product perspective, providing Treasurers and Collections Agents with a rich distribution over the invoice payment date expands the actions that can be taken to accelerate the process of collecting on that invoice.

2.2 Predicting Future Invoice and Bill Dates

To our knowledge, ours is the first work to predict future invoice and bill dates from existing customers and bills respectively.

2.3 Forecasting Account Receivables and Account Payables

Existing business processes for building cash forecasts are fragmented across internal Excel Models, business intelligence workstations, and ERP vendors. All of these processes involve manual ingestion of historical data and often limit the complexity of methodologies and datasets that can be used for forecasting [2].On the corporate finance and valuation front, Fader et al. have previously applied customer-level models to build valuations from public disclosed data of companies like Wayfair, Blue Apron, and Overstock. However, their models rely on business model specific assumptions about contractual nature of firms [5][6].Our work instead leverages advances Bayesian Structural Time Series Methods and Bayesian Additive Models for Time Series Forecasting to build models that dont require business specific assumptions and can handle uncertainty in a much more natural way [7][8].

3. DATASET DESCRIPTION

There are generally two types of business models: Business to Consumer (B2C) and Business to Business (B2B). In B2C models, businesses sells goods and services directly to a consumer. In general, B2C Internet businesses partner directly with payment processors to process payments from consumers. Payment Processors are obligated to pay the sum of the transactions to the business directly on the due date. On the other hand, B2B businesses sell goods and services to other businesses, who are then invoiced. While invoices have terms e.g. Net 15, Net 30, Net 90, etc. which creates contractual due date, partnering businesses often pay late in order to optimize their working capital. Our dataset consists of businesses with both B2B and B2C business models. We present results of our models on a subset of companies on the platform. For the sake of data privacy, we will not disclose the businesses but only their high-level business models. Table 1 details the companies and their respective business models.

Anonymized Company	Business Model
A	B2B
B	Hybrid B2B and B2C
C	B2B
D	B2B
E	B2B
F	B2B
G	Hybrid B2B and B2C

Table 1: Companies and their Business Models.

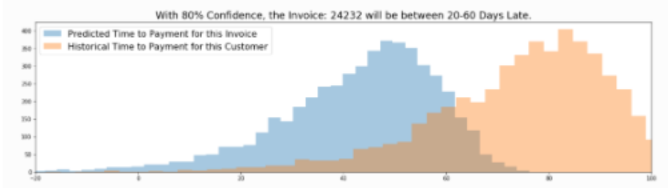

Figure 1: Output of Our Invoice Time to Pay Model

Our models are built on transaction level datasets from companies within our network. We leverage invoice and payment transactions to build invoice time to pay models, future invoice dates from existing customers, and forecast 12-month ahead AR. On the payables side, we leverage historical expenses incurred with existing vendors and G&A expenses such as payroll, tax, rent, etc. to build models future bills and 12-month ahead AP.

4. SYSTEM OVERVIEW AND RESULTS

Our system for cash management and forecasting consists of three main components: 1.) Predicting Payment Dates for Outstanding Invoices; 2.) Predicting future bills and invoices; and 3.) Forecasts for 12-month ahead AR and AP. Since the output of our transaction level models from 1.) and 2.) is a distribution, we can use them to generate more accurate and explainable cash flow forecasts.

4.1 Predicting Payment Dates for Outstanding Invoices

Problem: Invoices are often paid after their due date. Collections Managers need to know exactly when different invoices will be paid to optimize their collection processes. For a given invoice, we generate a distribution over their expected payment date. This allows us to quantify the uncertainty of our predictions and use the transaction level predictions to generate explainable short-term forecasts. See Figure 1 for example output of invoice time to pay model.

Models: We featurize each invoice based on invoice characteristics, customers past transaction history, due date of the invoice, and variety of other characteristics. Featurized invoices are fed into Gradient Boosted Machines (GBMs) which are used to parametrize a Weibull Distribution that serves as our final output. We benchmark our machine learned models against the median generated from Kaplan-Meier (KM) estimates of customer's historical time to pay.

Evaluation Metric: There are two metrics we use to evaluate the quality of this model. First, we evaluate the quality

Company	ML Model (MAE)	KM (MAE)	Lift(%)
A	**11.71**	19	**-38.37%**
B	**3.37**	3.61	**-6.55%**
C	**10.01**	14.0	**-28.50%**
D	**5.13**	10.56	**-51.40%**
E	**5.24**	17.79	**-70.55%**
F	**14.49**	26.13	**-44.55%**
G	**8.04**	10.45	**-23.10%**

Table 2: Performance of Our Best Model Against Baselines

Company	Global Model	Company Specific	Lift(%)
A	**11.71**	13.715	**-14.62%**
B	**3.37**	9.37	**-64.03%**
C	**10.01**	11.247	**-11.00%**
D	**5.13**	5.57	**-7.90%**
E	**5.24**	11.87	**-55.86%**
F	**14.49**	16.02	**-9.55%**
G	**8.04**	8.03	**0.12%**

Table 3: Performance of Our Global Machine Learning Models vs. Company Specific Machine Learning Models

of our predictions for when an invoice will get paid by calculating the Median Absolute Error (MAE) of the difference between our predicted payment dates and actual payment dates. Concretely, the metric is defined as follows where i is an invoice:

$$Median_i(|actual\_pymnt\_date_i - predicted\_pymnt\_date_i|)$$

Second, since we generate bottom-up, short-term cash flow forecasts by calculating the expected value of an invoice being paid off on any given day, we also evaluate the quality of the short-term cash flow forecasts generated from these transaction using out-of-sample, backtested Median Absolute Percent Error (MAPE).

Results and Discussion: In Table 2, we show that our best machine learned models outperform all standard survival baselines. More importantly, as Table 3 shows, for six out of the seven companies, our global machine learning model trained on all of the companies' dataset outperforms company specific models. This *network effect* associated with the data is especially prominent for companies that have very few number of invoices in the dataset. Beyond forecasting when invoices will be due, since we generate distributions over invoices, we can use transaction level models to generate daily cash flow forecasts like the ones shown in Figure 2. Most existing methodologies build such forecasts using top-level time series methods. By building these forecasts in a bottom-up fashion, we can expand the scope of actions that can be taken to *optimize* collections and expected cash inflows while restricting cash outflows. [Add note about performance of cash flow forecasts]

4.2 Future Invoice and Bill Predictions

Problem: Predicting payment dates of outstanding transactions is fundamentally limited to invoices and bills that are already entered into an ERP by the end user. As data

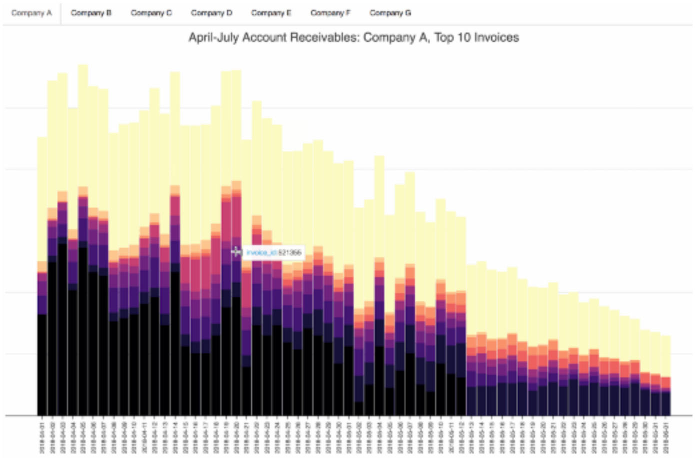

Figure 2: Daily Expected AR Forecasts for a Company (y-axis hidden for privacy reasons)

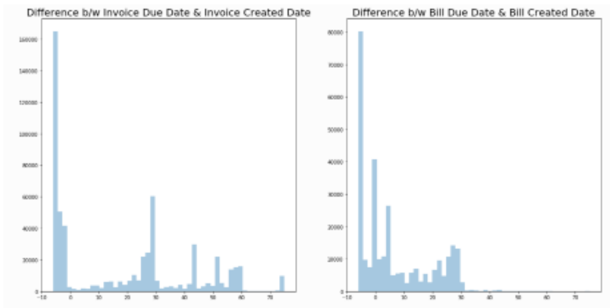

Figure 3: Transactions Entered into ERPs

in Figure 3 suggests that invoices and bills are typically entered into ERPs only months weeks in advance, which restricts the prediction horizon for the AR / AP forecasts to only one-month ahead. We likewise use historical invoices and bills from existing customers to predict future invoice and bills. The periodic nature of many partners business needs suggests that there is signal available to help infer when future transactions will be due and for how much. Often times billing cycles are transparently monthly, quarterly, or yearly. See Figure 4 for examples of simple invoicing and billing patterns. Other times, however, for a lot of SaaS businesses invoicing and billing patterns are much more complex as showing in 5. This warrants the utility of more complicated machine learning algorithms.

Models: The workhorse of our Future Invoice and Bill Predictions is Time Series Clustering. For each customer we learn the latent contractual details based on invoicing and billing patterns. This allows us to forecast the future dates and the amounts of the transaction for the customers and vendors. Furthermore, we also learn a model for partner churn. If a partner has churned, we shouldn't be predicting any future transactions for them. This model is a GBM trained on the features constructed from partner's past transaction history and time since last transaction.

Evaluation Metric: Our first metric for this model is the F1-Score on out-of-sample time periods between customers

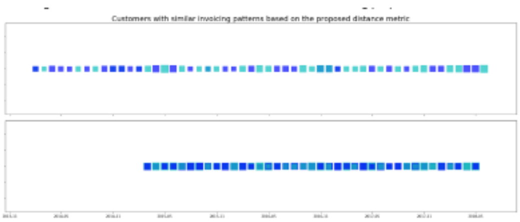

Figure 4: Example Simple Invoicing Patterns Across Customers. Each row is a visualization of the invoicing history of a specific customer. Each colored square represents a single bill. The x-axis is the dates when invoices were due. Amount billed is proportional to the size of the invoices squares. Squares are colored only to visually distinguish overlapping squares

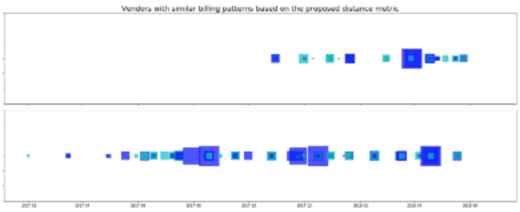

Figure 5: Example More Complex Billing Patterns Across Vendors.

and vendors who had a transaction in that time period and whether we predicted those vendors to have a transaction within that time period. We also use a the standard classification metric ROC AUC to evaluate quality of our customer and vendor churn and staleness algorithms.

Results and Discussion: As Figure 4 and 5 show, our algorithms can easily cluster partners of companies with similar contractual details. While the examples in Figure 4 have simple past and future transaction history, Figure 5 shows our algorithm's abilities to identify and cluster more complex contractual details. Table 4 and 5 shows how our clustering algorithms compare to baseline. While for most companies our clustering based approach works better, it doesn't *significantly* outperform simple baselines [Note: We exclude certain companies due to not having enough customer and vendor data to make clustering a fruitful effort]. However, because of this customer level clustering, we can also identify similar business segments and cohorts for a given business. Identification of business segments is useful in predicting medium-term AR and AP forecasts.

Company (AR)	Clustering Model	Historical Mean
A	**0.34**	0.26
C	0.31	**0.32**
D	**0.67**	0.66
F	**0.87**	0.84
G	**0.95**	0.94

Table 4: Averaged F1-Scores of Five quarters for Future Receivables Prediction: Q1 2017-Q2 2018

Company (AP)	Clustering Model	Historical Mean
A	**0.60**	0.59
C	**0.64**	0.61
D	**0.75**	0.73
F	**0.60**	0.57
G	**0.72**	0.70

Table 5: Averaged F1-Scores of Five quarters for Future Payables Prediction: Q1 2017-Q2 2018

Company	AR ROC AUC	AP ROC AUC
A	0.70	0.83
C	0.58	0.80
D	0.78	0.85
F	0.92	0.88
G	0.97	0.80

Table 6: ROC AUC Scores of Partner Inactivity for Companies

Likewise, we build classification models to evaluate whether customers and vendors of our companies have churned or not. The result of our machine learned algorithms are shown in 6. This sets the foundation for alerting treasurers and the sales force at our companies to take appropriate actions when managing partner relationships that are likely to churn. Furthermore, in the next stage of our modeling process, we can use customer churn models to build cash flow forecasts that aggregate customer and vendor level Customer Life Time Values.

4.3 12-month Ahead AR and AP Forecasts

Problem: The holy grail of financial forecasting is to build accurate forecasts for future cash flows. Accurate cash flow forecasts have been elusive due to disaggregated nature of financial data across businesses and obsolete forecasting methodologies. We generate 12-month ahead AR and AP Forecasts by: 1.) aggregating models from transaction time to pay and future invoice and bill predictions models; 2.) leveraging modern innovations in Bayesian time-series forecasting and machine learning.

Models: To forecast future AR and AP, we leverage distributions from invoice level models. This allows us to generate short-term cash flow forecasts that are 4-5 months ahead. Beyond that, we leverage the advances in Bayesian Time Series Forecasting methods [8] to generate 5-12 month ahead forecasts. We customize the advances in Bayesian Structural Time Series Methods and Bayesian Additive Models for Time Series Forecasting allow us to uncover latent variables to our needs. By utilizing advances in modern time series methods, our example forecasts uncover latent variables e.g. Seasonality as shown in Figure 6. Automatic seasonality detection is especially useful in helping optimize capital, human resource allocation, and liquidity constraints. We benchmark our models against: 1.) Traditional models in corporate finance which which calculate expected YoY growth rates of the business and multiply them by the previous Fiscal Years AR / AP; 2.) traditional time-series forecasting methods like ARIMA.

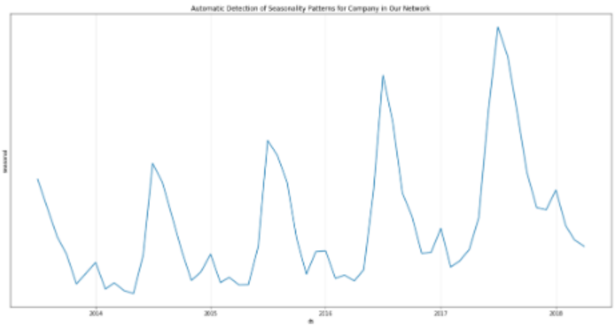

Figure 6: Automatic Seasonality Detection for One of the Companies in Our Network

Figure 7: Forecast Breakdown

Evaluation Metric: The standard metric for such regression problems is the Absolute Percent Error (APE) between the forecasted and actual amount.

$$APE = \frac{|forecasted\_amount - actual\_amount|}{actual\_amount}$$

Since the executives and investors make decisions at the yearly level, we calculate the Absolute Percent Error between the Sum of Forecasted AR / AP across twelve months against the Total AR / AP that came in during that time period. The TAPE is defined as follows:

$$TAPE = APE(\sum_{i=1}^{12} forecasted\_amount_i, \sum_{i=1}^{12} actual\_amount_i)$$

We report the rolling TAPE from the perspective of Q1 and Q2 in 2017.

Results and Discussion: As Tables 7 and 8 show, our models are beating the standard Time Series Forecasting and traditional methods used in corporate finance. Our best models provide a 4-43x reduction in the error rates depending on the company on the AR side. On the AP side, we perform significantly (2-11x) better for four out of the five companies on the AP Side and perform competitively for the other one. This is partially due to Company F having a lot of volatility in their vendor selection process leading to a lot of new vendors over time. In addition to superior performance, given that we are able to aggregate from the transaction level data, we can also connect our forecasts to the exact invoices and bills that can optimize future working capital as shown in Figure 7. Hence, beyond just predicting the future, we help the end-user *to optimize* their future cash flows by taking the appropriate actions in the present.

AR TAME	Our Model	YoY Growth	ARIMA
A	**0.08**	0.25	0.33
C	**0.02**	0.54	0.26
D	**0.05**	0.34	0.32
F	**0.009**	0.152	0.17
G	**0.03**	0.549	1.31

Table 7: Account Receivables (AR) TAME on 12 month ahead Backtested Projections for FY 2017-2018. *Note: B, E were two companies that had less than 6 months of relevant data to do a proper year long backtest and were likewise excluded.*

AP TAME	Our Model	ARIMA	YoY Growth
A	**0.14**	0.29	0.25
C	**0.15**	0.27	0.41
D	**0.02**	0.22	0.37
F	0.32	0.55	**0.31**
G	**0.06**	0.16	0.55

Table 8: Account Payables (AP) TAME on 12 month ahead Backtested Projections for FY 2017-2018. *Note: B, E were two companies that had less than 6 months of relevant data to do a proper year long backtest and were likewise excluded.*

5. CONCLUSION

In the paper, we presented a Cash Management and Forecasting platform that leverages advances in machine learning to align executive and investor level decision-making with the actions taken by Treasurers and Collections Agents on the ground. Our models: 1.) Provide a significant lift across all forecasting tasks; 2.) Leverage the network effect present in our datasets to improve quality of overall forecasts across all companies. Given the complexities of modern day business operations and accounting regulation, this work sets the foundation for more efficient capital allocation for businesses of all sizes.

6. ACKNOWLEDGEMENTS

We would like to thank Fabio Flietas and Peter Hadlaw on building integrations with NetSuite and helping edit this paper.

7. REFERENCES

[1] S. Chaudhari and A. Ghone. Erp software market: Global opportunity analysis and industry forecast, 2013 - 2020. *Allied Market Research*, 2015.

[2] C. Fields, L. Barbara, and J. C. Schmidt. Technology solutions for cash forecasting. May 18 2018.

[3] K. Firestone. General electric's flow of financial information has become fantastically muddled. *The Economist*, January 30 2018.

[4] P. Hu. Predicting and improving invoice-to-cash collection through machine learning. 2015.

[5] D. McCarthy and P. Fader. Valuing subscription-based businesses using publicly disclosed customer data. *SSRN*, December 10 2015.

[6] D. McCarthy and P. Fader. Customer-based corporate valuation for publicly traded non-contractual firms. *SSRN*, March 9 2018.

[7] S. L. Scott and H. Varian. Predicting the present with bayesian structural time series. June 28 2013.

[8] S. J. Taylor and B. Letham. Forecasting at scale. *PeerJ Preprints 5:e3190v2*, September 27 2017.

[9] S. Zeng, P. Melville, C. Lang, I. Boier-Martin, and C. Murphy. Using predictive analysis to improve invoice-to-cash collection. *14th Conference on Knowledge Discovery and Data Mining(KDD-08)*, August 24 2008.